UNIVERSITÄT BAYREUTH

Towards a Synthetic Protein Motor:

Engineering the Lactose Repressor as the Scaffold for a Non-Autonomous Bipedal Motor Architecture

DISSERTATION

Zur Erlangung des akademischen Grades
einer Doktorin der Naturwissenschaften (Dr. rer. nat.)
an der Fakultät für Biologie, Chemie und Geowissenschaften
der Universität Bayreuth

vorgelegt von

Anna Zink (M.Sc.)

geboren in Erlangen

Bayreuth, 2026

Die vorliegende Arbeit wurde im Zeitraum von Mai 2022 bis Mai 2026 in Bayreuth am Lehrstuhl Biochemie III unter Betreuung von Frau Professorin Dr. Birte Höcker angefertigt.

Vollständiger Abdruck der von der Fakultät für Biologie, Chemie und Geowissenschaften der Universität Bayreuth genehmigten Dissertation zur Erlangung des akademischen Grades einer Doktorin der Naturwissenschaften (Dr. rer. nat.)

Form der Dissertation:	Monographie
Dissertation eingereicht am:	21.05.2026
Zulassung durch Promotionskommission:	03.06.2026
Wissenschaftliches Kolloquium:	26.10.2026

Amtierender Dekan: Prof. Dr. Janosch Hennig

Prüfungsausschuss:

Prof. Dr. Birte Höcker (Gutachterin)

Prof. Dr. Andreas Möglich (Gutachter)

Prof. Dr. Jürgen Paul (Vorsitz)

Prof. Dr. Stephan Kümmel (Prüfer)

Die vorliegende Arbeit ist als Monographie verfasst.

Teile der Arbeit sind bereits in folgender Publikation erschienen:

Walking by design: how to build artificial molecular motors made of proteins

Patrik F Nilsson*, **Anna Zink***, Olivier M C Laprévote*, Chapin S Korosec, Nils Gustafsson, Neil O Robertson, Nancy R Forde, Ralf Eichhorn, Birte Höcker, Paul M G Curmi, *Nano Futures*, 2024, 8 (4), 042501 (* contributed equally to this work)

DOI: 10.1088/2399-1984/ad7d7e

Diese Publikation ist in der vorliegenden Arbeit mit der Literaturstelle [1] zitiert.

Table of Contents

List of Figures VIII
List of Tables XIV
List of Abbreviations XVI
Summary XXI
Zusammenfassung XXII
1 Introduction 25
1.1 The LacI/GalR protein family: an important part of gene expression 25
1.1.1 The model system LacI repressor 26
1.1.2 TreR: A trehalose-responsive member of the LacI/GalR Protein family 37
1.2 The periplasmic binding protein 39
1.2.1 PotF: a suitable natural clapping machine for motor design 39
1.3 Strategies for functional protein assemblies 41
1.3.1 Principles and applications of split-protein complementation 41
1.3.2 Covalent coupling strategies for modular protein assembly 43
1.4 The evolution of protein design 46
1.5 Natural molecular motors and artificial designs 52
1.5.1 Functional diversity and shared mechanistic principles of natural motor proteins 52
1.5.2 Artificial motors: from small molecules to proteins 58
1.6 Motivation and Aim of this work 65
2 Results 67
2.1 Architecture and mechanism of a novel artificial motor protein 67
2.1.1 Structural design principles of a LacI-based bipedal motor protein 67
2.1.2 Conceptual stepping mechanism of a LacI-based bipedal motor protein 70
2.2 Generating asymmetry through asymmetric DNA binding with a heterodimeric LacI 77
2.2.1 *In vitro* analysis of engineered LacI heterodimer variants reveals impaired stability. and insufficient heterodimer formation 77
2.2.2 Assessment of a splitGFP-based BiFC system for analysis of LacI oligomerisation states 84
2.2.3 SplitGFP promotes LacI assembly but does not yield a stable heterodimeric complex 87
2.2.4 A SpyTag/SpyCatcher-based strategy enables stable and functional heterodimerisation of monomeric LacI variants 92
2.3 Computational design of structural extension at responsive site in LacI 101

2.3.1 Biochemical and functional characterisation of LacI variants with designed extensions 104
2.3.2 Structural validation of the designed LacI extension by crystallisation of a DNA binding-domain-truncated variant 109
2.4 Design and assembly of the first foot of the proposed motor protein 115
2.4.1 Implementation and screening of the connector module for modular assembly 115
2.4.2 Heterodimerisation of extension variants for motor-protein foot assembly 120
2.5 Design of a second foot for a bipedal motor protein 125
2.5.1 Identification and initial characterisation of the scaffold for the second foot 125
2.5.2 Extension design and initial functional characterisation of extended LLhT variants 131
2.5.3 Engineering of the connector element for the second foot 135
2.5.4 Engineering of a monomeric LLhT precursor for heterodimer formation 137
2.5.5 Combination of design elements for the generation of a heterodimeric LLhT-based foot variant 140
2.5.6 Assessment of SpyCatcher integration into the structural extension LacI-Lev1 143
2.6 Design and biochemical characterisation of the PotF-based clapping machine 151
2.7 Assembly of the motor 157
3 Discussion 161
3.1 Context dependence and modular control of engineered LacI heterodimerisation 161
3.1.1 Divergence between *in vivo* assignment and *in vitro* behaviour 161
3.1.2 Distinct effects of splitGFP and SpyTag/SpyCatcher on engineered LacI heterodimerisation 162
3.2 RFdiffusion-guided structural extension of an allosterically active LacI domain 164
3.3 Selective effector recognition as a basis for ligand orthogonality 165
3.4 Kinetic origins of impaired DNA binding in heterodimeric motor feet 167
3.5 Thermal stability costs of implementing motor-protein design features 170
3.6 Effects of connector modules and structural extension on oligomeric state 172
3.7 Plasticity of the LacI interface and its implications for monomerisation of LLhT 174
3.8 Orthogonal Tag/Catcher systems as modular connectors for motor assembly: opportunities and architectural limitations 175
4 Conclusion and Outlook 177
5 Materials and Methods 179
5.1 Materials 179
5.1.1 Bacterial Strain 179

5.1.2 Plasmids 179
5.1.3 Sequencing Primers 179
5.1.4 Cloning Primers 180
5.1.5 Molecular Weight Marker 182
5.1.6 Enzymes 183
5.1.7 Chemicals 183
5.1.8 Kits 184
5.1.9 Consumables 184
5.1.10 Equipment 185
5.1.11 Culture Medium 187
5.1.12 Buffer and solutions 187
5.1.13 Software 189
5.2 Methods 190
5.2.1 Experimental Methods 190
5.2.2 Computational Methods 205
6 References 209
7 Supplementary Information 239
7.1 Additional information about constructs 239
7.2 Additional purification data: chromatograms and SDS-PAGE analyses 253
7.3 Characterisation of IPTG binding with fluorescence titration 281
7.4 Analysis of secondary structure content with circular dichroism spectroscopy 290
7.5 Analysis of thermal stability evaluated with a change of molar ellipticity determined with circular dichroism spectroscopy 292
7.6 Thermal unfolding and Ligand binding analysis 296
7.7 SEC-MALS evaluations determining the molecular weight 298
7.8 DNA binding analysis with SPR 302
7.9 Additional crystallographic data and evaluations 307
7.10 Fluorescence measurements 310
7.11 Wildtype PotF 311
Acknowledgement 312
(Eidesstattliche) Versicherungen und Erklärungen 315

List of Figures

Figure 1. Structure of tetrameric LacI (PDB:1LBG[28]) in the DNA-bound repressor state..27
Figure 2. Overview of the natural operator sequences (O_1-O_3) and a symmetric operator sequence 28
Figure 3. Protein-DNA interactions of LacI bound to its quasi-palindromic operator sequence. 29
Figure 4. Simplified mechanism of LacI target-site search and specific DNA binding.....30
Figure 5. Comparison of non-specifically bound and specifically bound LacI-DNA complex.....31
Figure 6. Structure of LacI effector molecules. 32
Figure 7. Ligand interactions within the LacI effector-binding site. 32
Figure 8. Comparison of hydrogen bond networks in the LacI effector binding site.....33
Figure 9. Structural comparison of LacI in the state without ligand and the inducer-bound state. 34
Figure 10. Structure of dimeric LacI lacking the DNA binding domain (PDB: 2P9H[30]), highlighting the C-terminal dimerisation interface.....35
Figure 11. Interaction patterns between monomer I and II of LacI and TreR.....36
Figure 12. Crystal structure of the TreR effector-binding domain in complex with T6P.....38
Figure 13. Structural comparison of PotF in the closed and open conformation 40
Figure 14. Stimulus-controlled and spontaneous assembly pathway of split-protein systems. 41
Figure 15. Overview of GFP and splitGFP complementation strategies.....42
Figure 16. Overview of SpyTag/Catcher and DogTag/Catcher systems and their natural origins, CnaB and RrgA. 44
Figure 17. Proposed mechanism of the isopeptide bond formation. 45
Figure 18. Overview of protein design strategies. 46
Figure 19. Conceptual comparison of input-output relationships in protein design and prediction methods. 49
Figure 20. Comparison of the workflows of the structure generation approaches used in this work. 50
Figure 21. Geometry of motion of molecular motor proteins. 52
Figure 22. Schematic illustration of two principal mechanochemical coupling mechanisms in motor proteins.....53
Figure 23. Stepping mechanisms of linear motor proteins. 55
Figure 24. Structural comparison of representative linear motor proteins. 56
Figure 25. Conceptual terminology for artificial motor protein design.....59
Figure 26. Examples of protein-based artificial molecular motors. 62
Figure 27. Comparison of processivity and speed across natural and artificial molecular motors.....63
Figure 28. Schematic overview of the stepwise design strategy pursued in this thesis toward a bipedal artificial motor protein. 65
Figure 29. Schematic overview of the LacI-based bipedal motor and its core elements. ...67
Figure 30. Proposed mechanism of the non-autonomous LacI-based walker controlled by sequential ligand pulses.....71
Figure 31. Proposed mechanism for an autonomous LacI-based walker. 74
Figure 32. Schematic representation of LacI engineering towards heterodimeric variants. 77
Figure 33. SDS-PAGE analysis of co-expressed LacI interface variants.....79
Figure 34. SEC-MALS analysis of variants with modified monomer-monomer interfaces. .80

Figure 35. Temperature-dependent changes in molar ellipticity of the LacI variants with modified monomer-monomer interface 82
Figure 36. Fluorescence-based analysis of IPTG binding to LacI variants. 83
Figure 37. Control constructs used to evaluate splitGFP complementation as an *in vivo* readout for LacI heterodimer formation. 85
Figure 38. SDS-PAGE analyses of splitGFP fusion constructs. 86
Figure 39. SEC and SEC-MALS analysis of cleaved and uncleaved LacI-splitGFP variants 88
Figure 40. CD-based thermal unfolding analysis of LacI-splitGFP variants. 89
Figure 41. SPR sensograms of LacI-GFP binding to asymmetric and symmetric DNA. 91
Figure 42. Design strategy for the generation of a heterodimeric LacI construct base on SpyTag/SpyCatcher coupling. 92
Figure 43. Three-step purification of Het-SpyTag×Catcher. 93
Figure 44. SEC-MALS characterisation of SpyTag- and SpyCatcher-fused LacI variants 94
Figure 45. IPTG binding analysis of SpyTag/SpyCatcher-linked LacI constructs by fluorescence spectroscopy. 96
Figure 46. SPR analysis of Het-SpyTag×Catcher binding to different operator DNA sequences 97
Figure 47. CD-based thermal unfolding analysis of SpyTag/SpyCatcher-linked LacI constructs. 98
Figure 48. Identification of conformationally responsive regions in LacI based on backbone dihedral-angle changes. 101
Figure 49. Overview of two computational design pipelines for engineering structural extension into LacI. 102
Figure 50. Final set of selected LacI extension designs generated using RFdiffusion and ProteinMPNN. 103
Figure 51. Far-UV CD spectra of LacI variants carrying designed extensions. 104
Figure 52. CD-based thermal unfolding analysis of LacI extension variants and wild-type LacI. 105
Figure 53. SEC-MALS characterisation of LacI extension variants and wild-type LacI 106
Figure 54. Fluorescence-based analysis of IPTG binding to LacI variants carrying designed extensions. 107
Figure 55. SPR sensorgrams of LacI structural extension variants binding to symmetric operator DNA. 108
Figure 56. Differential scanning calorimetry profiles of LacI-Lev1 and PBP1. 110
Figure 57. Comparison of the PBP1 AlphaFold2 model with the experimentally determined crystal structure. 111
Figure 58. Structural comparison of ligand occupancy and conformational states in PBP1. 112
Figure 59. Schematic representation of DogTag integration into the LacI-Lev1 structural extension. 115
Figure 60. CD-based thermal unfolding analysis of LacI DogTag-linker variants and wild-type LacI. 117
Figure 61. SPR analysis of linker variants' interaction with symmetric DNA. 119
Figure 62. Schematic representation of the modular assembly of a LacI heterodimer. 121
Figure 63. Experimental validation of heterodimer formation and oligomeric state. 122
Figure 64. DSC analysis of LacI-Genti and LacI in the presence of different ligands. 126
Figure 65. SEC-MALS characterisation of LLhT, LLhF, and LacI. 128

Figure 66. DSC analysis of LLhT and LacI in the presence of different ligands. 129
Figure 67. SPR analysis of LacI and LLhT binding to symmetric operator DNA in the presence of different ligands. 130
Figure 68. Schematic representation of two different methods to generate structural extension into the lower PBP domain of LLhT. 131
Figure 69. Final set of selected LLhT extension designs generated using RFdiffusion and ProteinMPNN. 132
Figure 70. SEC-MALS characterisation of LLhT and LLhT extension variants. 133
Figure 71. SPR analysis of LLhT-Lev7 binding to symmetric operator DNA in the absence and presence of trehalose-6-P. 134
Figure 72. Schematic representation of SpyCatcher integration into the PBP domain of LLhT 136
Figure 73. Analytical size-exclusion chromatography analysis of LLhT interface variants 138
Figure 74. Schematic representation of the modular assembly of LLhT-based heterodimer. 140
Figure 75. SEC-MALS characterisation of LLhT interface and DogCatcher fusion variants 142
Figure 76. Schematic representation of the modular assembly of a LacI heterodimer mediated by DogTagxCatcher. 143
Figure 77. CD-based thermal unfolding analysis of LacI DogTagxCatcher heterodimer and its corresponding monomeric subunits. 145
Figure 78. SEC-MALS analysis of LacI-A2-Link1-DogTag and LacI*-DogCatcher. 146
Figure 79. Validation of Het-Link1-DogTag×DogCatcher formation by SDS-PAGE and SEC-MALS. 148
Figure 80. SPR analysis of Het-Link1-DogTagxCatcher binding to asymmetric DNA operator sequence. 149
Figure 81. Conceptual and structural basis of the proposed clapping machine. 152
Figure 82. SEC and SDS-PAGE analysis of the PotF-based clapping-machine construct. 152
Figure 83. SEC-MALS analysis of the two CM-PotF species separated during purification 153
Figure 84. DSC analysis of the clapping-machine construct in the absence and presence of putrescine. 154
Figure 85. Schematic representation of the annealing steps. 157
Figure 86. SDS-PAGE and SEC-MALS analysis of stepwise motor assembly. 159
Figure 87. Structural comparison of disaccharide ligands of LacI and variants analysed in this thesis. 166
Figure 88. Purification of LacI-A1 by Ni^{2+}-IMAC and SEC. 253
Figure 89. Purification of LacI-B1 by Ni^{2+}-IMAC and SEC. 254
Figure 90. Purification of LacI-A2 by Strep-Tactin AC. 254
Figure 91. Purification of LacI-B2 by Ni^{2+}-IMAC and SEC. 255
Figure 92. Purification of LacI-B3 by Ni^{2+}-IMAC and SEC. 255
Figure 93. Purification analysis of co-expressed LacI-A1 and LacI-B1 by Ni^{2+}-IMAC and SEC. 256
Figure 94. Purification analysis of co-expressed LacI-A2 and LacI-B2 by Ni^{2+}-IMAC and Strep-Tactin AC. 257

Figure 95. Purification analysis of co-expressed LacI-A2 and LacI-B3 by Ni^{2+}-IMAC and Strep-Tactin AC. 258
Figure 96. Chromatograms of splitGFP fusion constructs used as controls for the design of a BiFC assay. 259
Figure 97. Purification and proteolytic processing of LacI-splitGFP. 260
Figure 98. Chromatograms of the three-step purification of Het-SpyTagxCatcher. 261
Figure 99. Purification of LacI*-SpyCatcher by Strep-Tactin AC and SEC. 261
Figure 100. Purification of LacI-A2-SpyTag by Ni^{2+}-IMAC and SEC. 262
Figure 101. Purification of LacI-Lev1 by Ni^{2+}-IMAC, reverse Ni^{2+}-IMAC after TEV cleavage and SEC. 263
Figure 102. Purification of LacI-Lev2 by Ni^{2+}-IMAC, reverse Ni^{2+}-IMAC after TEV cleavage and SEC. 264
Figure 103. Purification of PBP1 by reverse Ni^{2+}-IMAC after TEV cleavage and SEC. 265
Figure 104. Purification of PBP2 by reverse Ni^{2+}-IMAC after TEV cleavage and SEC. 265
Figure 105. Purification of LacI-Link1 by Strep-Tactin AC and SEC. 266
Figure 106. Purification of LacI-Link2 by Strep-Tactin AC and SEC. 266
Figure 107. Purification of LacI-Link3 by Strep-Tactin AC and SEC. 267
Figure 108. Purification of LacI-Link4 by Strep-Tactin AC and SEC. 267
Figure 109. Purification of LacI-Link5 by Strep-Tactin AC and SEC. 268
Figure 110. Purification of LacI-A2-Link1-SpyTag by Ni^{2+}-IMAC and SEC. 268
Figure 111. Purification of LacI-A2-Link2-SpyTag by Ni^{2+}-IMAC and SEC. 269
Figure 112. Purification of Het-Link1-SpyTag×Catcher by sequential Strep-Tactin AC, Ni^{2+}-IMAC, and analytical SEC. 269
Figure 113. Purification of Het-Link2-SpyTag×Catcher by sequential Strep-Tactin AC, Ni^{2+}-IMAC, and analytical SEC. 270
Figure 114. Purification of LacI-Gentio by Ni^{2+}-IMAC, reverse Ni^{2+}-IMAC after TEV cleavage, and SEC. 271
Figure 115. Purification of LLhT by Strep-Tactin AC and SEC. 272
Figure 116. Purification of LLhF by Strep-Tactin AC and SEC. 272
Figure 117. Purification of LLhT-Lev1 by Ni^{2+}-IMAC and SEC. 273
Figure 118. Purification of LLhT-Lev7 by Ni^{2+}-IMAC and SEC. 273
Figure 119. Purification of LLhT-Link6 by Ni^{2+}-IMAC and SEC. 274
Figure 120. Purification of LLhT-Link7 by Ni^{2+}-IMAC and SEC. 274
Figure 121. Purification of LLhT-L244A by Strep-Tactin AC and SEC. 275
Figure 122. Purification of LLhT-L268A by Strep-Tactin AC and SEC. 275
Figure 123. Purification of LLhT-F271A by Strep-Tactin AC and SEC. 276
Figure 124. Purification of LLhT-L272A by Strep-Tactin AC and SEC. 276
Figure 125. Purification of LLhT-L272S by Strep-Tactin AC and SEC. 277
Figure 126. Purification of LLhT-4Mut by Strep-Tactin AC and SEC. 277
Figure 127. Purification of LLhT-6Mut by Strep-Tactin AC and SEC. 278
Figure 128. Purification of LLhT-6Mut-DogCatcher by Ni^{2+}-IMAC and SEC. 278
Figure 129. Purification of LacI*-DogCatcher by Ni^{2+}-IMAC and SEC. 279
Figure 130. Purification of LacI-A2-Link1-SpyCatcher-DogTag by Ni^{2+}-IMAC and SEC. 279
Figure 131. Purification of Het-Link1-DogTagxCatcher by Ni^{2+}-IMAC and SEC. 280
Figure 132. Purification of CM-PotF by Ni^{2+}-IMAC and SEC. 280
Figure 133. Fluorescence-based analysis of IPTG binding to ligand-deficient variant LacI-R197G. 281
Figure 134. Fluorescence-based analysis of IPTG binding to LacI interface variants. 282

Figure 135. Fluorescence-based analysis of IPTG binding to variants of heterodimeric LacI. 283
Figure 136. Fluorescence-based analysis of IPTG binding to LacI extension variants. 284
Figure 137. Fluorescence-based analysis of IPTG binding to truncated LacI extension variants lacking the DNA binding domain. 285
Figure 138. Fluorescence-based analysis of IPTG binding to LacI extension variants with DogTag. 287
Figure 139. Fluorescence-based analysis of IPTG binding to heterodimeric Linker variants. 288
Figure 140. Fluorescence-based analysis of IPTG binding to heterodimeric Linker variants. 289
Figure 141. Circular dichroism analysis of LacI linker variants. 290
Figure 142. Circular dichroism analysis of heterodimeric LacI linker variants. 290
Figure 143. Circular dichroism analysis of LLhT structural extension variants. 291
Figure 144. Circular dichroism analysis of LLhT linker variants. 291
Figure 145. CD-based thermal unfolding analysis of truncated PBP variants and wild-type LacI. 292
Figure 146. CD-based thermal unfolding analysis of heterodimeric LacI linker variants and their monomeric counterparts. 292
Figure 147. CD-based thermal unfolding analysis of chimeric LacI variants and wild-type LacI. 293
Figure 148. CD-based thermal unfolding analysis of LLhT extension variants and parental LLhT. 293
Figure 149. CD-based thermal unfolding analysis of monomeric LLhT and parental LLhT. 294
Figure 150. CD-based thermal unfolding analysis of monomeric LLhT-DogCatcher and parental LLhT. 294
Figure 151. Differential scanning calorimetry analysis of LLhT-Lev7 in the absence and presence of T6P. 296
Figure 152. Differential scanning calorimetry analysis of LLhT-6Mut in the absence and presence of T6P. 296
Figure 153. SEC-MALS characterisation of LacI-linker variants. 298
Figure 154. SEC-MALS analysis of LLhT DogTag insertion variants. 298
Figure 155. SEC-MALS characterisation of interface mutations of LLhT. 299
Figure 156. SPR analysis of LacI-D-cleavedGFP binding to symmetric DNA (pink) and asymmetric DNA (purple). 302
Figure 157. Concentration-dependent SPR analysis of LacI and Het-SpyTag×SpyCatcher binding to symmetric and asymmetric operator DNA. 302
Figure 158. SPR analysis of LacI*-SpyCatcher and LacI-A2-SpyTag binding to different operator DNA sequences. 303
Figure 159. Concentration-dependent SPR analysis of LacI-Lev1 and LacI-Lev2 binding to symmetric operator DNA. 304
Figure 160. Concentration-dependent SPR analysis of LacI-Link1 and LacI-Link2 binding to symmetric operator DNA. 304
Figure 161. SPR analysis of Het-Link1-SpyTagxCatcher and Het-Link2-SpyTagxCatcher binding to asymmetric operator DNA. 305
Figure 162. Concentration-dependent SPR analysis of LLhT binding to symmetric operator DNA. 305

Figure 163. Concentration-dependent SPR analysis of LLhT-Lev7 binding to symmetric operator DNA. 306
Figure 164. SPR analysis of Het-Link1-DogTagxCatcher binding to different operator DNA sequences. 306
Figure 165. Analysis of the crystal structure of chain B 308
Figure 166. Structural comparison of PBP1 with wild-type LacI. 309
Figure 167. OD-normalised GFP fluorescence of splitGFP control constructs 310
Figure 168. SEC chromatogram of PotF after Ni^{2+}-IMAC. 311

List of Tables

Table 1. Sequence conservation between selected repressors and monomeric periplasmic binding proteins 37
Table 2. Overview of LacI variants with modified monomer-monomer interface analysed for heterodimer formation. 78
Table 3. Apparent melting temperature (T_{Melt}) of the LacI variants with varied monomer-monomer interface. 82
Table 4. Apparent melting temperatures (T_{Melt}) of LacI-splitGFP-derived fractions and reference LacI variants. 89
Table 5. Apparent dissociation constants (K_D) for IPTG binding to heterodimeric LacI (Het-SpyTagxCatcher) and its monomeric subunits (LacI-A2-SpyTag, LacI*-SpyCatcher) 96
Table 6. Kinetic parameters of DNA binding for the heterodimeric construct and wild-type LacI were determined by SPR. 98
Table 7. Apparent melting temperatures (T_{Melt}) of SpyTag/SpyCatcher-linked LacI constructs. 98
Table 8. Apparent melting temperatures (T_{Melt}) of LacI extension variants and wild-type LacI. 105
Table 9. Apparent dissociation constants (K_D) for IPTG binding to LacI extension variants and wild-type LacI. 107
Table 10. Kinetic parameters of DNA binding for the LacI extension variants and wild-type LacI determined by SPR. 108
Table 11. Apparent melting temperatures (T_{melt}) of truncated PBP variants, corresponding full-length LacI extension constructs, and wild-type LacI. 109
Table 12. Apparent melting temperatures (T_{Melt}) of LacI DogTag-linker variants and the parent construct LacI-Lev1. 117
Table 13. Kinetic parameters of DNA binding for the LLhT and wild-type LacI were determined by SPR. 130
Table 14. Kinetic parameters of DNA binding for the LLhT-Lev7 with LLhT in comparison to LacI-Lev1 and wild-type LacI were determined by SPR. 135
Table 15. Apparent melting temperatures (T_{Melt}) of LacI DogTagxCatcher heterodimer and its corresponding monomeric subunits 145
Table 16. Thermal transition temperatures of the PotF-based clapping-machine construct 154
Table 17. Overview of the used sequencing primers. 179
Table 18. Overview of the primers used for SDM. 180
Table 19. Overview of primer used for the amplification of linearised pDUET with RE-site overhang within the first and second reading frame. 181
Table 20. Overview of primers used for the amplification of the inserts for the incorporation into both reading frames of pDUET 181
Table 21. Overview of primers used for the amplification of linearised pBAD-HisA with *SalI* and *XhoI* RE-sites 181
Table 22. Overview of primers used for the amplification of SpyCatcher, adding RE-sites *SalI* and *XhoI*. 182
Table 23. Overview of primers used for the amplification of linearised pCDF-BAD with *SalI* and *EcoRI* RE-sites 182
Table 24. Overview of primers used for the amplification of SpyTag, adding RE-sites *SalI* and *EcoRI*. 182

Table 25. Overview of primers used for the removal of the DNA-binding domain for LacI-Lev1 and LacI-Lev2 to generate PBP1 and PBP2 variants. 182
Table 26. Composition of PCR for site-directed mutagenesis 190
Table 27. PCR cycling parameters for site-directed mutagenesis 190
Table 28. Composition of PCR for the incorporation of restriction enzyme recognition sites 191
Table 29. PCR cycling parameters for the incorporation of restriction enzyme recognition sites 191
Table 30. Composition of cleavage reaction mix for the reaction with restriction endonucleases (RE I and RE II). 192
Table 31. Summary of restriction endonucleases used for vector-specific cloning 192
Table 32. Composition of the ligation reaction for a molar ratio of 1:3 193
Table 33. Summary of primers used for colony PCR 194
Table 34. Colony PCR reaction composition 194
Table 35. Colony PCR cycling parameters 194
Table 36. DNA sequences used for SPR measurements 203
Table 37. Summary of the purification steps performed for each construct. 239
Table 38. Overview of the analysed protein constructs and their calculated molecular properties. 240
Table 39. Overview of the constructs and corresponding amino acid sequences. 242
Table 40. Apparent melting temperatures (T_{Melt}) derived from CD-based thermal unfolding experiments for LacI heterodimer with Linker extension and LLhT based variants. 295
Table 41. Thermal transition temperatures and unfolding enthalpies of LacI, LacI-Gentio and LLhT-variants determined by DSC. 297
Table 42. Molecular weights determined by SEC-MALS for variants generated to achieve a heterodimeric LacI 299
Table 43. Molecular weights determined by SEC-MALS for variants generated to generate a heterodimeric foot with LacI structural extension 300
Table 44. Molecular weights determined by SEC-MALS for variants generated to generate a heterodimeric foot based on LLhT 301
Table 45. Molecular weights determined by SEC-MALS for the clapping machine 301
Table 46. Molecular weights determined by SEC-MALS for the assembly of the motor protein 301
Table 47. Kinetic parameters of DNA binding for LacI linker variants were determined by SPR. 305
Table 48. X-ray crystallographic data collection and refinement statistics of PBP1. 307

List of Abbreviations

General Abbreviations and Technical Terms

AAA+	ATPases Associated with diverse cellular Activities
AS	Amino acid
BiFC	Biomolecular fluorescence complementation assay
CD	Circular dichroism spectroscopy
CV	Column volume
ddH_2O	Double-distilled water
dNTPs	Deoxynucleotide triphosphates
DNA	Deoxyribonucleic acid
DSC	Differential scanning calorimetry
eGFP	Enhanced green fluorescent protein
ESM	Evolutionary scale modelling
fw	forward
GPCRs	G protein-coupled receptors
HTH	Helix-turn-helix motif
IMAC	Immobilised metal affinity chromatography
IPA	invariant point attention
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Lysogeny broth
LOV	Light-oxygen-voltage
MPNN	Message passing neural network
MSA	Multiple sequence alignment
MW	Molecular weight
MWC	Monod-Wyman-Changeux
MWCO	Molecular weight cut-off
NMR	Nuclear Magnetic Resonance
o/N	Overnight
ONPF	o-Nitrophenyl-β-D-fucopyranoside
ONPG	o-Nitrophenyl-β-D-galactopyranoside
PBP	Periplasmic binding protein
PCR	Polymerase chain reaction
pLDDT	Predicted Local Distance Difference Test
PPI	Protein-protein interaction
QM/MM	Quantum mechanics/molecular mechanics

RE	Restriction endonuclease
RNA	Ribonucleic acid
rpm	Revolutions per minute
rv	reverse
SDM	Site-directed mutagenesis
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
SEC-MALS	Size-exclusion chromatography coupled to multi-angle light scattering
sfGFP	Superfolder green fluorescent protein
SI	Supplementary Information
splitGFP	Split green fluorescent protein
SPR	Surface Plasmon Resonance
T6P	Trehalose-6-phosphate
TAE	Tris-acetate-EDTA
TEV	Tobacco etch virus
x g	Relative centrifugal force
X-Ray	X-radiation

Natural proteins

ABP	Androgen-binding protein
ALBP	Adipocyte lipid binding protein
CcpA	Catabolite control protein A
CPR/CRP	Catabolite receptor protein
CytR	Cytidine regulator
FKBP-FRB	FK506-binding protein and FKBP-rapamycin binding domain of mTOR
FruR/Cra	Catabolite repressor activator
GalR	Galactose repressor
GFP	Green fluorescent protein
LacI	Lactose repressor
MglB	Methyl-galactoside-binding protein
PotF	Putrescine-binding periplasmic protein
PurR	Purine repressor
RbsB	Ribose-binding protein
RegB/RegA	Two-component regulator
TreR	Trehalose repressor

Protein constructs

A1-splitGFP	Monomeric control for the BiFC assay
CM-PotF	Clapping machine PotF
DogTag-PotF-SpyCatcher	Clapping machine PotF
Het-Link1-DogTagxCatcher	Heterodimer based on LacI-Link1 with DogTagxCatcher as the dimerisation module
Het-Link1-SpyTagxCatcher	Heterodimer based on LacI-Link1 with SpyTagxCatcher as the dimerisation module
Het-Link2-SpyTagxCatcher	Heterodimer based on LacI-Link2 with SpyTagxCatcher as the dimerisation module
Het-SpyTagxCatcher	Heterodimer wild-type control
LacI*	Monomeric LacI with point mutation Y282S
LacI*-SpyCatcher	Subunit for wild-type heterodimer
LacI-A1	Designed monomer-monomer interface variant A1
LacI-A2	Designed monomer-monomer interface variant A2
LacI-A2-Link1-SpyTag	Subunit for heterodimer Het-Link1-SpyTagxCatcher
LacI-A2-Link2-SpyTag	Subunit for heterodimer Het-Link2-SpyTagxCatcher
LacI-A2-SpyTag	Subunit for wild-type heterodimer Het-SpyTagxCatcher
LacI-B1	Designed monomer-monomer interface variant B1
LacI-B2	Designed monomer-monomer interface variant B2
LacI-B3	Designed monomer-monomer interface variant B3
LacI-Dimer-cleavedGFP	LacI variants with removed splitGFP to initially trigger the heterodimer formation of monomer-monomer interface variants, dimeric peak
LacI-Genti	Modified LacI to bind gentiobiose
LacI-Lev1	LacI with structural extension design 1
LacI-Lev2	LacI with structural extension design 2
LacI-Link1	LacI with structural extension and DogTag attached with ProteinMPNN sequence optimisation design 1
LacI-Link2	LacI with structural extension and DogTag attached with ProteinMPNN sequence optimisation design 2
LacI-Link3	LacI with structural extension and DogTag attached with ProteinMPNN sequence optimisation design 3
LacI-Link4	LacI with structural extension and DogTag attached with GS_4-Linker
LacI-Link5	LacI with structural extension and DogTag attached with $(EP_4)_2$-Linker

LacI-Monomer-cleavedGFP	LacI variants with removed splitGFP to initially trigger the heterodimer formation of monomer-monomer interface variants, monomeric peak
LacI-R197G	LacI with a point mutation to diminish IPTG binding
LacI-splitGFP	LacI variants with splitGFP to initially trigger the heterodimer formation of monomer-monomer interface variants
LLhF	LacI DNA-binding domain with the FruR effector domain
LLhT	LacI DNA-binding domain with TreR effector domain
LLhT-4Mut	LLhT with four-point mutations in the interface to generate a monomeric LLhT
LLhT-6Mut	LLhT with six-point mutations in the interface to generate a monomeric LLhT
LLhT-6Mut-DogCatcher	Subunit for LLhT-based heterodimer
LLhT-Lev1	LLhT with structural extension used for LacI-Lev1
LLhT-Lev7	LLhT with structural extension designed with RFdiffuion and ProteinMPNN design 1
LLhT-6Mut-Lev7-SpyCatcher-DogTag	Subunit for LLhT-based heterodimer
LLhT-Link6	LLhT with structural extension of LacI-Lev7 with SpyCatcher attached to the structural extension design 1
LLhT-Link7	LLhT with structural extension of LacI-Lev7 with SpyCatcher attached to the structural extension design 2
PBP1	LacI-Lev1 design without the DNA-binding domain
PBP2	LacI-Lev2 design without the DNA-binding domain
RBP-splitGFP	Positive control for BiFC assay, heterodimeric ribose binding protein with splitGFP attached to each subunit
WT-splitGFP	Control for BiFC assay, LacI with splitGFP halves attached to LacI WT

Summary

Converting chemical energy into directional mechanical work is the most crucial task of motor proteins in living organisms. Common functional principles, including controlled track binding, coordinated conformational coupling and directional stepping, are present for all natural motors despite their structural and mechanistic diversity. The ArtMotor project aims to recreate motor protein function from molecular building blocks, without the inherent motor protein function, by reproducing the principles of natural motor proteins.

The lac repressor (LacI), one of the most extensively studied transcriptional repressors, was repurposed in this work as the primary scaffold for designing an artificial bipedal motor protein. The results presented here establish several fundamental design requirements for both non-autonomous and autonomous LacI-based motor architectures. For this purpose, a modular assembly concept was developed that allows the integration of motor-relevant functional elements into a higher-order protein architecture, including a DNA-binding module, an oligomerisation module, and coupling elements.

A key advance was the establishment of a strategy in which LacI monomers were functionalised with an assembly domain to enable the formation of a heterodimeric LacI species. A heterodimeric architecture is essential for generating a bipedal motor design with asymmetric DNA-binding properties and therefore constitutes a critical prerequisite for directional motor function. AI-assisted protein design (e.g. RFdiffusion) enabled the introduction of rigid structural extensions into the N-terminal PBP-like domain of LacI as a first step towards establishing a connector element that couples LacI to an effector module. As the extension is functionally linked to the stability of the LacI-DNA complex, it represents a particularly relevant site for mechanical coupling and conformational signal transmission. The resulting constructs provide an important proof of principle that motion-relevant structural elements may be integrated into an allosterically active protein scaffold while preserving key aspects of its functional integrity. Screening of related LacI/GalR-family regulators identified variants with ligand orthogonality, thereby establishing a basis for selective control in future non-autonomous stepping mechanisms. However, current limitations of the system include impaired DNA binding, altered oligomeric states and the reduced thermal stability induced by different design routes.

This thesis establishes the key molecular components and design principles required for a LacI-based artificial motor protein. In addition to its relevance for the ArtMotor project, this thesis provides new insight into LacI as an engineerable allosteric scaffold and into the underlying modifications required to construct dynamic protein machines.

Zusammenfassung

Motorproteine erfüllen zentrale Funktionen in lebenden Organismen, indem sie chemische Energie in gerichtete mechanische Bewegung umwandeln. Trotz ihrer strukturellen und mechanistischen Vielfalt beruhen natürliche Motorproteine auf gemeinsamen funktionellen Prinzipien, darunter die kontrollierte Bindung an ein Substrat („Track“), die koordinierte Kopplung von Konformationsänderungen sowie gerichtete Schrittbewegungen. Das ArtMotor-Projekt zielt darauf ab, ein Motorprotein zu rekonstruieren, indem es molekulare Bausteine, die keine Motorfunktion besitzen, modifiziert und kombiniert.

LacI, einer der am häufigsten untersuchten Repressoren, wurde in dieser Arbeit als zentrales Gerüst zur Konstruktion eines künstlichen bipedalen Motorproteins verwendet. Die vorliegenden Ergebnisse definieren mehrere grundlegende Anforderungen sowohl für nicht-autonome als auch für autonome LacI-basierte Motorarchitekturen. Hierzu wurde ein modulares Assemblierungskonzept entwickelt, das die Kombination motorrelevanter funktioneller Elemente in eine übergeordnete Proteinarchitektur ermöglicht, einschließlich eines DNA-Bindungsmoduls, eines Oligomerisierungsmoduls sowie Kopplungselementen. Die Entwicklung einer Strategie, bei der LacI-Monomere mit zusätzlichen Assemblierungsdomänen funktionalisiert wurden, ermöglichte die Bildung eines heterodimeren LacI. Eine solche heterodimere Architektur ist essenziell für die Realisierung eines bipedalen Motor-Designs mit asymmetrischer DNA-Bindung und stellt damit eine zentrale Voraussetzung für eine gerichtete Motorfunktion dar.

KI-gestütztes Proteindesign (z.B. mittels RFdiffusion) ermöglichte die Einführung rigider struktureller Erweiterungen in die N-terminale PBP-ähnliche Domäne von LacI als ersten Schritt zur Etablierung eines Verbindungselements zur Kopplung von LacI an ein Effektor-Modul. Da diese Erweiterung funktionell mit der Stabilität des LacI-DNA-Komplexes verknüpft ist, stellt sie einen besonders relevanten Ansatzpunkt für die mechanische Kopplung und die Übertragung konformationeller Signale dar. Die resultierenden Konstrukte liefern einen wichtigen Machbarkeitsnachweis dafür, dass bewegungsrelevante Strukturelemente in ein allosterisch aktives Proteingerüst integriert werden können, ohne dabei zentrale funktionelle Eigenschaften vollständig zu verlieren.

Das Screening verwandter Regulatoren aus der LacI/GalR-Familie identifizierte Varianten mit Liganden-Orthogonalität und schafft damit eine Grundlage für die selektive Steuerung zukünftiger nicht-autonomer Schrittmechanismen.

Gleichzeitig definieren die erhaltenen Ergebnisse zentrale Limitierungen, darunter eine verringerte DNA-Bindung infolge der Verknüpfung mehrerer Designmerkmale, veränderte Gleichgewichte der Oligomerisierung sowie eine reduzierte Stabilität. Diese Erkenntnisse liefern eine rationale Grundlage für die zukünftige Optimierung der Motorfunktion, indem sie die molekularen Eigenschaften identifizieren, die für eine effiziente und zielgerichtete

Bewegung weiter verbessert werden müssen.
Diese Arbeit etabliert die wesentlichen molekularen Bausteine und Gestaltungsprinzipien für ein LacI-basiertes künstliches Motorprotein. Über ihre Bedeutung für das ArtMotor-Projekt hinaus liefert sie neue Einblicke in LacI als technisch nutzbares allosterisches Gerüst sowie in die grundlegenden Modifikationen, die für die Konstruktion dynamischer Proteinmaschinen erforderlich sind.

1 Introduction

1.1 The LacI/GalR protein family: an important part of gene expression

Under changing environmental and growth conditions, organisms rely on regulatory systems that adjust gene expression to meet cellular demands. An important part of this process is transcriptional control. Jacob and Monod established the conceptual basis for such transcriptional control in 1961, whose work on the *lac* operon provided one of the first mechanistic frameworks for bacterial gene regulation.[2] The LacI/GalR family constitutes one of the largest and most diverse groups of bacterial transcriptional regulators. Comparative genomic analyses have identified more than 1,000 homologs, consistent with broad phylogenetic representation and functional diversification within this family.[3] Members of the LacI/GalR family control the catabolic synthesis as well as the biosynthesis of nucleotides or the expression of toxins.[4–6] LacI/GalR-based repressors can function as activators of gene expression by enhancing binding of the RNA polymerase (e.g. the catabolite control protein A CcpA[7], the catabolite repressor activator FruR/Cra[8] or the two-component regulator RegB/RegA[9]), and general transcription factors that repress gene expression by blocking the transcription of DNA by the RNA polymerase (e.g. lactose repressor (LacI)[10], galactose repressor (GalR)[11], purine repressor (PurR)[12], trehalose repressor (TreR)[13]). The effector ligands allosterically control DNA binding. A widely used conceptual framework for describing repressor allostery is the Monod-Wyman-Changeux (MWC) model, which outlines the regulation as a ligand-driven shift of a pre-existing equilibrium between conformational states that differ in DNA-binding affinities.[2,14] Ligands can either decrease (inducer) or increase the DNA binding (anti-inducer), which differs for repressors of the LacI/GalR protein family.

Despite their functional diversity, LacI/GalR regulators share a conserved modular architecture comprising three principal regions[15]: (i) an N-terminal DNA-binding domain (Figure 1, red), (ii) a connecting segment (Figure 1, yellow), and (iii) a C-terminal effector-binding domain (Figure 1, blue), which are described in more detail below.

(i) In members of the LacI/GalR protein family, the DNA-binding domain consists of the helix-turn-helix motif (HTH).[15,16] This fold represents one of the most prevalent architectures for sequence-specific DNA recognition in bacterial transcription factors.[17] The motif comprises two helices connected by a short turn that constrains their relative orientation. The second helix, commonly known as the recognition helix, inserts into the major groove of the DNA and forms base-specific and sugar-phosphate contacts. The first helix can provide additional stabilising interactions and contribute to the geometry of DNA binding.[16,18–20]

(ii) A flexible linker consisting of around 18 residues connects the DNA-binding headpiece to the regulatory core. In approximately 60% of LacI/GalR homologues, this segment contains a conserved motif capable of forming a hinge helix.[21–24] The hinge helix is functionally important because it undergoes a disorder-to-order transition initiated by changes in the effector domain.

This provides an essential allosteric communication element within the protein family.[24] Even though many positions within the linker are not conserved on a sequence level, they are still important for transducing the allosteric function.[25,26]

(iii) The C-terminal regulatory domain adopts a fold related to the periplasmic binding protein (PBP) superfamily.[27] The effector domain corresponds to a type-I PBP-like architecture composed of two subdomains (N-terminal (Figure 1, light blue), and C-terminal (Figure 1, dark blue)) that are organised similarly and connected by a hinge.[27] Each subdomain exhibits an α/β/α layered fold, comprising a six-stranded parallel β-sheet flanked by α-helices.[28] Three polypeptide crossovers connect the subdomains.[29] The effector-binding site lies at the interface of the two subdomains. Binding results in a conformational change in the relative orientation of the subdomains, resembling a Venus-flytrap.[21,23,30] Across bacteria, LacI/GalR regulators display substantial diversity in effector specificity. In some cases, repressors share ligand preferences with homologous PBPs; for example, GalR and methyl-galactoside-binding protein (MglB) both bind D-galactose and ribose repressor (RbsR), and ribose binding protein (RbsB) both recognise D-ribose.[31] Beyond carbohydrate effectors, LacI/GalR family members also respond to chemically distinct regulatory inputs, exemplified by guanine-dependent regulation of PurR and the modulation of CytR activity through cooperative interaction with the catabolite receptor protein (CRP).[32,33]

1.1.1 The model system LacI repressor

1.1.1.1 Functional role of LacI as a transcriptional repressor

LacI is a canonical molecular switch and the central regulatory protein of the lactose pathway in prokaryotes.[10] It was identified as a regulatory factor by Jacob and Monod in 1961[2] and was isolated and purified by Gilbert and Müller-Hill in 1966[10] for the first time. Since then, LacI has evolved into one of the best-understood regulatory proteins, with extensive mutational[34–39], biochemical[40–44], structural[22,28,30], biophysical[45–48] and computational studies[49–52] elucidating the molecular basics of allosteric regulation, DNA recognition, DNA looping and ligand binding. Beyond its fundamental *in vivo* role, LacI has also been repurposed as a modular regulatory element in synthetic biology. This includes the use in engineered genetic logic circuits[53,54] and as a fluorescent tagging system for visualising chromosomal loci and studying segregation[45,55].

LacI is a homotetramer composed of four identical ~36.5 kDa subunits organised as a dimer of dimers (Figure 1).[28] As part of the LacI/GalR protein family, each monomer consists of the three key domains (i)-(iii). In contrast to many members of the protein family, where no tetramerisation domain is present, the tetramerisation domain of LacI is located at the C-terminus (Figure 1, purple), leading to the formation of dimers of dimers (Figure 1). In LacI, allolactose acts as a small-molecule inducer that binds within the effector-binding domain (Figure 1, blue). It alters the affinity of the spatially separated DNA-binding domains for operator DNA (Figure 1, red). A main contributor to the change in stability of the DNA-operator

complex is the hinge helix (Figure 1, yellow). It forms the connection between the DNA-binding domain and the effector-binding domain and adopts an ordered helical conformation upon operator binding, but becomes destabilised into a flexible loop upon inducer binding.[28,56] The order-to-disorder transition of the hinge-helix plays a crucial role in allosteric regulation, coupling inducer binding to a ~20-fold reduction in operator affinity.[57,58]

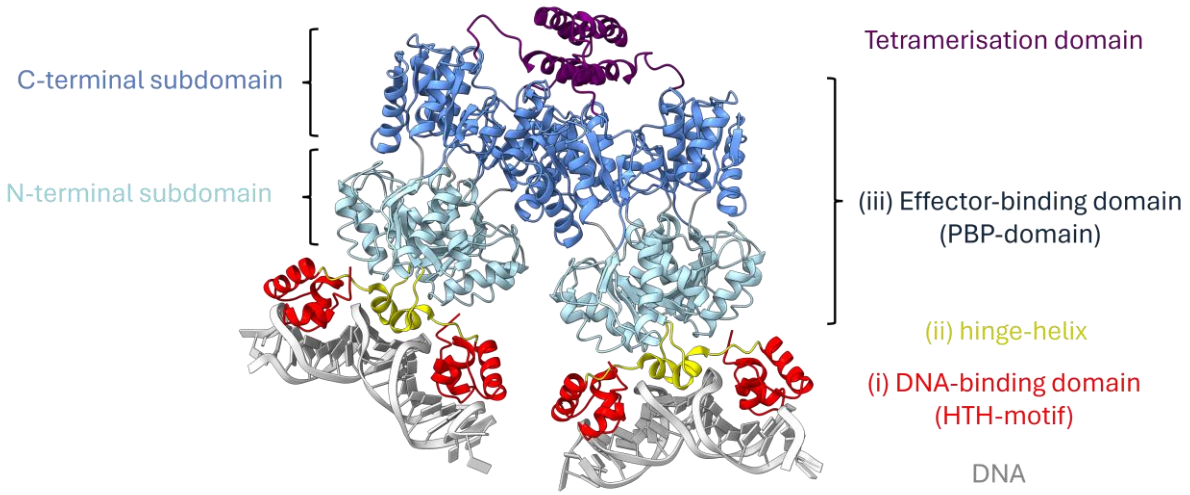


Figure 1. Structure of tetrameric LacI (PDB:1LBG[28]) in the DNA-bound repressor state. LacI is shown bound to the operator DNA (grey). The main structural modules are colour-coded: the DNA-binding domain (red), hinge-helix (yellow), effector-binding domain, subdivided into the N-terminal subdomain (light blue) and the C-terminal subdomain (blue), and the tetramerisation domain (purple).

1.1.1.2 Molecular mechanism of DNA recognition and repression

In nature, LacI binds with the highest affinity to the primary operator O_1.[59] This region overlaps with the lac promoter, thereby sterically blocking RNA polymerase binding and initiation. In addition to O_1, two auxiliary operator sites (O_2 and O_3) are located 401 bp downstream and 92 bp upstream of O_1, respectively, and play a critical role in gene regulation (Figure 2A).[60–63] While O_1 binds LacI with dissociation constants in the low picomolar range,[64] the auxiliary operators exhibit weaker affinities, approximately 10-fold (O_2) and up to 300-fold (O_3) lower compared to O_1 due to variations in the operator sequence (Figure 2B).[59] Despite their reduced intrinsic affinity, the simultaneous binding of tetrameric LacI to O_1 and an auxiliary operator induces DNA looping, enhancing transcriptional repression *in vivo*.[59,63,65,66]

Early low-resolution crystal structures did not reveal significant asymmetries between the left and right DNA-binding domains of LacI, even though LacI binds a quasi-palindromic operator containing a central G:C base pair (Figure 2B, box), as well as additional sequence differences between the two operator half-sites at positions 7 and 9 (Figure 2B, grey highlights).[67]

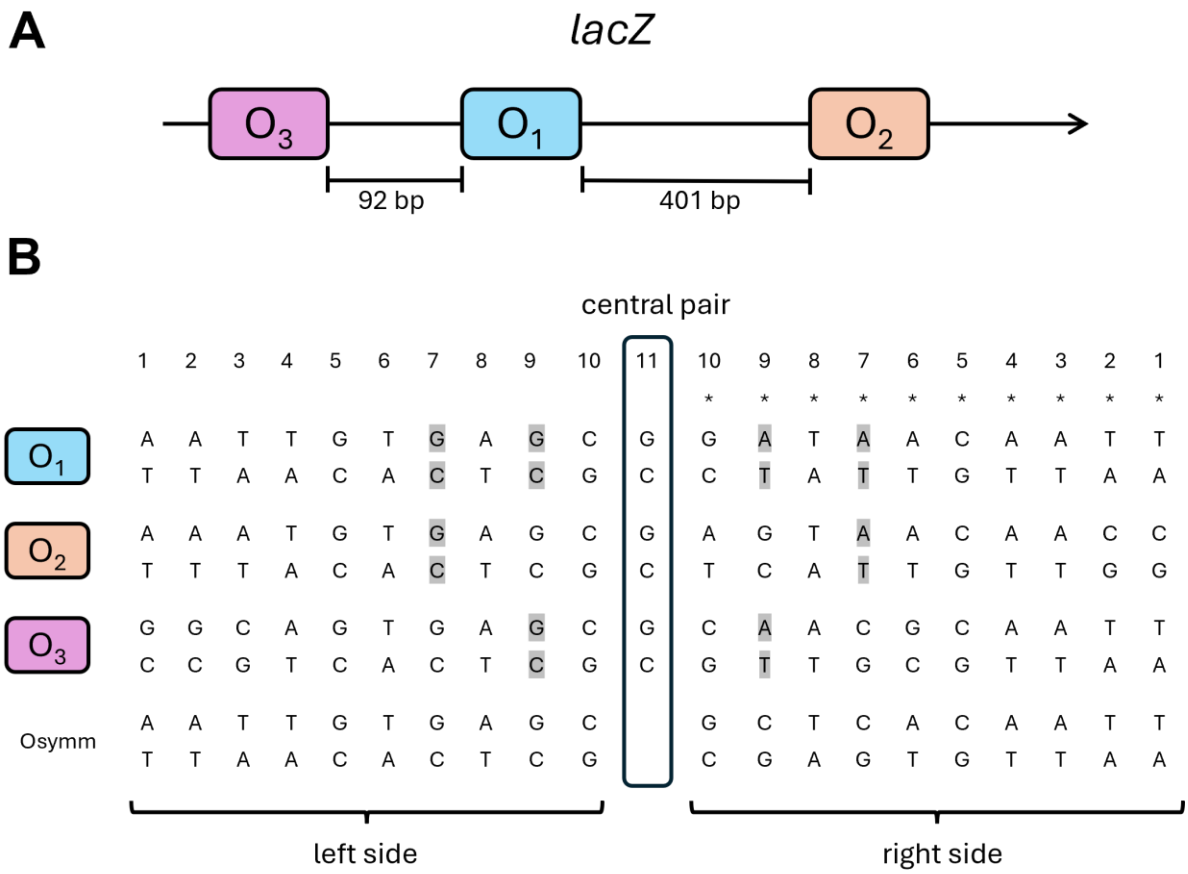


Figure 2. Overview of the natural operator sequences (O_1-O_3) and a symmetric operator sequence used for *in vivo* studies.[68] A) Schematic representation of the lacZ regulatory region, illustrating the arrangement of the auxiliary operator sites O_3 (pink) and O_2 (orange) relative to the primary operator O_1 (blue). Inter-operator distances are indicated below the schematic. B) Sequence alignment of O_1, O_2, O_3, and the symmetric operator O_{symm}. The left and right half-site flank the central G:C pair, which is boxed. Positions within the right half-site are marked by asterisks (*), and nucleotides that differ between the two half-sites are highlighted in grey.

Higher-resolution NMR analyses revealed pronounced asymmetry in the overall arrangement of the DNA-binding domain and the local interactions formed with the left and right sides of the quasi-palindromic operator sequence.[69] The hinge helix, specifically residue L56, is a major contributor to binding the central C:G pair and crucial for the cross-interaction between both sides (Figure 3, pink). L56 of the LacI subunit, which interacts mainly with the right operator side, also forms interactions with the base at position 10 of the left and right operator sites. Other residues of the hinge helix form more extensive contacts on the left side of the DNA than on the right (Figure 3, light green). Due to variations in the base at position 9, residues L6 and Y7 lose base-specific interaction with the right side of the operator compared to the left side (Figure 3, orange). Residues of the recognition helix, specifically Y17 and Q18 (Figure 3, blue), reorient their base-specific interactions to compensate for the changes at base 7 and 9. This is possible due to a 12° reorientation of the recognition helix between the left and right protein-DNA complexes, which alters the π-stacking arrangements involving Y7, Y17 and Q18.[69] Despite their adaptability when interacting with different bases on both operator halves, Y17

and Q18, those residues have been shown to play an essential role in specific operator recognition, highlighted by structural[22,28,30,70] and mutational studies[71–77].

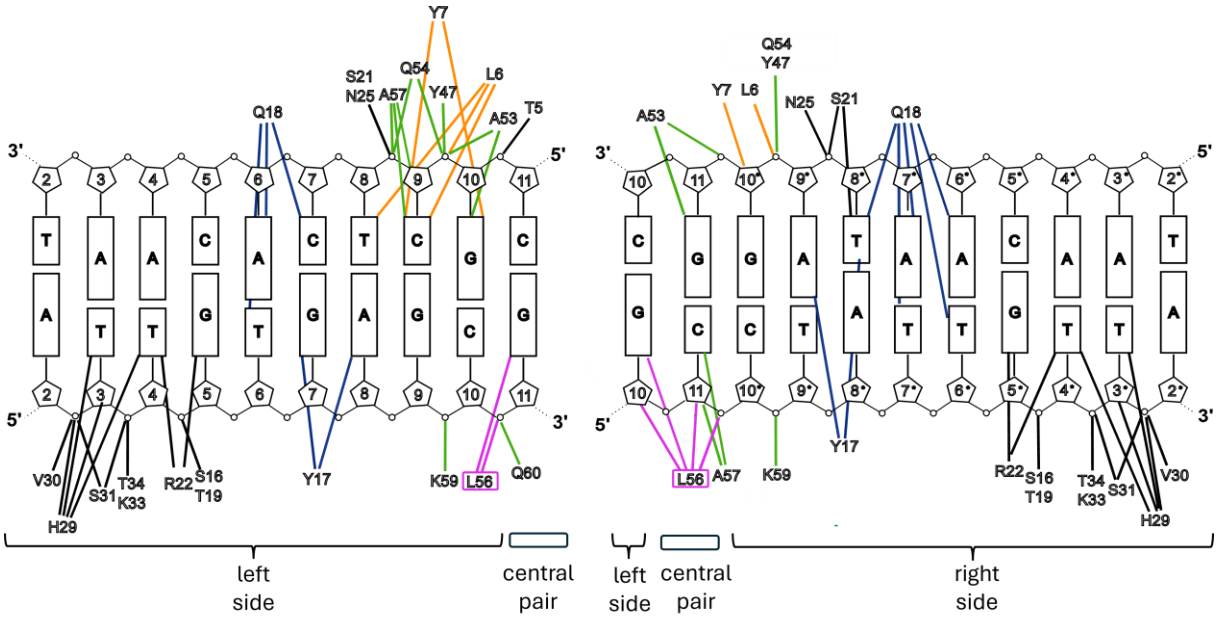


Figure 3. Protein-DNA interactions of LacI bound to its quasi-palindromic operator sequence. Interactions formed by the two LacI subunits are shown on the left and right sides of the figure, corresponding to the subunits that predominantly contact the left and right operator half-sites, respectively. The nucleotides of the left and right half-sites, together with the central base pair, are indicated at the bottom. Owing to a slight rotation of the recognition helix, the subunit predominantly associated with the right half-site also contacts base-pair position 10 within the left half-site (pink). This results in distinct interaction patterns across the two halves of the operator. Interactions are colour-coded: L56, pink; hinge helix, light green; L6 and L7, orange; residues Y17 and Q18 contacting nucleotides that differ between the left and right half-sites, blue. Adapted from Kalodimos *et al.*[69].

The critical missing piece in the physicochemical mechanisms of the specific protein-DNA recognition process is the association of regulatory proteins with nonspecific DNA, and their sliding along DNA until they bind specifically. The faster-than-diffusion association rates *in vivo* of LacI with operator-containing DNA have long been recognised. Kinetic studies by Hsieh *et al.* and Winter *et al.* demonstrated that LacI associates with its operator approximately two orders of magnitude faster than predicted by three-dimensional diffusion alone.[78–81]

Single-molecule imaging of *in vivo* DNA association, together with complementary *in vitro* studies, has established facilitated diffusion as an efficient and biologically relevant mechanism for interactions with the operator.[45,46,82] In this mechanism, LacI first undergoes three-dimensional diffusion until it encounters DNA, then LacI binds non-specifically to the DNA backbone (Figure 4, bottom left). Once non-specifically bound, the repressor diffuses one-dimensionally along the DNA through sliding, thereby efficiently locating its cognate operator sequence (Figure 4, bottom right).[79,81,83] In addition to facilitated diffusion, it has been proposed that tetrameric LacI engages in intersegmental transfer, allowing direct exchange between distant DNA segments or looped DNA without full dissociation into solution.[82]

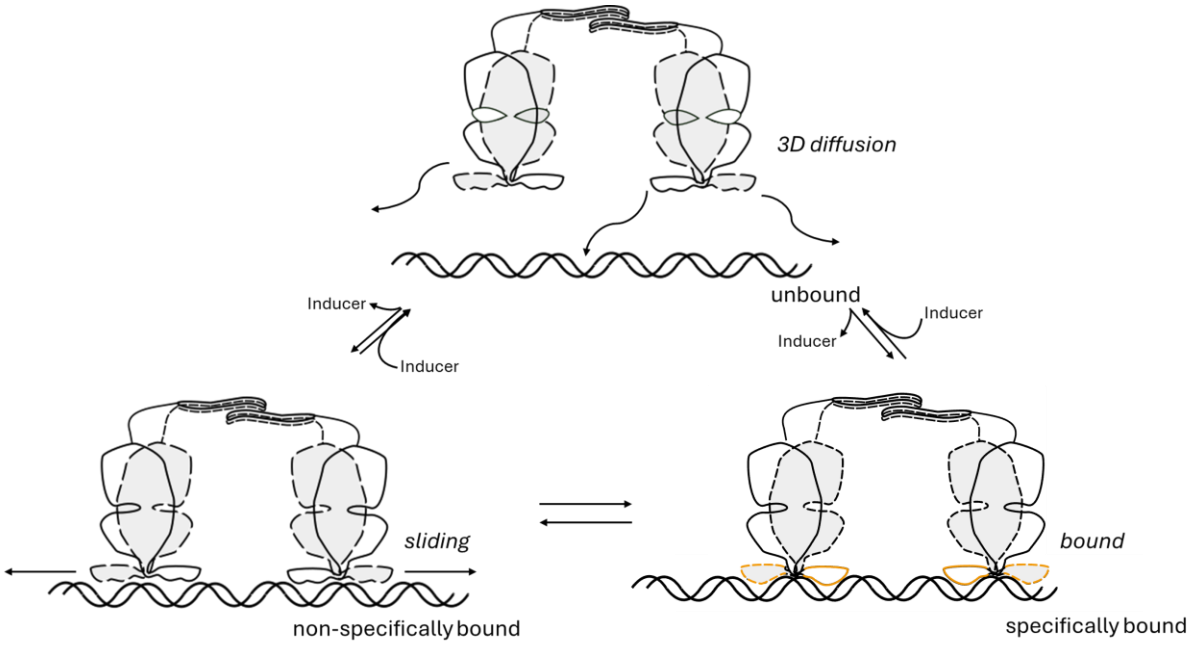


Figure 4. Simplified mechanism of LacI target-site search and specific DNA binding. The scheme illustrates the transition from three-dimensional diffusion to one-dimensional sliding along DNA and subsequent formation of the specifically bound state. Changes in ligand occupancy and the associated association and dissociation kinetics between ligand-bound and ligand-free states, as well as DNA-looping, are not explicitly considered. The DNA-binding interactions are highlighted in orange.

Upon specific operator recognition, the heat capacity (C_p) changes from −200 cal·mol^{-1}·K^{-} to −1110 cal·mol^{-1}·K^{-1}, indicating a thermodynamically driven induced-fit mechanism.[84] This mechanism involves the extensive conformational rearrangement of the solvent accessible surface area during the formation of the specific protein-DNA complex.[84] Changes in the arrangement of the DNA headpiece are supported by structural evidence. Upon specific operator binding, the recognition helix of the HTH motif (Figure 5A, blue) is positioned within the major groove of the DNA. At the same time, the hinge region undergoes a disorder-to-order transition to form a structured hinge helix (Figure 5A, green).[22] This hinge helix is disordered in the non-specifically bound state (Figure 5A, light green).[22] Additionally, non-specific LacI-DNA interactions are dominated by electrostatic forces[84], while specific operator binding is characterised by the formation of an extensive network of new interactions (Figure 5B, bold lines). Some of the interactions present in the non-specific complex are retained (Figure 5B, thin lines), and some electrostatic interactions are lost during the transfer from a non-specific to a specifically bound state (Figure 5B, dotted lines).

Looking more closely into the position of the changing interactions upon specific DNA recognition, it is shown that the first helix of the HTH-motif contributes to interactions with the DNA phosphate backbone as well as base-specific contacts (Figure 5B, bold orange). Additionally, the newly ordered hinge-helix forms seven new contacts with the DNA minor groove of the DNA (Figure 5B, bold green), predominantly involving the phosphate backbone, while residues 53, 56 and 57 (Figure 5B, green box) additionally engage in base-specific interactions with bases 10 and 8, respectively.[22]

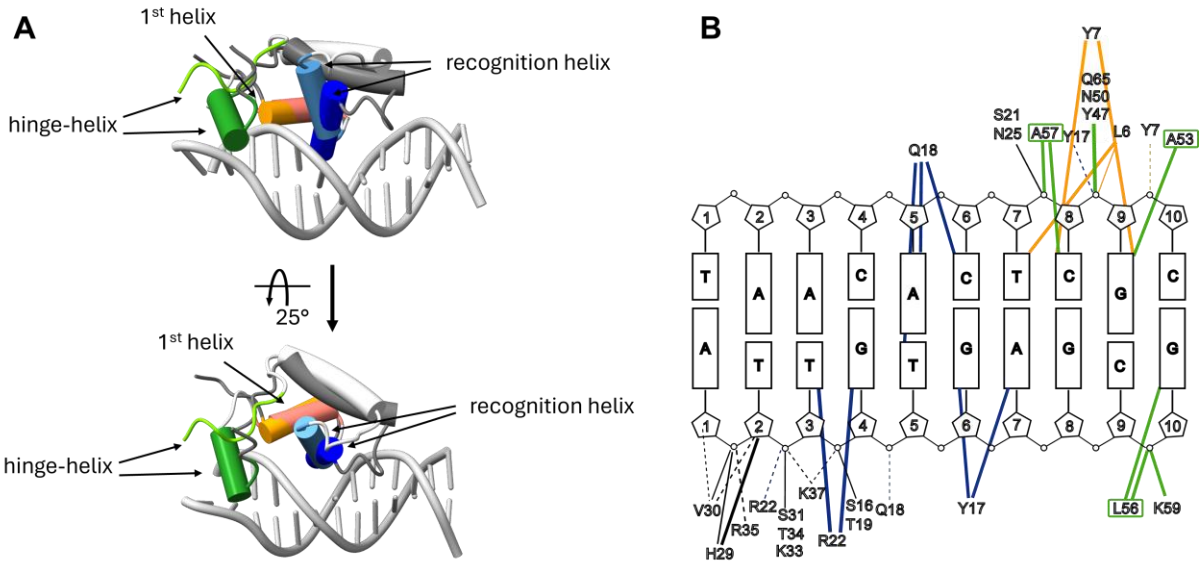


Figure 5. Comparison of non-specifically bound and specifically bound LacI-DNA complex. A) NMR structures comparing LacI specifically bound to the natural operator O_1 (PDB: 1L1M[69]) with non-specifically bound LacI (PDB:1OSL[85]). In the specific complex, the hinge helix is ordered (green), whereas it remains disordered in the non-specific complex (light green). The recognition helix is coloured dark blue and light blue, and the remaining protein dark orange and salmon, for the specific and non-specific states, respectively. B) Comparison of the interaction patterns between the non-specific and specific LacI-DNA complexes. Interactions unique to the non-specific complex are shown as dotted lines, interactions present in both states as thin lines, and interactions unique to the specific complex as bold lines. Colour coding indicates interactions of the hinge helix (green), the first helix (orange), and key residues of the recognition helix (blue), while additional interactions involving the recognition helix and the third helix of the DNA-binding domain are shown in black. Adapted figure from Kalidimos *et al.*[22].

Most important for base-specific interactions in the specifically bound LacI-DNA complex are residues 17, 18, and 22, located within the recognition helix. They form multiple new base-specific contacts (Figure 5B, blue) within the major groove. Given the importance of interactions between the left and right sides of the natural operator sequence, as well as for specific DNA interactions, these positions are crucial for specific DNA interactions. The potential of repurposing LacI to recognise modified operator sequences utilised for logical gates or asymmetric DNA binding has highlighted the importance of these three positions for specific operator recognition.[68,86,87]

1.1.1.3 Allosteric regulation by inducer binding to the PBP-like core domain

The effector-binding domain is subdivided into the N- and C-terminal subdomains, connected by three crossover regions[28,88], and serves as the primary monomer-monomer interface as well as the ligand binding site.[30,89] In nature, LacI is allosterically regulated by allolactose (Figure 6A), which serves as its physiological inducer.[90] Early binding studies[91] demonstrated that LacI recognises a range of galactosidase-related ligands, which are functionally grouped into three classes[47]: i) inducers that reduce the operator affinity (e.g. Isopropyl-β-D-thiogalactopyranosid (IPTG) (Figure 6B)), ii) neutral ligands that bind without measurably altering operator binding (e.g. o-Nitrophenyl-β-D-galactopyranoside (ONPG) (Figure 6C)), and

iii) anti-inducers that stabilise the repressed state and thereby enhance operator binding (e.g. o-Nitrophenyl-β-D-fucopyranoside (ONPF) (Figure 6D)).

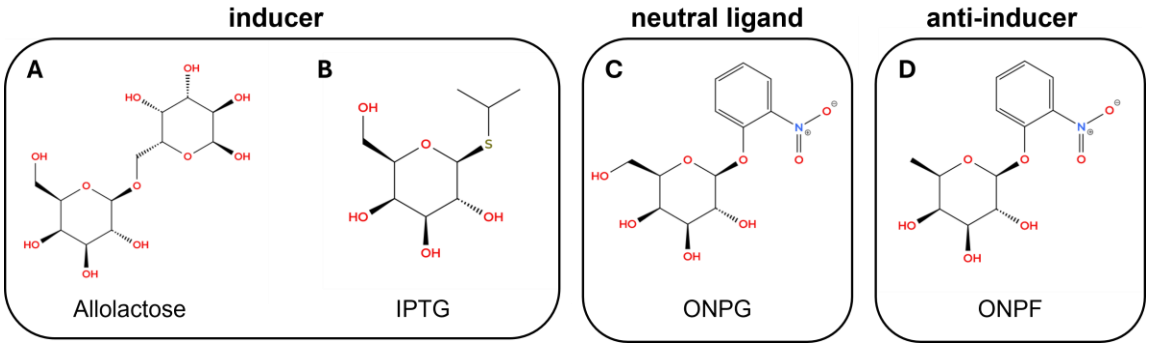


Figure 6. Structure of LacI effector molecules. A) The natural inducer allolactose, B) the synthetic inducer IPTG, C) the neutral ligand ONPG, and D) the anti-inducer ONPF. Molecular structures were generated using MolView[92].

LacI already binds a variety of different effectors, but with focused libraries or computational design, new LacI variants that respond to lactulose[93] or other sugars, such as gentiobiose[94], have been created. However, the main challenge is to remove the responsiveness to the inducer IPTG.[93,94]

Comparing the interactions of the inducer, neutral ligand, and anti-inducer has furthered our understanding of key interactions within the effector binding site (Figure 7).

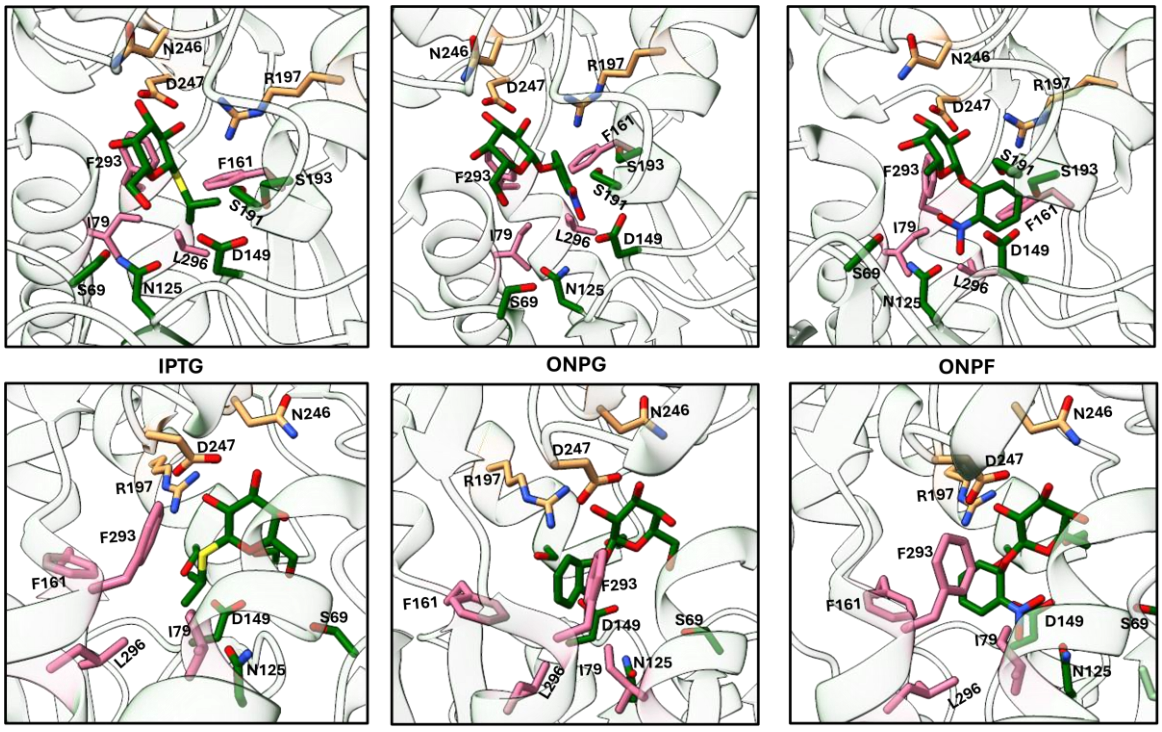


Figure 7. Ligand interactions within the LacI effector-binding site. The interactions between LacI and the inducer IPTG (left, PDB:2P9H[30]), the neutral ligand ONPG (middle, PDB:2PE5[30]), and the anti-inducer ONPF (right, PDB:2PAF[30]). Upper row: interactions with the O2, O3, and O4 hydroxyl groups; lower row: hydrophobic pocket accommodating the isopropyl or nitrophenyl group. Interacting residues are shown as sticks and colour-coded: O2/O3-contacting residues, orange; hydrophobic pocket, pink; O6 hydrogen-bonding network, green.

Comparing the different effectors, conserved interactions between O2 and O3 hydroxyl groups of the small molecules and R197, N246, and D274 of LacI are present in all complexes (Figure 7, orange). Furthermore, the ring oxygen O5 interacts with N125. The O4 hydroxyl participates in water-mediated hydrogen bonds involving the A75 backbone and N246.[28,30,95,96] A key difference in effector recognition arises from the distinct functional groups presented by the different ligands, which give rise to different interactions with the hydrophobic pocket formed by I79, F161, F293, and L296 (Figure 7, pink). Accordingly, this pocket accommodates either the isopropyl group of IPTG or the nitrophenyl ring of ONPF and ONPG. [28,30,95,96] A further difference in effector recognition lies in how the individual ligands interact with the O6 hydroxyl position (Figure 7 green; Figure 8). The interactions are particularly important for determining the distinct allosteric responses triggered upon effector binding.

A key step in the allosteric response is the formation of an extended hydrogen-bond network involving residues of the N- and C-terminal subdomains, namely S69, D149, N125, S191, and S193 (Figure 8). In comparison to the binding of the anti-inducer and the neutral ligand, additional hydrogen bonds form between S69 and D149 and the O6 hydroxyl oxygen of IPTG (Figure 8, blue lines), building a key nucleation point for an extended hydrogen-bond network.[28,30,95,96] ONPG lacks tight interactions with the hydrophobic pocket (Figure 7, pink), which prevents the formation of the hydrogen-bond network with the O6 hydroxyl, leading to no response upon effector interaction.[30] The anti-inducer ONPF lacks the O6 hydroxyl. However, its anti-inducing behaviour is best understood not as the consequence of a single missing interaction, but as the result of a distinct ligand orientation and interaction network within the binding pocket that stabilises a different allosteric ensemble.[30,40]

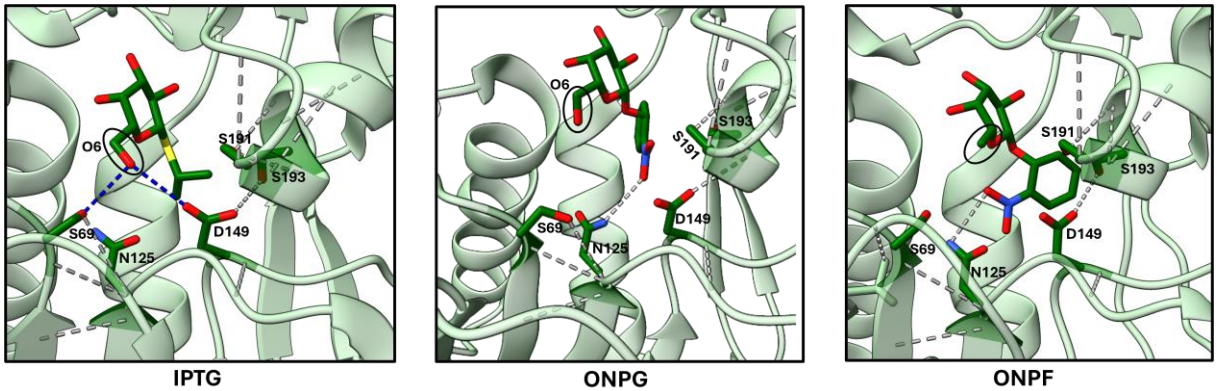


Figure 8. Comparison of hydrogen bond networks in the LacI effector binding site. Hydrogen bonds formed between key binding-site residues and IPTG (left, PDB:2P9H[30]), the neutral ligand ONPG (middle, PDB:2PE5[30]), or the anti-inducer ONPF(right, PDB:2PAF[30]) are shown. Grey dotted lines indicate hydrogen bonds; contacts involving O6, which are associated with IPTG-induced allosteric signalling, are highlighted in blue.

These observations define a coherent mechanistic framework for allosteric regulation in LacI. The central determinant is the ligand-specific engagement of the O6 hydroxyl group, which, in the case of the inducer IPTG, nucleates an extended hydrogen-bond network spanning both the N- and C-terminal subdomains. Residue-level description of the binding pocket has

revealed that 10% more intramolecular hydrogen bonds are formed in the inducer-bound state.[30] The ability of this region to reconfigure and extend hydrogen-bond connectivity is essential for allosteric regulation.[49,97] The network formation increases intramolecular connectivity and promotes a pronounced rigidification of the effector-binding pocket and the adjacent N-terminal subdomain.[40]

Structural characterisation of the inducer-bound (holo) and effector-free state (apo) of LacI revealed that the rigidification is associated with a structural change in the N-terminal subdomain. In contrast, the C-terminal subdomain showed no significant changes between the inducer-bound and operator-bound forms (Figure 9).[30,67] Involved in the rigidification of the N-terminal subdomain is K84. In the DNA-bound (open) state, K84 is buried within an apolar interfacial environment. In the inducer-bound state (closed), its reorientation towards a more solvent-exposed conformation is promoted (Figure 9, right, bold residue).[98–101] A broader reorganisation of the monomer-monomer interface accompanies this local conformational change. This is observed in the formation of a continuous intersubunit β-sheet, which brings two β-strands from opposing monomers (I and II) approximately 4 Å closer together (Figure 9, right). Residues 94-98 on either side of the monomer reorient and stabilise the formation of the intersubunit β-sheet of the N-terminal lobe.[102]

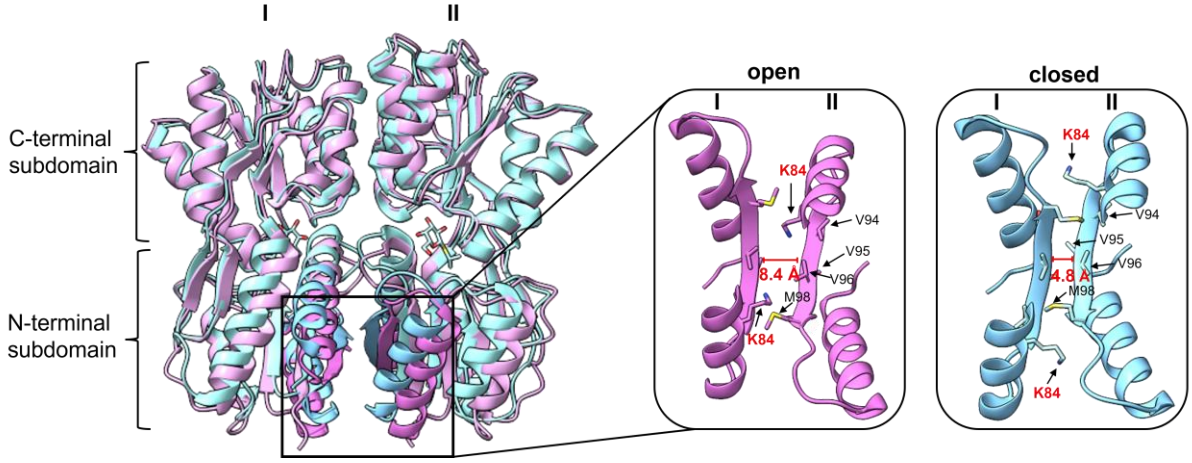


Figure 9. Structural comparison of LacI in the state without ligand and the inducer-bound state. The structures of the effector binding domain of LacI in the absence (pink, PDB:1JWL[95]) and presence of the inducer IPTG (blue, PDB:2P9H[30]) are superimposed on the upper lobe, highlighting conformational differences within the N-terminal subdomain of the effector domain. A close-up view emphasises interactions that stabilise the IPTG-bound closed conformation. L84 becomes more solvent-exposed in the IPTG-bound state, while IPTG binding reduces the distance between two opposing β-strand edges, allowing residues 94-98 to stabilise a continuous β-sheet.

The inducer-dependent rigidification of the N-terminal subdomain constitutes a central element of allosteric regulation in LacI, as it alters the structural coupling to the DNA-binding domain. Reducing the stabilising interactions between the effector domain and the hinge helix in the inducer-bound state lowers the stability of the structured hinge-helix, required for high-affinity operator binding, as the helical conformation of the hinge helix contributes to minor-groove interactions as well as to connections between the DNA-binding domain and the ligand-binding

domain.[56,95] In this way, ligand binding in the effector pocket is translated into reduced DNA affinity at the operator site.

The availability of residue-resolved structural models and defined allosteric communication nodes has enabled the extensive re-engineering of the LacI switch into anti-repressors. This has expanded the regulatory repertoire for mechanistic studies and synthetic biology, particularly in the context of synthetic genetic circuits.[103–105]

1.1.1.4 Dimerisation of LacI mediated by the PBP-like core domain

Both the N- and C-terminal subdomains of the PBP-like domain contribute to stabilising the monomer-monomer interface of LacI.[106] However, the N-terminal subdomain appears to play a more important role in enabling a functional allosteric response rather than the stabilisation of the dimer interface (see Chapter 1.1.1.3).[107] Mutations at the N-terminal dimer interface illustrate a key design principle: increased dimer stability is achieved at the expense of allosteric performance.[100,102,108] These observations suggest that the N-terminal interface is not solely optimised for maximising dimerisation stability, but instead strikes a balance between stability and the conformational flexibility necessary for regulation. In contrast, for the C-terminal subdomain, key interaction partners mandatory for dimerisation are identified. Structurally, the C-terminal portion of the monomer-monomer interface is formed by three α-helices (residues 222–235, 247–260, and 276–281)[30] (Figure 10, highlighted by darker colours).

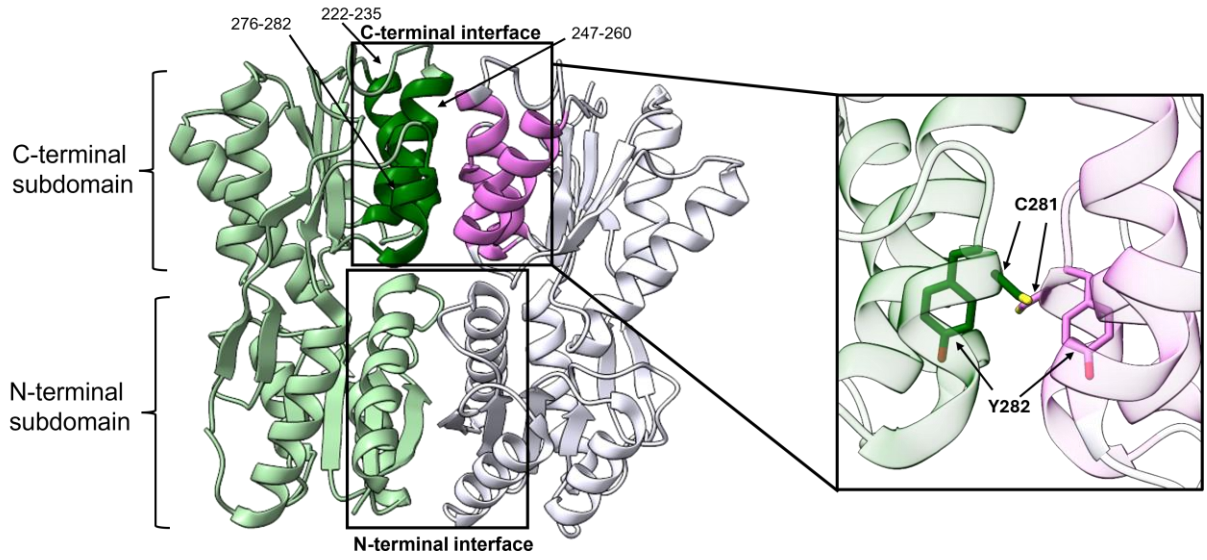


Figure 10. Structure of dimeric LacI lacking the DNA binding domain (PDB: 2P9H[30]), highlighting the C-terminal dimerisation interface. The three α-helices forming the monomer-monomer interface are highlighted in darker colours and are central to stabilisation of the dimer. A close-up of interface residues 281 and 282 is shown on the right.

For repressors belonging to the LacI/GalR protein family, dimerisation is stabilised by interactions at the monomer-monomer interface. Still, residues found at this interface are conserved only at the structural level and exhibit low sequence similarity compared to

members of the LacI/GalR protein family.[106] Despite sequence divergence among LacI/GalR family members, the inter-monomer contact pattern is highly conserved. Relative to the conserved interaction network characteristic of this family, TreR shows only three divergent contacts and LacI only five divergent contacts (Figure 11, thin lines). In contrast, the remaining interfacial interactions are retained in all family members (Figure 11, bold lines) or in some family members (Figure 11, bold dashed lines).

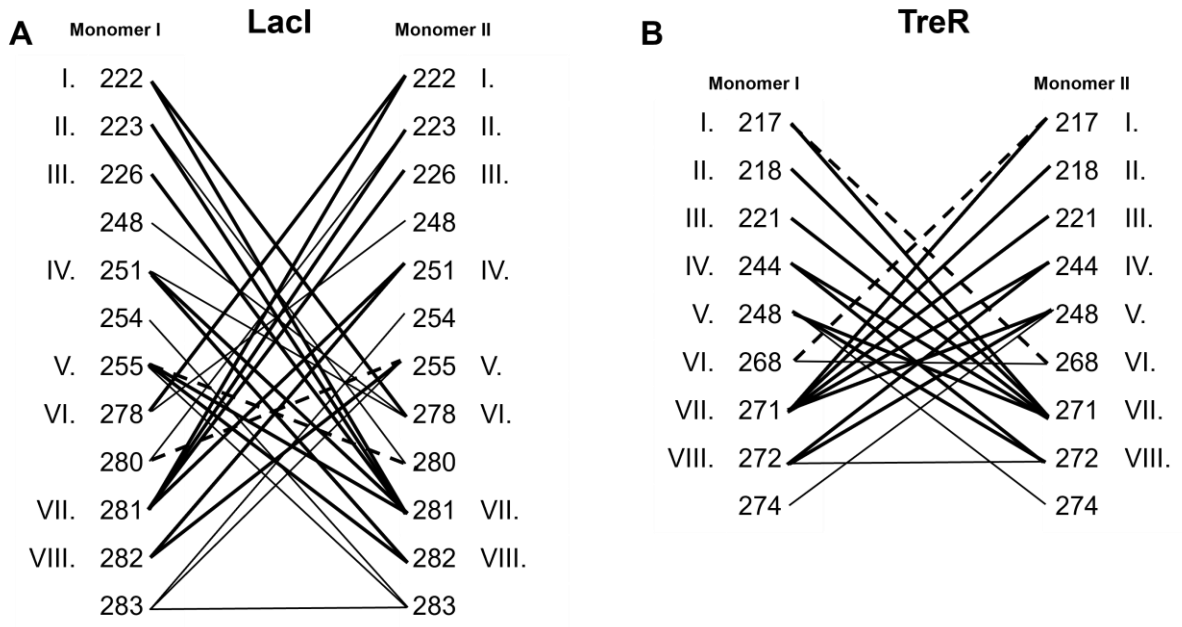


Figure 11. Interaction patterns between monomer I and II of LacI and TreR. Interactions at the monomer–monomer interface are shown for A) LacI and B) TreR. Roman numerals indicate structurally conserved positions within the interface. Line thickness distinguishes conserved from repressor-specific contacts: bold lines indicate interactions also observed in other LacI/GalR family members, whereas thin lines denote interactions unique to the respective repressor. Line style indicates the degree of conservation across the family: solid lines represent interactions present in all family members, and dotted lines represent interactions observed only in a few. Adapted from Swint-Kruse *et al*.[106].

The small deviations in the interface interactions shows that members of the LacI/GalR protein family can exhibit alternative yet functionally compatible interaction solutions consistent with pronounced interface plasticity studied for LacI.[106] Plasticity within the monomer-monomer interface is particularly well documented for LacI, where the deleterious effects of mutations at position 282 (VIII) are compensated for by local rearrangements in the vicinity of position 282, stabilising the monomer-monomer interface.[106] Additionally, several mutations have been described, including substitutions on the backside of the inducer-binding site (e.g., A133T/V, V150I, S151P) and within the binding pocket (e.g., D149N, S191F), which have similar stabilising effects.[109] Some mutations, especially outside the C-terminal subdomain, restore dimerisation but also impair regulatory function.[35,106,109]

Despite the low sequence similarity between the LacI/GalR family in the dimerisation interface, four positions within the C-terminal interface exhibit a conserved hydrophobicity (Table 1, grey). Besides focusing on members of the LacI/GalR family, it is worth investigating the conservation

of these positions between monomeric periplasmic binding. The conserved hydrophobicity of the LacI/GalR family is only retained at two positions of monomeric periplasmic binding proteins (Table 1, orange). Amino acids at positions VII and VIII do not retain hydrophobic character, indicating that they are crucial for dimerisation.

Table 1. Sequence conservation between selected repressors and monomeric periplasmic binding proteins at structurally conserved positions within the C-terminal monomer-monomer interface. Grey shading marks positions conserved as hydrophobic among repressors, whereas orange shading indicates positions where hydrophobic character is also retained in the monomeric periplasmic binding proteins. Adapted from Swint Kruse *et al.*.[106]

	I	II	III	IV	V	VI	VII	VIII
LacI	A	M	F	L	R	D	C	Y
PurR	P	E	Y	M	C	N	Y	F
TreR	M	K	Y	L	K	L	F	L
RbsR	F	N	F	V	Q	L	F	A
RbsB	R	I	L	L	R	D	K	A
ABP	I	P	F	L	R	-	S	E
ALBP	R	I	L	M	Q	E	K	M

Early mutational studies further highlighted the importance of positions VII and VIII for the oligomeric state of LacI.[35,109,110] Position 281 (VII) exhibits substantial tolerance for hydrophobic, aromatic, and even charged substitutions, whereas substitutions such as P, K, or G promote monomer formation.[109,110] Position 282 (VIII) is considerably more restrictive; only F or L exchanges are tolerated, while most other substitutions shift the equilibrium toward monomers rather than dimers.[109,110] Mutations that retain the dimer interface are found in other natural repressors that belong to the same protein family.[106] This asymmetric sensitivity of residue C281 and Y282 is consistent with the structural context of these residues. While residue Y282 is located near the centre of the interaction network, residue C281 occupies a more peripheral position (Figure 10, right side).[30,67]

Mechanistic insights into lethal dimerisation mutations and interface plasticity have been applied to engineer LacI variants with rearranged monomer-monomer interfaces.[68,111] Biasing the formation of heterodimers over homodimers through the introduction of complementary, specific inter-subunit interactions enabled the formation of heterodimers *in vivo*.[68]

1.1.2 TreR: A trehalose-responsive member of the LacI/GalR Protein family

Compared with LacI, the dimeric trehalose repressor (TreR), another member of the LacI/GalR protein family, has received comparatively little attention. It was first described in 1995 as the transcriptional regulator of the transmembrane phosphotransferase TreB and the cytoplasmic trehalose-6-phosphate hydrolase TreC.[112] Structurally, TreR conforms to the conserved modular architecture that defines the LacI/GalR family, comprising an N-terminal helix-turn-helix (HTH) as the DNA-binding domain and a C-terminal PBP-like domain as the effector

domain.[33] TreR recognises palindromic operator sequences, consistent with binding of the repressor as a dimer.[113,114] As in other members of the LacI/GalR protein family, the effector molecule binds within a cleft formed between the N- and C-subdomains of the regulatory domain (Figure 12), triggering conformational rearrangements that modulate operator affinity.[115] In TreR, the natural effector Trehalose-6-phosphate (T6P) acts as an inducer, while trehalose is a neutral ligand.[115] A key mechanistic insight into the ligand binding and allosteric response of TreR emerged from the crystal structure of its effector-binding domain in complex with T6P, which revealed a dual recognition of the sugar and its phosphate moiety. The trehalose scaffold, especially the hydroxyl groups of the sugar moiety are stabilised by an extensive hydrogen-bonding network involving residues such as E77, R147, D187, T189, and T190 (Figure 12, orange). In contrast, the phosphate group is coordinated by a dense cluster of polar and positively charged residues, including R71, S74, S76, T240, T242, and Y284 (Figure 12, green), forming an electrostatically favourable environment for interactions with the phosphate group that is further stabilised by water-mediated interactions and has functional significance for the allosteric regulation.[115]

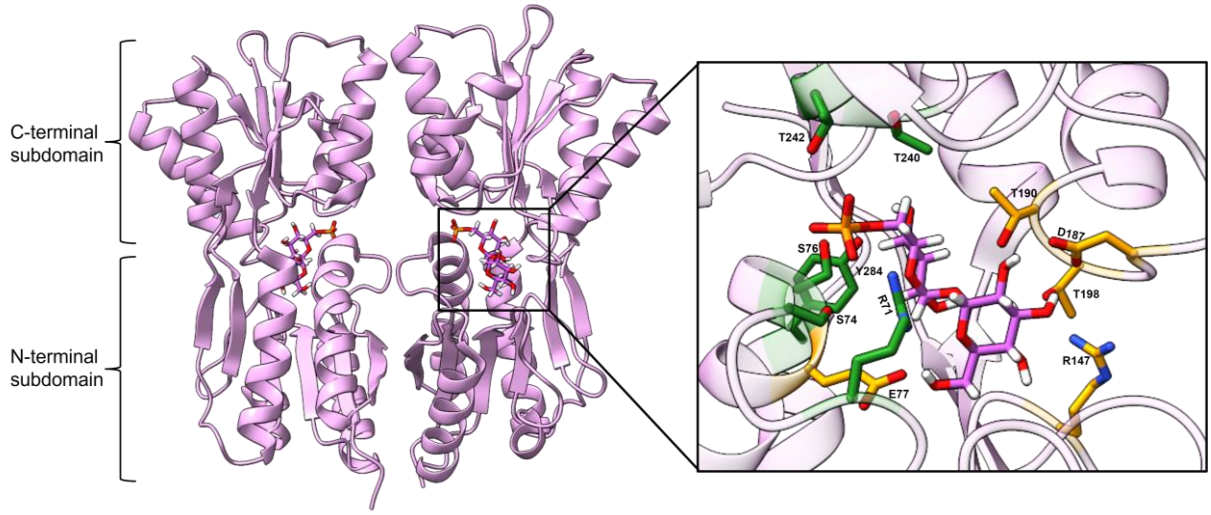


Figure 12. Crystal structure of the TreR effector-binding domain in complex with T6P (PDB:4XXH[115]). The ligand-binding site is shown in close-up, with T6P depicted as a stick model and coloured pink. Residues interacting with the hydroxyl groups of the trehalose moiety are highlighted in orange, whereas residues contacting the phosphate group are shown in green.

Exchanging T6P with its non-phosphorylated analogue trehalose shows that while sugar binding was not impaired, the absence of the phosphate group abolishes the allosteric regulation of DNA binding.[115] No controlled release from the DNA can be detected, highlighting that the phosphate moiety is important to couple ligand recognition to conformational rearrangements within the DNA binding domain.[115]

1.2 The periplasmic binding protein

As part of the ABC transporter, periplasmic binding proteins (PBPs) play an important role in the active transport of water-soluble ligands.[116,117] The PBPs are divided into three different classes, depending on the number of crossovers between the N- and C-terminal subdomains. Class I has three crossovers, while class II has two and class III connects the C- and N-subdomains only via one crossover.[118] A difference between those classes is the motion observed after ligand binding: while large conformational changes are observed for class I and II, class III hardly shows any change in the overall architecture.[119] Members of class I PBPs share the same structural architecture as the LacI/GalR protein family's effector domain (see Chapter 1.1).[28,29] Class II PBPs consist of two αβα lobes. However, compared to their precursor class I, they only have five β-strands and two crossovers between the N- and C-terminal subdomains.[27] Upon ligand binding, these proteins undergo a pronounced twisting and closing motion, while the small molecule binds in the hinge region between the two β/α lobes.[120]

1.2.1 PotF: a suitable natural clapping machine for motor design

PotF, a member of the class II family, has already been utilised as a starting construct for protein design challenges.[121,122] Both the N- and C-terminal subdomains are composed of a five-stranded central β-sheet. However, the N-terminal subdomain contains six additional α-helices, while the C-terminal subdomain contains eight (Figure 13).[123] Two antiparallel β-strands link the subdomains.[123] Different polyamines bind in the hinge region of PotF. However, their affinities and induced conformational change vary.[122] After ligand interaction of the natural ligand putrecine, the opening angle and a twisting angle change by 40° or 50°, respectively. Thereby, the distance between the C- and N-termini changes by 12 Å (Figure 13).

The conformational changes of protein domains are crucial for functions such as catalysis, regulation, and the transport of metabolites, and also play a major role in the design of both autonomous and non-autonomous motor proteins (see Chapter 2.1.2). Conformational changes occurring after ligand binding vary across protein families and range from small to large motions. These include motions such as shear, the reorientation of secondary structure elements, and hinge motions, which involve the reorientation of a whole domain or segment by altering the torsion angle of a small hinge element. PotF and other members of the PBP family are classified as domain-based hinge motion by performing a conformational change upon binding a ligand.[27] Hinge motion seems to be the most suitable type of motion for a designed motor protein 'clapping machine' for artificial motor protein design since shear motion is usually insufficient to produce large-scale domain motion (see Chapter 2.1.2). [1,124]

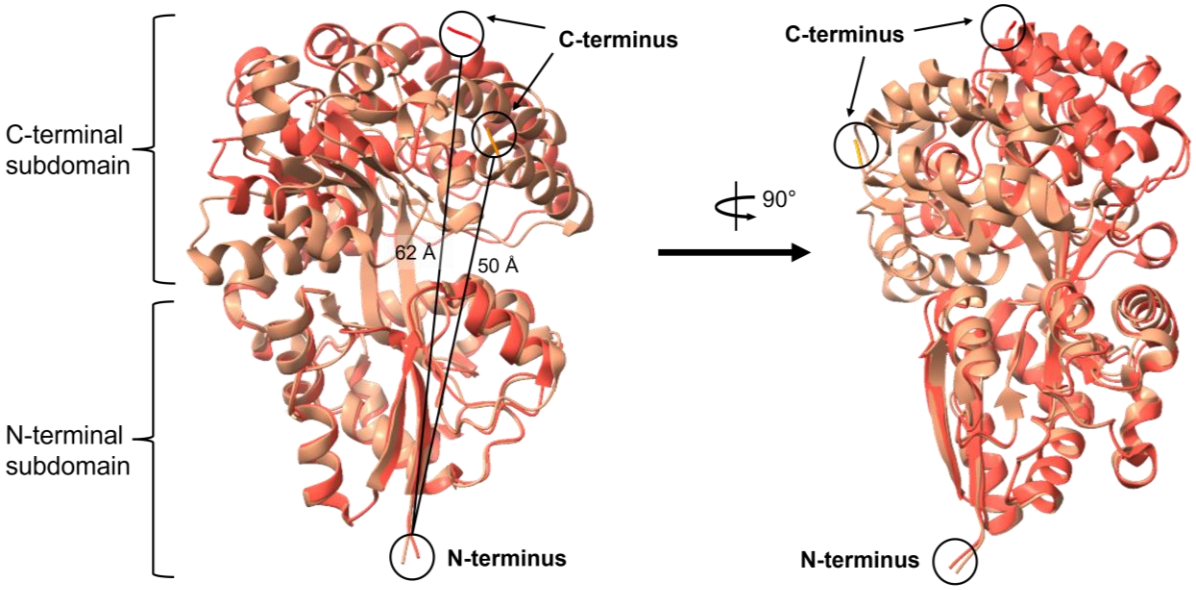


Figure 13. Structural comparison of PotF in the closed and open conformation. The closed PotF structure (orange, PDB: 6YEB[122]) and open structure (red, PDB:6YED[122]) are superimposed based on the lower lobe and shown in two orientations. The N- to C-terminal distance is indicated to highlight the structural rearrangement between both states.

Server key criteria need to be considered to identify the optimal scaffold for a clapping machine. Important parameters are the magnitude of the conformational change in the presence and absence of ligands, the possibility to tune the kinetics of ligand binding, a monomeric architecture and the possibility to incorporate an enzymatic reaction at later design steps of the autonomous motor. Comparing PotF to other protein classes revealed that the change in the distance between N- and C-terminus is sometimes greater than what is observed for PotF. Calmodulin, a member of the EF-hand calcium-binding protein family, for example, undergoes a conformational change of 26 Å between its apo and Ca^{2+}-bound forms.[125] Even though the conformational change would be ideal for use in the context of the motor protein, Ca^{2+} is not a suitable ligand to design enzymatic function. The peptide-bound form of Calmodulin, however, only exhibits a 12 Å change.[125] Furthermore, the dissociation kinetics of the peptides are high, which also makes them unsuitable for use in a ligand-controlled clapping machine, where fast dissociation times are crucial for the rapid stepping of a non-autonomous motor.[1,126] In comparison, the ligand-binding site of PotF is well-characterised, enabling the design of specific affinities and kinetics.[122] In comparison, ATP sulfurylases with a change of 9 Å are around 30 kDa larger than PotF in one subunit.[127,128] ATP sulfurylases are a hexamer in nature, which makes them even less suitable as a component for a defined bipedal motor protein *in vitro*.[127,128] The monomeric character of PotF, its ligand-binding tunability, and the hinge motion upon ligand binding make it an ideal candidate for use as a clapping machine in artificial motor proteins (Chapter 2.6).

1.3 Strategies for functional protein assemblies

1.3.1 Principles and applications of split-protein complementation

Splitting proteins has been widely used both to control functions and to interrogate biological processes *in vivo*.[129] Conceptually, split-protein platforms are usually implemented as two- or three-part systems, where the individual fragments are inactive on their own, but can regain native activity through complementation.[130] These complementations can be controlled, meaning that assembly occurs only with an external trigger, or occur spontaneously, meaning that reassembly occurs without an external trigger (Figure 14).[129,131] In conditionally activatable designs, split fragments are fused to regulatory domains that only bring the fragments into proximity under defined stimuli. Light-oxygen-voltage (LOV) domains, for example, undergo light-induced conformational changes[132] and have been used to control Cre recombinase[133] or RNA polymerase activity[134,135]. Similarly, ligand- or temperature-controlled reassembly modules, such as the FK506-binding protein and FKBP-rapamycin binding domain of mTOR (FKBP-FRB)[136] as well as the temperature-sensitive TlpA protein[137], have been used to regulate split systems such as Cas12 protein[138], ubiquitin-based reporters[139] and *Tobacco Etch Virus* (TEV) protease[140]. The objective of these approaches is to regulate gene expression, signal pathways and protein stability in response to external stimuli.[129,133–135,138–140].

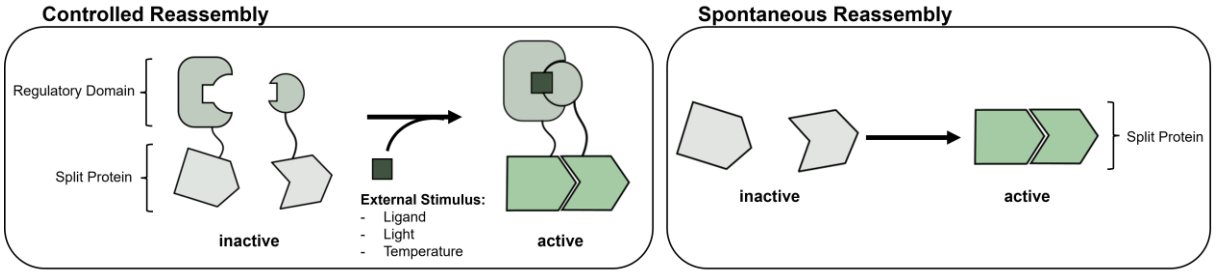


Figure 14. Stimulus-controlled and spontaneous assembly pathway of split-protein systems. The scheme summarises different modes of split-protein reconstitution, including externally triggered assembly and spontaneous reassembly of the protein fragments. Modified from Bae *et al*.[129].

The ability to self-assemble spontaneously allows split-protein systems to regain functionality without needing an outside trigger. SplitGFP is a prime example of this, having been derived from the green fluorescent protein (GFP), one of the most widely used fluorescent reporters for cellular imaging[141,142], protein localisation studies[143,144] and gene expression reporting[141,142]. GFP adopts an 11-stranded β-barrel architecture that encloses a central α-helix (Figure 15A).[143] The chromophore is embedded within this helix and is largely shielded from the surrounding solvent by the barrel structure, which supports high quantum yield and photostability.[144] Fluorescence only emerges after chromophore maturation, a post-translational process that converts three internal residues (Ser65, Tyr66 and Gly67) into a conjugated chromophore.[144] Since its discovery in *Aequorea victoria* in 1962,[145] GFP has been extensively optimised through mutational screening to improve its spectroscopic properties

and folding behaviour.[146–148] *A. victoria* thrives in relatively cold waters (9-12 °C). Early engineering efforts to enhance folding robustness at higher temperatures led to the development of enhanced green fluorescent protein (eGFP)[146] and superfolder green fluorescent protein (sfGFP)[147], both of which exhibit increased folding stability.

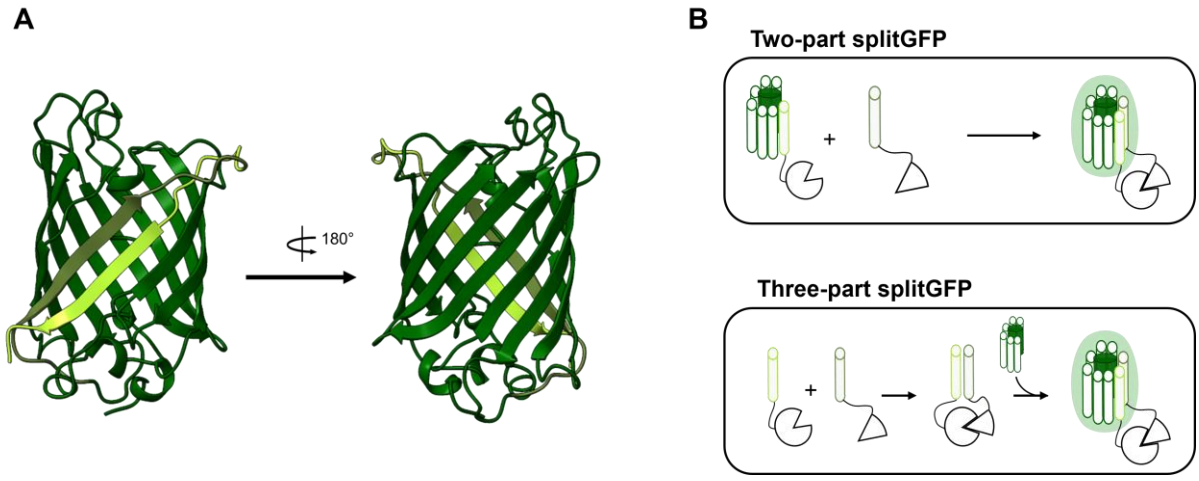


Figure 15. Overview of GFP and splitGFP complementation strategies. A) Crystal structure of *Aequorea victoria* GFP (PDB: 1EMA[143]). B) Schematic representation of two-part and three-part splitGFP complementation strategies.

sfGFP has been repurposed into different splitGFP systems. Three-part systems, typically consisting of GFP1-9 (Figure 15, green) plus separate β-strands 10 and 11 (Figure 15, khaki and light green), have been successfully implemented for read-outs of enzymatic activity in living cells[149] or to analyse protein-protein interactions *in vitro*[150]. Two-part designs are divided into a large fragment GFP1-10 (Figure 15, green and light green) and a small fragment comprising strand 11 (Figure 15, khaki).[151] The fragments complement each other independently with a high level of efficiency[152], which is why splitGFP was repurposed as an aggregation sensor when its fragments are fused to aggregation-prone proteins.[153–155]

Furthermore, split fluorescent proteins have been widely used to analyse protein-protein interactions (PPIs) both *in vivo*[156–159] and *in vitro*[160–162] Assays analysing PPIs based on split fluorescent reporters are commonly referred to as bimolecular fluorescence complementation (BiFC).[163] Those assays do not build on the spontaneous self-assembly of split-GFP, but rather on the concept of enhanced assembly by attaching proteins of interest that form PPIs. A general limitation of split fluorescent proteins in the use of BiFC assays is their tendency to undergo concentration-dependent spontaneous complementation, as demonstrated for splitGFP, whose self-assembly has been reported to be approximately 3-5 fold slower than diffusion-limited association.[152] Even though this can result in basal reconstitution in the absence of a specific interaction, split proteins can still be used for protein-protein interaction (PPI) analysis in living cells, because genuine PPIs substantially increase complementation abundance relative to random encounters.[156,157] This is overcome by engineering split fluorescent protein variants with reduced intrinsic self-association. Reducing basal

complementation and preserving interaction-driven reconstitution have been achieved through site-directed mutagenesis[164,165] or spacer-insertion screening[166]. Besides using split fluorescent proteins, PPIs have been analysed in a cellular context using split reporters that are luminescent[167,168] or that exhibit enzymatic outputs (e.g., HRP[169,170], β-lactamase[171], or aminoacyl-tRNA synthetases[172]).

In general, PPIs can be defined as specific associations between proteins in a cellular context that depend on local concentration and subcellular localisation and occur either directly, via physical contact between two proteins, or indirectly, through intermediates within a larger protein complex.[173] PPIs are central determinants of cellular physiology and have been implicated in numerous diseases[174,175], as they regulate a wide range of biological processes, including transcription[176], replication[177], signal transduction[178], membrane transport[179,180], and cell-cell communication[181].

The LacI monomer-monomer interface represents homomeric PPI that stabilises the functional dimer (see Chapter 1.1.1.4). In this work, BiFC-derived approaches are repurposed to identify interactions at the monomer-monomer interface that mediate heterodimer formation. Moreover, the intrinsic self-assembly of split GFP[152] is used in the splitGFP-guided complementation to achieve the formation of heterodimers from monomeric constructs (Chapter 2.2.2), thereby restoring the LacI DNA-binding activity in split constructs (Chapter 2.2.3).

1.3.2 Covalent coupling strategies for modular protein assembly

The design principle of modularity is pervasive. It enables the separation of discrete functional units and the generation of highly complex systems through the combinatorial assembly of protein parts.[182] A range of site-specific conjugation strategies have been discovered. The two most prominent categories of conjugation are split protein domain-catalysed conjugations and chemical ligation methods.[183] For native chemical ligation that yields an amide bond, the practical challenges are the generation of the required reactive C-terminal thioester, the low biocompatibility associated with this process, and selectivity issues in complex biological environments.[184]

Other alternatives, such as several enzymatic and protein-domain-catalysed conjugation systems, are also not without limitations. These systems vary in the segments of the protein they fuse to and in the mechanism they use to catalyse bond formation.[185] Sortase-mediated ligation catalysing peptide-bound formation between a C-terminal recognition motif (LPXTG) and an N-terminal glycine residue exhibits modest reaction efficiency and calcium dependence.[186,187] Another widely used strategy, involving transglutaminases to generate isopeptide bonds between lysine and glutamine side chains, is highly biocompatible but has limited site selectivity.[188–191]

Limitations in biocompatibility, efficiency, and selectivity have been addressed by the

development of Tag/Catcher systems, which were pioneered by the SpyTag/SpyCatcher system.[192] It originates from the CnaB-like domain of the pilin protein Spy0128 from *Streptococcus pyogenes*, which naturally forms an irreversible intramolecular isopeptide bond (Figure 16A). The first engineered two-component system comprises a short peptide (SpyTag) and a complementary protein partner (SpyCatcher), which reconstitute the reactive interface and spontaneously covalently couple via an isopeptide bond between a lysine side chain and an asparagine. This process is catalysed by a conserved glutamic acid (Figure 17).[192]

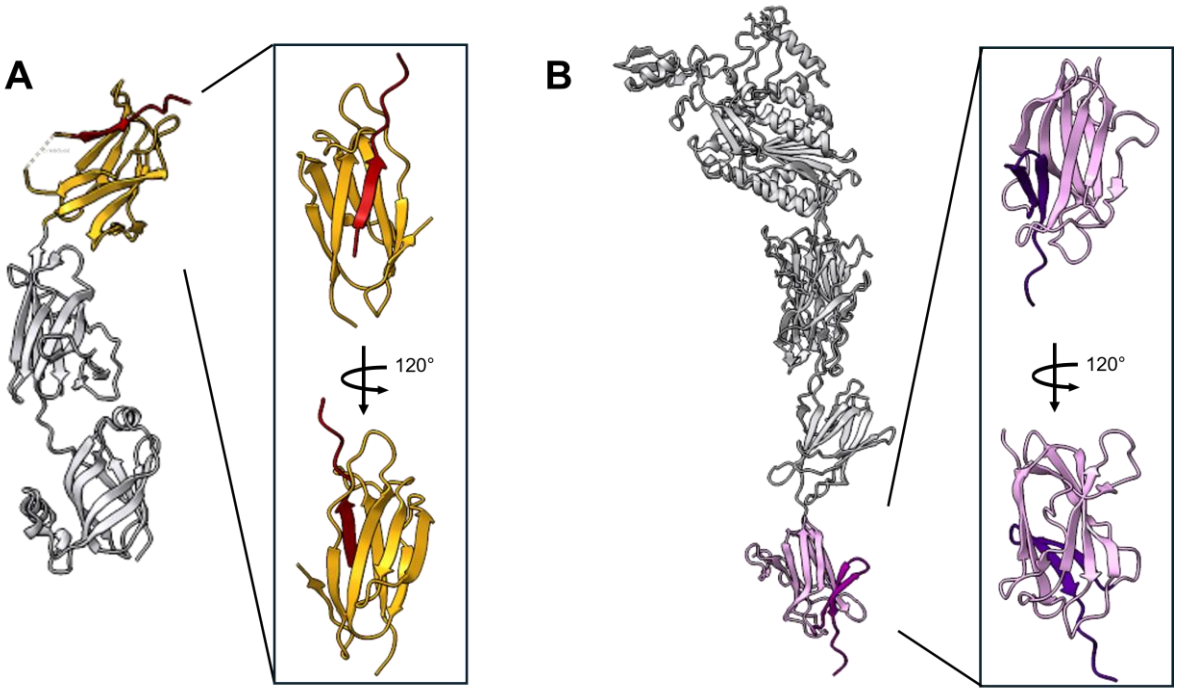


Figure 16. Overview of SpyTag/Catcher and DogTag/Catcher systems and their natural origins, CnaB and RrgA. A) Structure of the CnaB domain from the *Streptococcus pyogenes* ancillary pilin (PDB:4OQ1[193]), highlighting the natural domain that was repurposed to engineer SpyTag (red) and SpyCatcher (orange) (PDB:2MLI[194]). A boxed close-up shows the SpyTag/SpyCatcher complex in two different orientations. B) Structure of RrgA from *Streptococcus pneumoniae* (PDB:2WW8[195]), highlighting the natural domain that was repurposed to engineer DogTag (dark purple) and DogCatcher (pink). A boxed close-up shows the DogTag/DogCatcher complex in two different orientations.

Since the initial development of these systems, numerous Tag/Catcher pairs have been engineered. The main differences between these systems lie in the source protein scaffold and the position of the split site. Conceptually, split designs are grouped into[183,196]: (i) systems, in which a small C-terminal tag is generated by splitting after the first β-strand[192]; (ii) systems, which split after the first two N-terminal β-strands. This class of split Tag/Catcher proteins are ideal candidates for loop insertion and internal display[197,198] like DogTag/Catcher which originates from the protein RrgA from *Streptococcus pneumoniae* (Figure 16B); and (iii) multi-fragment designs, in which the domain is partitioned into three components: two Tag segments which carry the residues that form the isopeptide bond, and the Catcher provides the complementary reactive environment and the catalytic glutamic acid residue that promotes

bond formation.[199,200] The Tag/Catcher conjugation process is efficient over a broad range of conditions, including temperature ranges between 4 and 37 °C, diverse buffer compositions, and pH ranges.[183,192] It is also compatible with *in vitro* and *in vivo* applications.[183,192]

Mechanistic insight into Tag/Catcher reactivity has been obtained through a combination of NMR spectroscopy[201], X-ray crystallography[202], and QM/MM calculations[203], which together support a conserved catalytic reaction (Figure 17). The key reactive residues, namely lysine, asparagine/aspartate and glutamic acid, are situated within a predominantly hydrophobic microenvironment, with strict geometric conservation across related systems, and are frequently embedded within an aromatic cluster that stabilises the reactive configuration. A nucleophilic attack by the ε-amino group of Lys on the C_γ of Asn/Asp forms a zwitterionic intermediate state (Figure 17, step 1). Using Glu as a proton shuttle, a neutral tetrahedral intermediate is formed (Figure 17, step 2). The release of ammonia or water results in the intermediate state breaking down into the final, stable, irreversible isopeptide bond (Figure 17, step 3).[183,201,203,204]

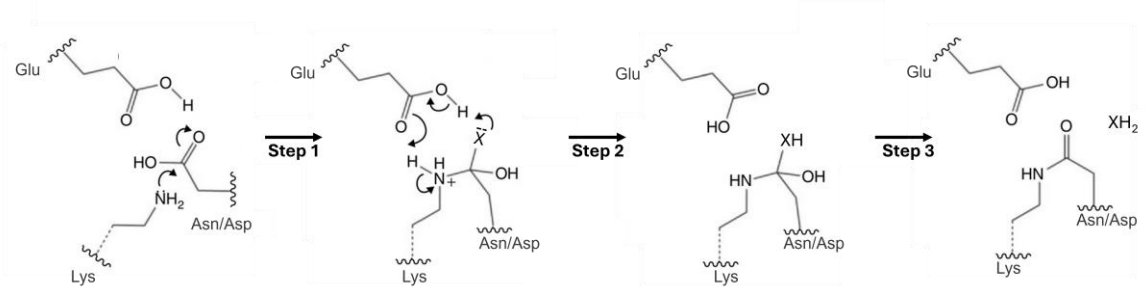


Figure 17. Proposed mechanism of the isopeptide bond formation. The scheme highlights nucleophilic attack by the ε-amino group of Lys (step 1), proton shuttling (step 2), and formation of the stable isopeptide bond with concomitant release of water or ammonia (step 3). Modified mechanism from Reddington *et al.*[204].

The spontaneous, efficient, and stable nature of the conjugation reaction, along with its robustness under a variety of reaction conditions, makes the Tag/Catcher system an ideal choice for a variety of applications. Tag/Catcher systems have been used for hydrogel formation[205,206], immobilisation on nanoparticles[153] or quantum dots[207,208], affinity purification tags[209], cellular applications such as the specific labelling of living cells[192,210], polymerisation[196,211–213], and many more. In the context of artificial motor protein assembly, the SpyTag/Catcher system has already been used,[214] and is also used in the underlying study as a connection tool to restore monomer-monomer interactions as well as a tool to connect different building blocks.

1.4 The evolution of protein design

Since the early 1970s, structure prediction and protein design have undergone massive changes, especially in terms of speed of success.[215,216] The achievements have been highlighted by the 2024 Nobel Prize being awarded to the three scientists David Baker, Demis Hassabis and John Jumper, who have had a significant impact towards more successful protein design.[217,218] Starting with the idea of understanding the sequence-to-structure relationship, protein design has evolved to encompass the design of entirely new proteins with existing or novel folds as well as optimised or new functions.[215,216] Before outlining the key milestones and modern methods, it is helpful to distinguish between two broad design paradigms: protein engineering[219] and *de novo* protein design[216] (Figure 18).

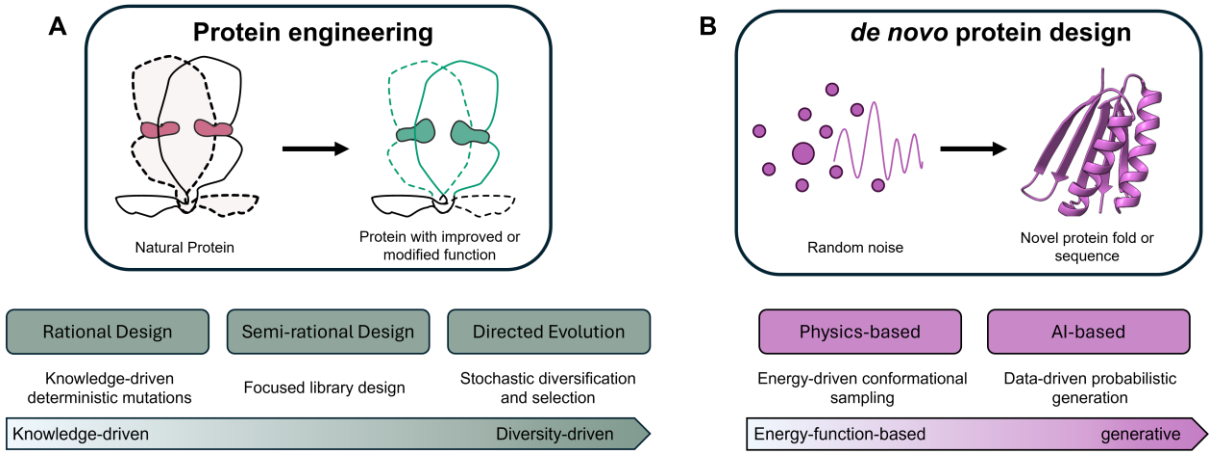


Figure 18. Overview of protein design strategies. A) Protein engineering strategies and methodological approaches are classified along a spectrum from knowledge-driven to diversity-driven methods. B) *De novo* protein design strategies are grouped into two broad classes of design principles, ranging from energy-function-based to generative approaches.

Protein engineering starts with natural proteins and modifies them to improve, alter or repurpose their properties, such as stability, binding or activity. This is commonly achieved through rational design, directed evolution or semi-rational strategies (Figure 18A).[219] By contrast, *de novo* protein design does not typically begin with a natural scaffold. Instead, the scaffold itself was the initial design target (Figure 18B), and functional features were introduced during a later design step.[220]

By this classification, the reconstruction of the repressor monomer-monomer interface shown in this study is classified as protein engineering, as are the modifications of the repressors LacI and TreR, and the periplasmic binding protein PotF, to create or extend their functions for use as building blocks in engineered motor protein systems. However, the design steps involved in constructing novel structural elements within these repressors extend beyond classical protein engineering workflows, incorporating modern computational and AI-based methods mostly used for *de novo* applications. Consequently, parts of this study fall into a conceptual grey zone between protein engineering and *de novo* design.

The vast combinatorial size of the sequence space of 20^n sequences for a polypeptide of length

n and the comparatively small free-energy gap between folded and unfolded states suggested that designing proteins from scratch would be impractical.[215,221–223] Early efforts to create proteins with a predetermined structure and function before the 1980s largely failed to yield robust designs that folded reproducibly.[224,225] Since then, an increased understanding of protein folding and rapidly improving computational methods have transformed *de novo* protein design into a mature discipline capable of producing proteins that are increasingly functional and structurally sophisticated.

Initial successes were driven by the application of physicochemical principles of secondary-structure formation and the use of mathematical scoring functions to specify backbone geometry and side-chain repacking. Seeking out low-energy sequences led to the design of the first small proteins designed using a *de novo* approach.[220,226]

A major increase in success between 2000 and 2015 was driven by physics-based approaches that used all-atom energy functions to evaluate the energy of a given design.[227,228] Rosetta tools perform computational backbone generation through coarse-grained blueprints of the target fold, utilising fragment-based assembly from PDB-derived structural elements.[229,230] This process is coupled with iterative sequence design and structural relaxation to minimise free energy, and the all-atom energy calculations have led to the successful design of proteins with high thermodynamic stability.[229–232] (Figure 19A) Over subsequent years, continued improvements in energy functions, sampling strategies, and the integration of experimental feedback have further increased accuracy, enabling the design of elaborate architectures[233–236] and proteins with versatile, tailored functions[237–240]. In addition to Rosetta, other structure- and energy-based design frameworks further supported the development of computational protein design by providing specialised strategies for sequence optimisation within predefined structural contexts.

OSPREY, for example, combines discrete rotamer sampling with energy-based scoring and provable search algorithms to identify low-energy sequence-conformation combinations (Figure 19A).[241] A particular strength of OSPREY is the explicit consideration of conformational flexibility and conformational ensembles, which allows mutations to be evaluated not only as static side-chain replacements but in the context of alternative low-energy states.[241] This enabled applications in protein redesign, ligand-binding optimisation, and protein-protein interface engineering, thereby contributing to the transition from purely rational mutagenesis towards more systematic computational exploration of sequence space.[242–244]

Another program, called Triad, has been used as a structure-guided protein engineering platform to evaluate and prioritise mutations based on their compatibility with a given protein scaffold or functional site.[245] It is best described as a structure-guided protein engineering platform for evaluating and prioritising mutations based on their compatibility with a given protein scaffold or functional site (Figure 19A).[245] The software applies rotamer optimisation

using a Monte Carlo with simulated annealing search algorithm, thereby sampling alternative side-chain conformations and identifying energetically favourable sequence–structure combinations.[245] Rather than designing completely new protein folds, such approaches are particularly useful for redesigning existing proteins, for example, by improving packing interactions, stabilising active or binding-site geometries, or modifying local structural environments.[245–247]

However, deep learning methods that generate sequences, structures or both in an integrated manner have been a major step towards solving the protein design problem.[248,249] Deep learning-based protein design methods can be broadly grouped by the type of information used as input and the type of output they generate (Figure 19B).

Protein language models adapt architectures and training principles originally developed for natural language processing to protein sequences. By learning statistical dependencies and evolutionarily shaped patterns from large collections of natural protein sequences, these models enable *de novo* protein sequence generation without requiring an explicit structural backbone as input, as exemplified by ProtGPT2[250,251]. (Figure 19B)

In the following work, sequence generation tools with fixed backbones are more important, as the backbone plays a key role in the design. Among the wide range of neural network-based techniques, ProteinMPNN[252] has become the go-to tool for fixed-backbone sequence design. ProteinMPNN[252] applies a message passing to predict amino acid identities for a defined local geometric context. The geometric context is defined as the graph of proximal residues, where nodes represent residues and edges represent connections between neighbouring residues. The system is conditioned by a three-dimensional backbone (Figure 19B). It is a message-passing neural network (MPNN) comprising three encoder and three decoder layers with 128 hidden dimensions. It generates amino acid identities autoregressively, conditioned on the geometric features of the backbone. The inputs include inter-residue distances (e.g. Cα-Cα), relative orientations/frames and backbone torsion angles. This allows the model to output a probability distribution over the 20 amino acids at each position and to iteratively decode a sequence that is expected to be compatible with the target fold and soluble expression.[252] (Figure 19B)

Another key challenge that has been revolutionised by rapidly evolving AI-based techniques is structure prediction. Success increased by utilising AI-based methods for structure prediction and subsequently adapting them for design.[253,254] A notable example is the structure prediction tool RoseTTAFold[253], which was expanded to encompass more than just fixed-sequence structure prediction through additional training on fixed-backbone sequence design and inpainting tasks. RoseTTAFold is a three-track neural network, which integrates sequence information, pairwise residue relationships, and three-dimensional structural representations within a unified framework, allowing these levels of information to be refined simultaneously

and iteratively.[253] In its classical prediction mode, amino acid sequence information serves as input, and the corresponding three-dimensional protein structure constitutes the output, whereas in fixed-backbone design mode, a structural framework can instead be provided as input to generate compatible sequences (Figure 19C).

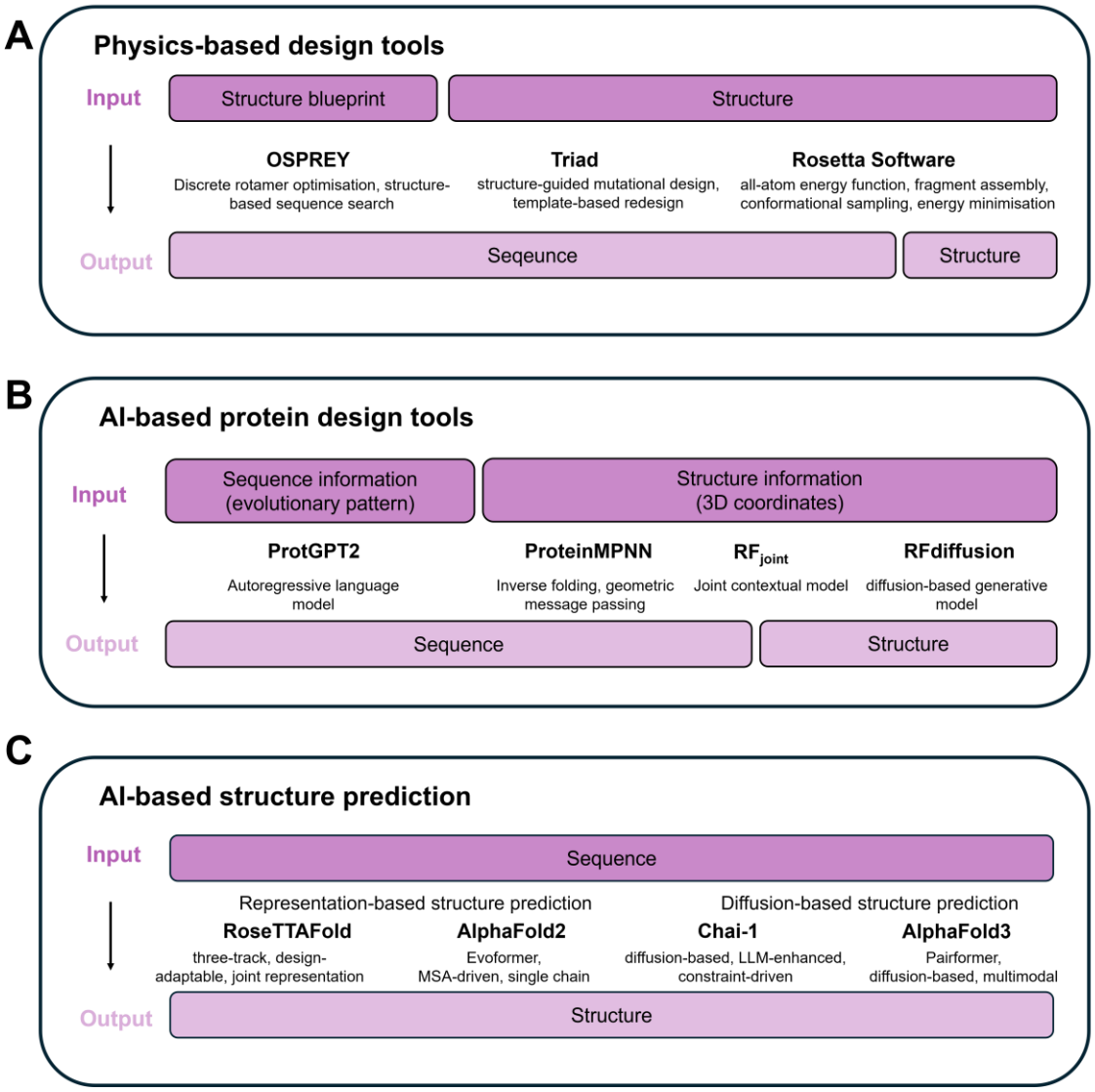


Figure 19. Conceptual comparison of input-output relationships in protein design and prediction methods. A) Physics- and energy-function-based protein design approaches, exemplified by Rosetta[229], OSPREY[241], and Triad[245], use predefined structural blueprints, scaffolds, or existing protein structures to identify energetically favourable sequence variants and compatible structural conformations. B) AI-based protein design approaches leverage sequence-derived or structural inputs through distinct generative frameworks, for example, present in ProtGPT2[255] for sequence generation, ProteinMPNN[252] for structure-conditioned sequence design, RF_{joint}[256] for joint sequence-structure modelling, and RFdiffusion[257] for *de novo* structure generation. C) Structure prediction methods use sequence information to predict the structure by using different classes of neural-network architectures, including representation-based predictors such as RoseTTAFold[253] and AlphaFold2[258] and generative or diffusion-based approaches such as Chai-1[259] and AlphaFold3[260].

Further fine-tuning the structure prediction tool RoseTTAfold[253] into a denoising network, called RF_{joint}[256], has been shown to help redesign the structure of TIM barrels[261], transmembrane nanopores[262] and many more[256,263]. RF_{joint}[256] extends RoseTTAFold[253] into a conditional denoising framework that jointly represents sequence and structure. RF_{joint} uses partial three-

dimensional coordinate information, optionally together with incomplete sequence information, as input and outputs reconstructed sequence-structure segments that remain compatible with the surrounding scaffold (Figure 19B).[256] RF_{joint} learned to iteratively recover coherent sequence-structure pairs from partially corrupted inputs, thereby enabling the context-conditioned inpainting of the missing segments. The model fills the missing protein segments in a structure-aware manner, making sure that the inpainted regions remain compatible with the fixed protein environment (Figure 20A).[256] Sequence generation is also possible with RF_{joint}[256], but inverse-folding models such as ProteinMPNN[252] outperform RF_{joint}'s sequence generation capabilities.[248,252,257] Designing an extension onto an existing repressor structure is an ideal showcase for RF_{joint}[256].

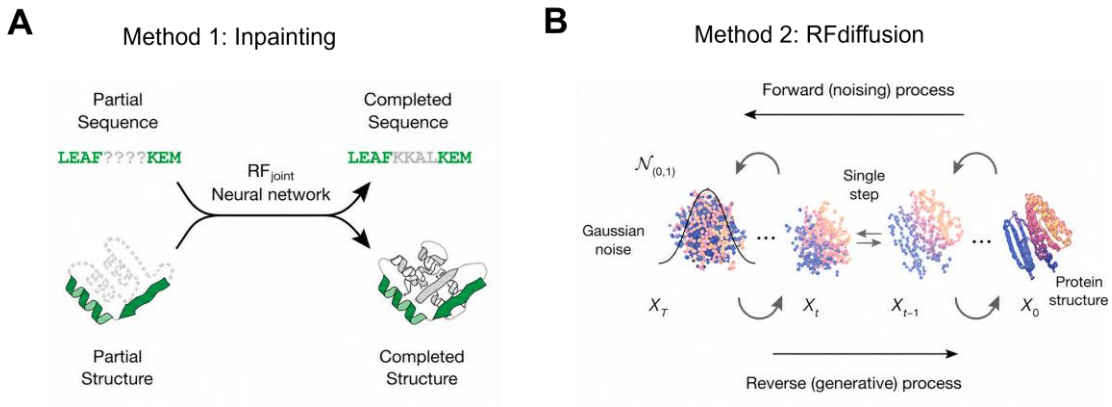


Figure 20. Comparison of the workflows of the structure generation approaches used in this work. A) In the RFjoint-based inpainting approach, known sequence and structural information (green) are provided as input, and the network completes the missing sequence and structural regions (grey). B) In RFdiffusion, novel structures are generated by iteratively reversing a corruption process through stepwise denoising. Self-conditioning is implemented by feeding the previous prediction back into the model as a template input for the next denoising step. Figure modified from Wang *et al.*[256] and Watson *et al.*[257]

A major step towards the successful design of diverse backbones and conditional designs, such as motif scaffolding or topological constraints, was achieved by fine-tuning RoseTTAFold[253] into a denoising diffusion probabilistic model (RFdiffusion).[257] RFdiffusion[257] learns to recover native-like structures from progressively corrupted inputs analogous to diffusion-based image generation. It operates directly on protein geometry by encoding three-dimensional backbone coordinates and their spatial relationships. Depending on the task, RFdiffusion uses either random noise or conditional inputs such as structural motifs, binding-site geometries, or topological constraints, and outputs complete three-dimensional protein backbone structures that satisfy these constraints. (Figure 19B) Through an iterative denoising process, the model refines noisy structural representations stepwise into coherent protein backbones, effectively sampling from a learned distribution over protein structures (Figure 20B). It has been widely adopted for tasks such as *de novo* binder design[264,265] and functional-site/motif scaffolding[266–269]. Furthermore, it is commonly paired with ProteinMPNN[252] to assign sequences to the generated backbones. Demonstrating increasing efficiency, RFdiffusion[257] can be utilised across various design steps in the development of an artificial motor protein

from natural components.

As noted above, advances in structure prediction were a key development for modern structure generation. Independently, they also represented a major milestone toward the long-standing protein folding problem. A wide variety of structure prediction tools have been developed in recent years. In this work, the prediction tools AlphaFold2[258] and Chai-1[259] were used to rank and validate design variants *in silico* before progressing further.

AlphaFold2[258] learns about the evolutionary, physical, and geometrical constraints of the training dataset using a deep neural network architecture. Its outrageous success relies on the development of Evoformer, which integrates multiple sequence alignment data and pairwise residue information to build detailed information about amino acid interactions in their geometric context. AlphaFold2 can therefore be considered a representation-based structure prediction method, in which sequence- and MSA-derived representations are progressively refined before being translated into three-dimensional atomic coordinates (Figure 19C). The refined representations of three-dimensional atomic coordinates are passed to the structure model, which converts the information into a three-dimensional protein model. Local coordinated frames of each residue are predicted (rigid-body frames) using invariant point attention (IPA) to place the atoms in 3D space. Prediction quality is enhanced by recycling, in which intermediate outputs serve as inputs to subsequent design rounds, continuously refining the structure. [258]

An open challenge for AlphaFold2[258] is the modelling of bound small molecules. This motivated the development of AlphaFold3[260], which introduces several fundamental architectural modifications. Evoformer is replaced by Pairformer, thereby simplifying the multiple sequence alignment process. Furthermore, the structure module, which transforms three-dimensional atomic coordinates into a local rigid-body frame for each residue, was replaced with a diffusion module that directly predicts raw atomic coordinates. (Figure 19C) This multiscale denoising process allows the model to generalise across a broad spectrum of chemical entities, including proteins, nucleic acids, small molecules, ions, and modified residues with minimal special casing. As a result, AlphaFold3 eliminates the need for explicit stereochemical losses and most bespoke bonding logic within the network architecture.[260]

Building on the architecture of AlphaFold3[260], Chai-1[259] also enables the prediction of protein-ligand complexes. It integrates protein language models to enable robust prediction even without MSAs (ESM) (Figure 19C). In the training data set, pocket, contacts, and docking constraints were present, which enable the prediction of protein-ligand complexes. Altogether, the approach significantly improves prediction for complex targets and protein-ligand interfaces where evolutionary information is limited.[259]

1.5 Natural molecular motors and artificial designs

1.5.1 Functional diversity and shared mechanistic principles of natural motor proteins

In living organisms, motion is driven by molecular-scale machines. They are crucial for most intracellular transport mechanisms and force-generating processes, making them essential for diverse cellular functions, including mitosis, DNA replication, ATP synthesis, and muscle contraction.[270] A common feature of the natural motor proteins is their ability to transduce an energy source, most commonly the free energy of nucleotide hydrolysis or ion gradients, into directed mechanical motion.[270]

Molecular motor proteins can be categorised according to their geometry of motion into revolutional, rotary, and linear motor proteins (Figure 21).[270]

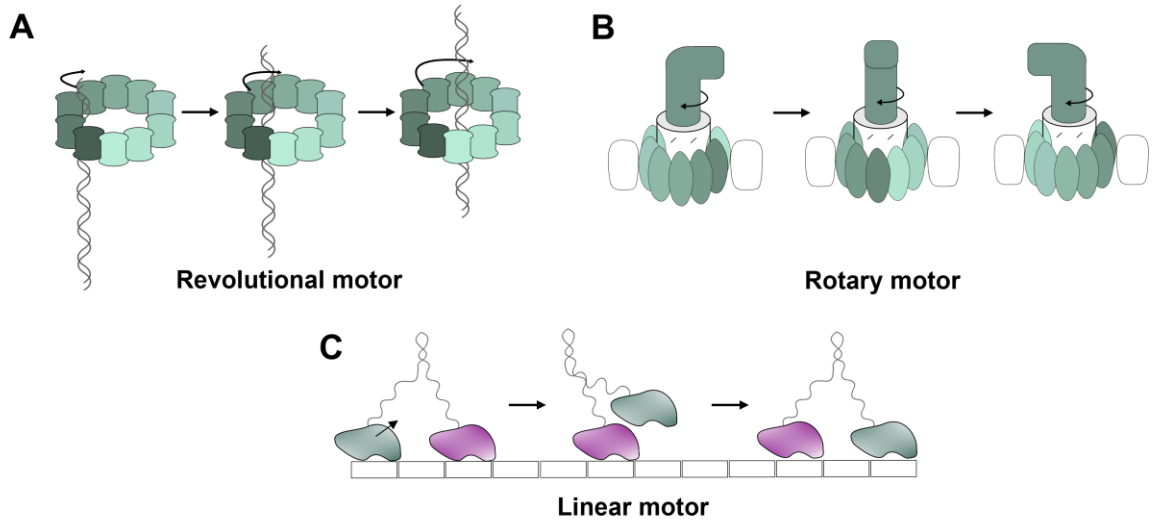


Figure 21. Geometry of motion of molecular motor proteins. A) Schematic representation of revolutionary motion. The substrate rotates around its own axis while undergoing orbital motion around a largely stationary motor. B) Schematic representation of rotary motion. Rotary motors generate rotation around their own axis. C) Schematic representation of linear motion. Linear motors translocate along a linear track.

Historically, only the two categories, rotary and linear, were distinguished. The third mode of revolutional motion was proposed in 2013 based on studies of the dsDNA translocase of the bacteriophage φ29 packaging motor.[271] In this mechanism, the substrate undergoes a revolution around a central axis without the motor rotating around its own axis (Figure 21A).[271,272] Motors assigned to this class are associated with the AAA+ (ATPases Associated with diverse cellular Activities) superfamily and harness ATP hydrolysis to drive tasks such as dsDNA packaging[271] into viral procapsids and chromosome segregation, exemplified by bacteriophage DNA packaging motors and the dsDNA translocases FtsK and SpoIIIE.[270,271] Before the introduction of the revolution model, several of these systems had been interpreted as rotary motors. Rotary motors generate motion by rotation around their own axis (Figure 21B), as performed in nature by the F_0F_1-ATP synthase[273] and bacterial flagellar motors[274]. Rotary mechanisms have also been invoked for a subset of dsDNA translocation motors, such as transcription termination factors[275] or helicases[276]. Finally, linear motor proteins convert

chemical energy into stepwise translocation along linear biological polymer tracks, such as actin filaments, microtubules, or nucleic acids (Figure 21C).[270] Prominent examples include nucleic acid-based motors such as polymerases and helicase[277] as well as cytoskeletal motors such as kinesins[278] and myosins[279].

In addition to classifying motors by their motion geometry, the underlying mechanochemical coupling mechanisms can also be used for classification (Figure 22). There are two commonly contrasting views of motion in motor proteins: the power stroke-based stepping (Figure 22A) and the Brownian ratchet-based stepping (Figure 22B).

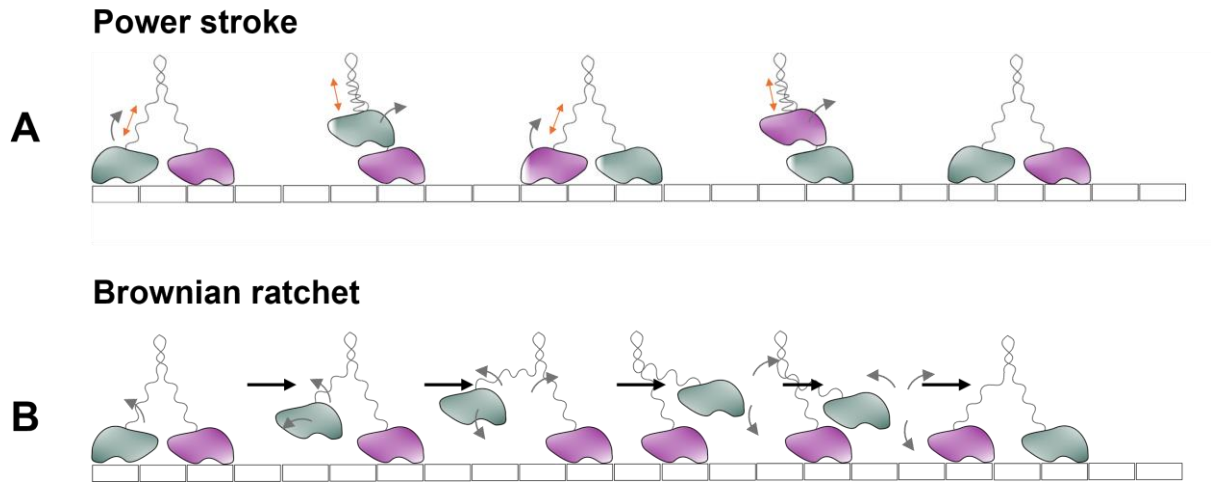


Figure 22. Schematic illustration of two principal mechanochemical coupling mechanisms in motor proteins. A) In a power-stroke mechanism, a force-generating conformational change (orange arrow) is directly converted into forward displacement (grey arrow). B) In a Brownian ratchet mechanism, stochastic thermal fluctuations (grey arrows) are rectified into motion.

In Howard's widely used observational definition, a power stroke is a "large, rapid structural change in a protein that is used to do mechanical work" (Figure 22A).[280] A power stroke is therefore envisaged as a conformational transition between two mechanical states that are coupled to a force-generation mechanism. The generated force is often viewed in terms of the resulting applied strain or the stored elastic energy released, which control the chemomechanical cycles.[281–283] Controlling the cycle is possible because the generated force reshapes the active site environment by modulating the activation barriers for substrate binding and product release, thereby generating kinetic asymmetry and directional flux.[283] A power stroke-based stepping has been shown to occur for many different motor proteins, with prime examples being the linear motor kinesin with its neck linker performing large conformational changes[284] as well as myosin with its lever arm swing[285]. In contrast, Brownian-ratchet stepping relies on diffusive exploration of neighbouring registers on the track, thereby undergoing thermally driven forward-backwards fluctuations over distances comparable to the step size (Figure 22B).[286] As the step size increases, rebinding to a forward site becomes less likely, since a larger volume of diffusional space must be explored before rebinding. A purely Brownian-ratchet-based mechanism is therefore only possible for small step sizes. Larger

steps require additional directional bias, for example, due to conformational gating, elastic strain, or power-stroke-like contributions.[287,288] This enables the forward positions to be preferentially stabilised, while backwards stepping is disfavoured, by building a kinetic and thermodynamic gate that rectifies Brownian motion into biased forward translocation.[289,290] A few molecular machines are ratchet-dominated, meaning their stepping relies primarily on thermally driven fluctuations between neighbouring registers. At the same time, chemical transitions within the molecular motor domain act as kinetic-gating steps that bias forward motion.[286,291] RNA polymerase[289,290] or ribosomal translocation[292] are examples of linear motor proteins following a Brownian-ratchet dominated stepping. In most biological motors, stepping is better viewed as a hybrid mechanism: conformational changes often provide a power-stroke component that can contribute to force generation and bias forward motion. At the same time, Brownian fluctuations still play a central role in state-to-state transitions.[287] Consequently, power-stroke and Brownian-ratchet mechanisms are not mutually exclusive. Different steps within the same mechanochemical cycle combine power stroke and Brownian motion in varying degrees, adapting to distinct functional demands such as speed and processivity.[293] An additional classification can be performed regarding the walking mechanism. Focusing on linear motor proteins, three mechanisms are dominant (Figure 23). Conceptually, linear motor motion is decomposed into repeated cycles of detachment, forward displacement, and rebinding to the track. In the hand-over-hand stepping motion, the two track-binding domains alternate these states out of phase, such that each cycle produces an “overtaking” event and the leading site changes every step (Figure 23A).[294] This form of motion is observed in cytoskeletal motor proteins such as kinesin[295] or bipedal myosins like myosin V[279]. In contrast, linear nucleic acid motors, like the hexameric helicases (DnaB), perform a sequential hand-over-hand motion.[277] Sequential hand-over-hand motion describes a translocation mechanism in oligomeric ring motors in which individual subunits, one after another, undergo nucleotide-state-coupled cycles of release and re-engagement with the polymer, such that the track-contacting subunit effectively moves from the back to the front of the helical grip (Figure 23B).[277] In contrast, dimeric helicases, like SF1 and SF2 helicase, move in an inchworm gait. There, the two binding sites move in a coordinated but non-overtaking manner (Figure 23C). The same site remains the leader, and forward progress arises from two sequential substeps that first shift one site and then the other.[276,296,297]

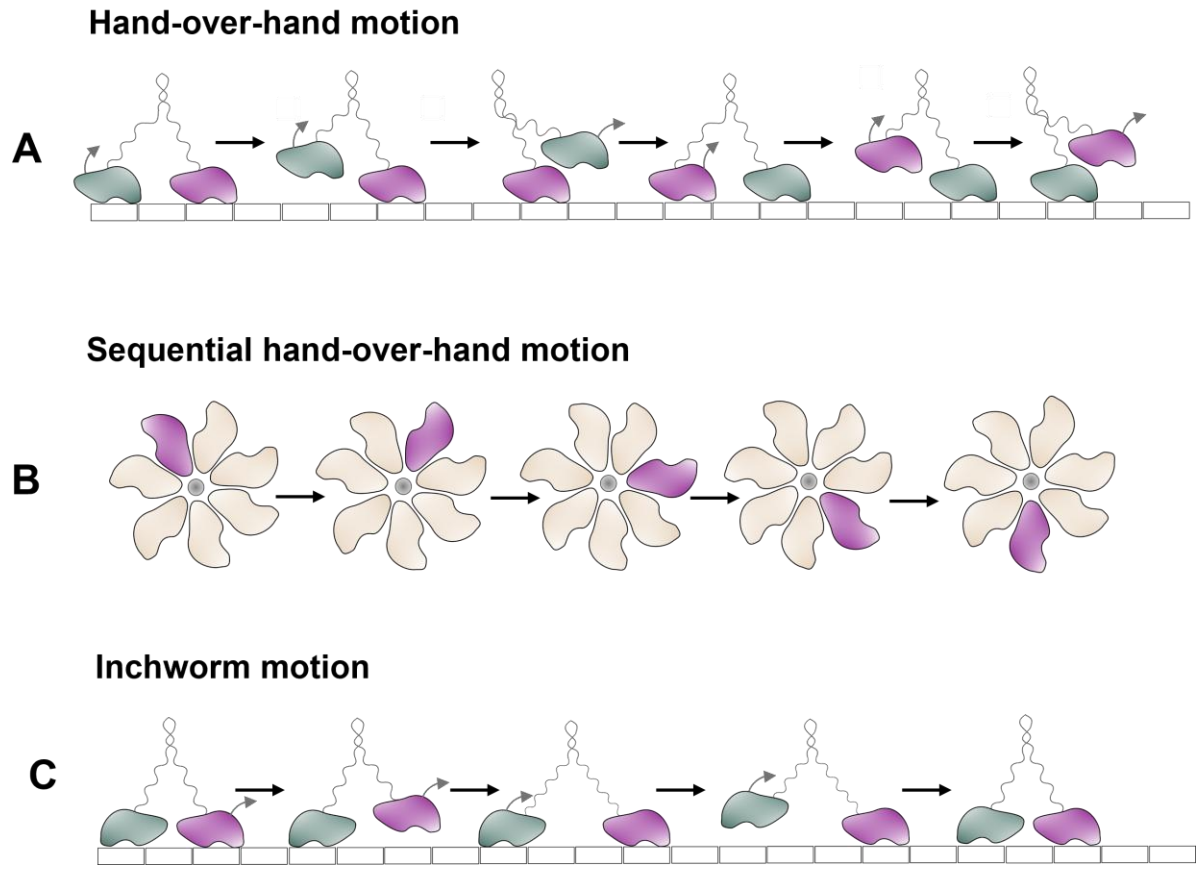


Figure 23. Stepping mechanisms of linear motor proteins. A) In hand-over-hand motion, two motor domains move in a coordinated manner and alternately overtake each other. B) Sequential hand-over-hand motion is characteristic of multimeric motors, in which individual subunits step sequentially along the track. C) In inchworm motion, two binding sites move coordinately while maintaining their relative order, without overtaking.

Linear motor proteins act in diverse biological contexts, on different tracks, with distinct biological functions. Even though walking mechanisms vary across motor superfamilies, the underlying biophysical principles are conserved.[295,298,299] The ability of the molecular motor to perform work depends on precise molecular building blocks that coordinate energy transduction and control directionality and processivity.[1] The overall architecture of linear motor proteins can be divided into the following regions: the catalytic motor domain (Figure 24, blue), the structural coupling element (Figure 24, pink), the dimerisation or oligomerisation domain (Figure 24, green), and the track binding or interface contact domain (Figure 24, orange).[293] Even though the regions are present in all linear motor proteins, the overall architecture and the given names for those domains vary (Figure 24).[278,293,298–300] A few functional requirements, such as converting chemical-free energy into mechanical work, generating directional bias, and maintaining processive track association, are achieved through the detailed coordination of functional modules. The architectural building blocks of different linear molecular motors are displayed in Figure 24.

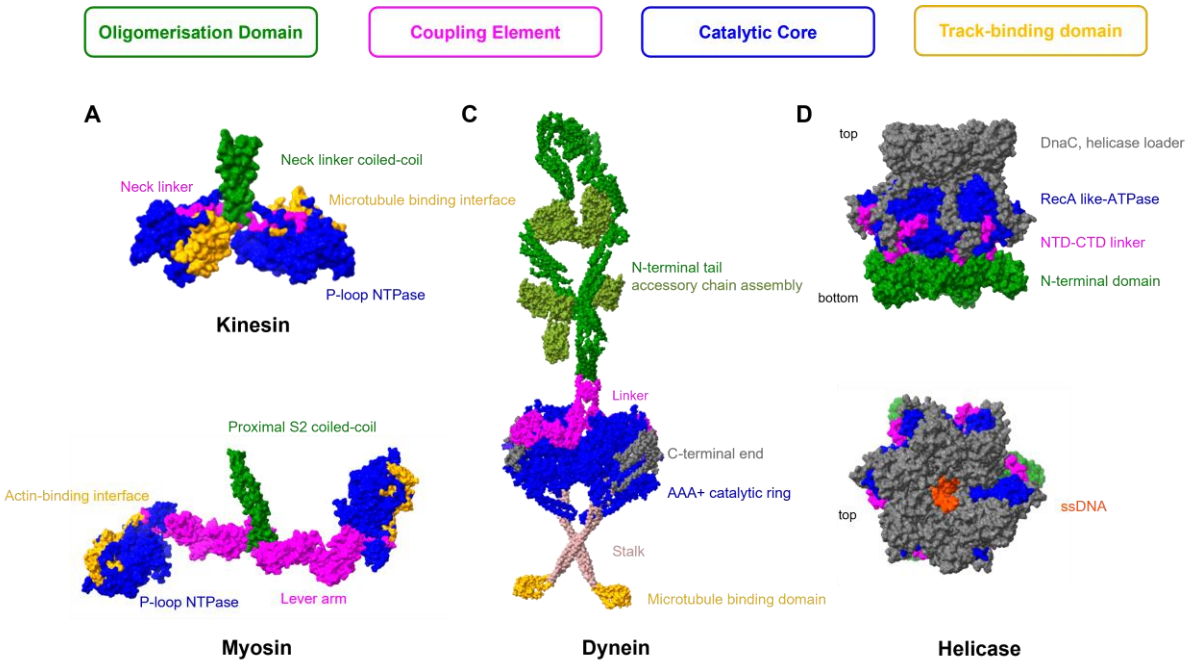


Figure 24. Structural comparison of representative linear motor proteins.Track-binding interfaces/domains are shown in orange, catalytic cores in blue, coupling elements in magenta, and oligomerisation scaffolds in green. A) Surface representation of the Kinesin motor-neck fragment (PDB 3KIN[301]), with the P-loop NTPase motor domains (blue) embedded microtubule-binding interfaces (orange), neck linkers (magenta), and proximal neck coiled-coil (green). The extended coiled-coil stalk and cargo-binding tail of full-length kinesin are not present in this structure. B) Surface representation of smooth-muscle heavy meromyosin fragment (PDB 3J04[302]): Following strucutreal elements are colour coded, the actin-activated ATPase motor domains (blue) with embedded actin-binding interfaces (orange), lever arm (pink), and proximal S2 coiled-coil (green). The extended filament-forming tail of full-length myosin is not present in this structure. C) Surface representation of full-length human dynein-1 (PDB 5NVU[303]): AAA+ ATPase rings (blue), linker domains (pink), N-terminal tail/accessory-chain assembly (green), and stalk-mounted microtubule binding domain (orange). D) Surface representation of the DnaB-DnaC helicase complex on ssDNA (PDB 6QEM[304]): hexameric DnaB with RecA-like ATPase domains (blue), NTD-CTD linkers (magenta), N-terminal oligomerisation domains (green), and ssDNA (dark orange) threaded through the central pore; helical loader DnaC is shown in grey.

For a motor to work, it is crucial that it can convert chemical-free energy into mechanical work. For many cytoskeletal linear motor proteins, the binding of ATP, its hydrolysis to ADP, and the subsequent product release drive an allosteric cascade in which different nucleotides induce discrete rearrangements within the catalytic core (Figure 24, blue). In kinesin (Figure 24A) and myosin (Figure 24B), this catalytic core is formed by a P-loop NTPase motor domain that contains the nucleotide-binding site and the cytoskeletal track-binding interface.[301,302,305] In dynein (Figure 24C), the catalytic core is formed by a ring-shaped AAA+ ATPase motor domain, in which ATP-dependent conformational changes are transmitted to a large linker domain that serves as the motor's principal mechanical element.[303,306] In hexameric replicative helicases (Figure 24D) such as DnaB, the catalytic core is formed by C-terminal RecA-like ATPase domains arranged into a hexameric ring.[307] These sub-nanometer shifts are subsequently transmitted to a structurally integrated coupling element (Figure 24, magenta). This element acts as the chemomechanical transducer, amplifying internal rearrangements of the catalytic core into large-scale mechanical displacements.[308] By alternating between distinct structural states such as the docking of a neck linker (kinesin, Figure 24A)[284], the swinging of

a lever arm (myosin, Figure 24B)[285], the nucleotide-dependent movement of the linker domain across the AAA+ ring (dynein, Figure 24C)[309] or the rotation of a ring-loop (helicases, Figure 24D)[310,311] in which ATPase-driven conformational changes are coupled to DNA-binding pore loops within the central channel, motor proteins effectively converts chemical free energy into a directed physical conformational change.

In linear cytoskeletal motor proteins, directionality is not an inherent property of ATP hydrolysis itself but emerges from the stereospecific interaction between the motor domain and its asymmetric substrate track (e.g., a polar cytoskeletal filament or a directional nucleic acid strand).[287,312,313] The modular geometry of motor protein and track allows binding in only one specific orientation.[312,314] Binding affinity to the track is coupled to changes in the nucleotide state. Overall, the geometry and the varying affinities create a kinetically inequivalent energy landscape.[315,316] In this environment, forward transitions are preferentially sampled and stabilised, whereas backward steps are strongly suppressed by higher activation barriers and reduced rebinding probability.[281,313,317] Within this framework, the motor effectively functions as a Brownian ratchet, in which stochastic thermal fluctuations are biased toward unidirectional progress.[287,318,319] The track-binding domain defines the position of interaction with the track, while the coupling element dictates the direction of the mechanical tension that prevents reversal and supports directional movement of the track-binding domains.[285,287,320,321] For hexameric helicases (Figure 24D), the track-binding interface is not a separate peripheral domain but is formed by DNA-binding loops that project into the central pore of the ring and engage the nucleic-acid substrate from within the oligomeric assembly.[299,307,322] In kinesin (Figure 24A) and myosin (Figure 24B), the track-binding interface (orange) is likewise embedded within the catalytic motor domain rather than being organised as a separate peripheral domain.[301,302] In contrast, dynein (Figure 24C) spatially separates the catalytic AAA+ ring from the microtubule-binding domain, which is positioned at the distal end of a long coiled-coil stalk.[303]

Processivity, the ability to undergo multiple catalytic cycles and mechanical steps before dissociation, is structurally facilitated by an oligomerisation scaffold (Figure 24, green).[323,324] Whether implemented as a coiled-coil stalk in bipedal motors (Figure 24A, Figure 24B)[325,326], the dimeric heavy-chain tail of cytoplasmic dynein (Figure 24C)[327] or a hexameric ring-shaped assembly in hexameric helicases (Figure 24D)[328]. These architectures physically tether multiple catalytic subunits, enabling multivalent track engagement and ensuring that at least one motor domain remains bound to the substrate while the others undergo translocation.[293] The structural connection alone is insufficient to prevent the complete detachment from the track. Inter-subunit communication, through strain-dependent gating of ligand affinities, is also crucial for precise coordination of stepping.[329,330] Intramolecular tension is transmitted through the coupling elements (Figure 24, magenta) or the scaffold to regulate the chemical cycles of

the subunits out of phase. Conformational states present in one subunit physically modulate the activation energy for substrate binding or hydrolysis in the adjacent subunit.[330] In kinesin (Figure 24A), the neck linker controls the nucleotide-dependent transitions of the two motor heads in a defined manner, thereby favouring alternating microtubule engagement rather than simultaneous detachment.[331] In processive myosin (Figure 24B), intramolecular strain transmitted through the lever arms gates nucleotide release and ATP-dependent detachment of the leading head, allowing one head to remain actin-bound while the partner advances.[332] For dynein (Figure 24C), processive minus-end-directed motility therefore depends not only on dimerisation of the two heavy chains but also on nucleotide-dependent coordination between the two AAA+ motor rings and their microtubule-binding domains.[327,333] In nucleic acid-based linear motors, substrate-mediated load and grip that gate forward versus reverse translocation prevent the dissociation of both feed from the track.[334] In hexameric helicases (Figure 24D), analogous coordination is achieved by sequential ATPase cycling around the ring, coupled to stepwise engagement of the nucleic-acid strand within the central pore.[322] The efficiency of a motor protein is defined by the tightness of its mechanochemical coupling, meaning the ratio of productive mechanical steps to ATP molecules hydrolysed.[335]

Natural motor proteins differ widely in motion geometry, mechanochemical coupling and stepping mechanism. Nevertheless, they share a common functional principle: directional, processive and efficient translocation emerges from the coordinated interplay of catalytic activity, track binding, coupling elements and oligomerisation scaffolds.

1.5.2 Artificial motors: from small molecules to proteins

1.5.2.1 Key terms for molecular motors

Before discussing specific design approaches of artificial motor proteins, it is useful to define several key terms that frequently appear in the design of artificial molecular motors. At the most general level, the essential architectural elements are the track, foot and leg (Figure 25A).[1] A track refers to the structured pathway along which a motor moves. In natural systems, tracks range from cytoskeletal filaments such as actin[336] and microtubules[337] to nucleic acid polymers, such as DNA[338,339] or RNA[340], and even unfolded polypeptide chains in proteolytic machines[341]. In synthetic systems, this may be an intramolecular axis[342,343], a periodic surface lattice[344], a polymeric template, such as DNA[214,345,346], or a surface-tethered substrate track[347]. The track provides spatially organised binding sites and defines the coordinate system in which motion occurs.[1] It can additionally encode positional information via sequence-specific interactions.[214,348,349] The term foot denotes the motor subunit that directly interacts with the track, mediating stepping through reversible binding and release.[1] The part that refers to a foot of artificial motor proteins is the track-binding domain in natural motor proteins.[333,350] In most synthetic systems to date, the motor foot primarily functions as a track-binding domain and typically lacks intrinsic catalytic activity.[214,346,351] The connector of the feet to the remainder of

the motor is the leg (also referred to as the arm).[1] In natural cytoskeletal motors, the leg (e.g. lever arm in myosin[285] or neck linker in kinesin[331]) transduces mechanical energy and controls the processivity.[285,331] In most designs, legs are implemented as linkers with defined length and flexibility and are connected to the feet and the motor body through short junction regions that function as hinges or joints.[352,353] Their mechanical properties are a central design variable, as they determine the accessible stepping geometry: leg length sets the reach between binding sites and thus, contributes to the step size, whereas leg stiffness and attachment geometry constrain the configurational space of the feet and can thereby influence directional bias.[1]

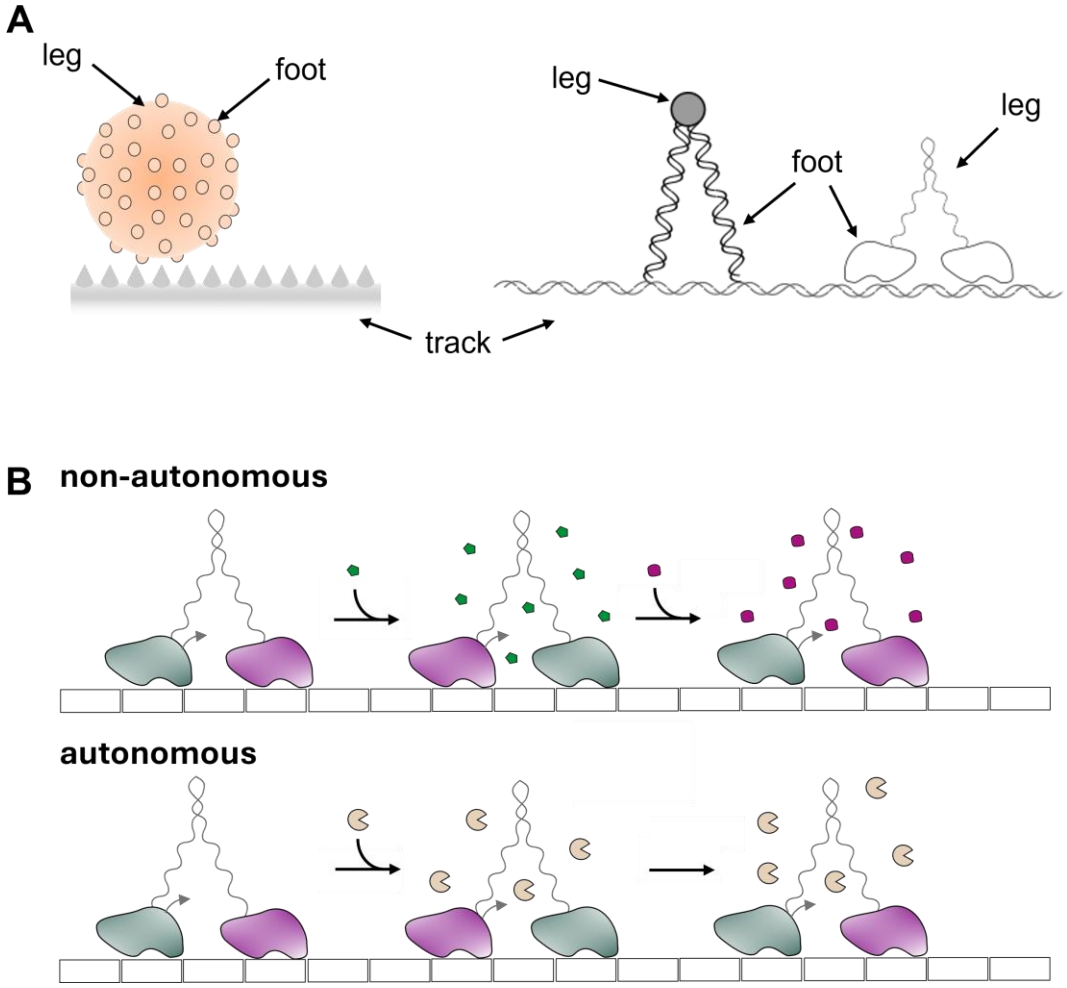


Figure 25. Conceptual terminology for artificial motor protein design. A) Modular building blocks of artificial motor proteins, illustrating alternative architectures of legs, feet, and tracks. B) Operating modes of artificial motor proteins. Non-autonomous, or clocked, motors require sequential external varying inputs, such as chemical, optical, or DNA-based stimuli, to progress through their motion cycle. In contrast, autonomous motors proceed after an initial input or fuel supply without further stepwise external control.

The operational logic of artificial motor proteins is commonly categorised into non-autonomous and autonomous mechanisms (Figure 25B). A non-autonomous motor requires discrete external intervention to progress through its stepping cycle.[354] Each transition between states is typically triggered by a specific externally applied stimulus, such as the addition of a chemical reagent[214], a strand-displacement event[355], a change in pH[356,357], an electrochemical

potential[358,359], or a light pulse[360,361]. Consequently, motion proceeds in a stepwise, externally gated manner. By contrast, an autonomous molecular motor undergoes repeated cycles once supplied with a continuous energy source.[354] After initiation, the system progresses without step-by-step external intervention, typically driven by consumption of a chemical fuel[362], constant irradiation[363,364], or an intrinsic catalytic cycle[347,365]. Such systems maintain net directional motion until the fuel is depleted, a structural failure occurs, or the machine dissociates from its track.

1.5.2.2 Evolution of artificial motor proteins

In chemistry and nanotechnology, the bottom-up construction of synthetic molecular machines has been a long-standing ambition[366,367], echoing Richard Feynman's quote "What I can not create, I do not understand".[368] The significance of this endeavour was underscored in 2016, when the Nobel Prize in Chemistry was awarded to Jean-Pierre Sauvage, J. Fraser Stoddart and Ben Feringa for pioneering key design principles of artificial molecular machines.[369] Building on these foundations, the field of molecular machines has produced a broad spectrum of controllable assemblies ranging from rotary motors[364] to molecular shuttles and switches[370,371], to directional transport systems, such as molecular pumps[372], ratchets[373], and surface-walking constructs[374,375]. Common to these approaches is the goal of converting external energy inputs into biased motion that breaks detailed balance and generates net mechanical work at the nanoscale. The biggest challenge in the design of synthetic molecular motors is processivity, speed, and robust directionality, which depend not only on how energy is supplied but also on how motion is coordinated along a defined pathway.[376,377] While several synthetic systems already operate on tracks in a broad sense, these tracks are often not readily programmable, modular, or extendable.[378–380] Using DNA as a molecular track enabled the programmability of binding sites, as sequence information can encode both binding-site identity and the energy landscape.[381–383] After the first DNA walker was developed in the late 1990s, 388 a variety of DNA walker architectures, including unipedal, bipedal, and multipedal walkers, have been developed.[345,346] The main difference between these systems lies in their stepping mechanisms, energy sources, and transduction.[345,346] In biological motors, the energy generation is directly coupled to a conformational rearrangement to generate a power stroke, whereas most DNA walkers operate on a chemical ratchet mechanism.[383,384] In systems where the movement depends on strand-displacement, the free energy released by forming a more stable DNA duplex reshapes the energy landscape of the motor and thereby drives the stepping.[385] The system can also be controlled by modifying the hybridisation equilibrium through light-induced structural changes or by irreversible phosphodiester bond cleavage, thereby altering the system's chemical free energy.[364,386] In all cases, a source of external energy biases stochastic motion rather than generating an intrinsic mechanical stroke.

The directionality of natural motor proteins derives from stereospecific interactions with

inherently polar tracks[387,388] and from strain-mediated mechanochemical gating.[389] DNA walkers lack both prerequisites for the directional motion of natural motor proteins.[345,346] But DNA walkers use alternative processes to achieve directionality, namely chemical asymmetry, such as burnt-bridge mechanisms that irreversibly degrade previously occupied sites[390], or kinetic bias introduced by toehold design and hybridisation energetics[385]. To bias Brownian motion into a net forward transport, fuel consumption or irreversible reactions are utilised.[391] In biological motors, processivity emerges from coordinated multivalent attachment mediated by strain-dependent gating between motor domains.[389] DNA walkers instead rely on kinetic control and track design. Optimised binding affinities and strand-exchange rates increase the probability of forward stepping before detachment[392], while burnt-bridge systems suppress backward transitions at the cost of track consumption.[390]

Cytoskeletal motors move with distinct mechanochemical gaits (e.g. hand-over-hand or inchworming) that are driven by power-stroke coupling.[280] In contrast, DNA walkers typically lack an intrinsic power stroke[345,346]; nevertheless, their geometry can be programmed to produce inchworm-like stepping[393] or hand-over-hand-like alternation[394]. In multipedal or multivalent formats, the resulting trajectories frequently resemble crawling or rolling along the substrate[395].

A key factor underlying the success of DNA walkers is the high programmability of both the track and the motor components.[381–383] With recent advances in protein design, particularly AI-assisted approaches[249,251], proteins have emerged as similarly programmable building blocks, motivating the development of protein-based walkers.

One of the earliest protein-based designs, the autonomous motor the Lawnmower[347], recapitulates the burnt-bridge principle previously established in DNA rollers[396] (Figure 26A). This artificial protein motor consists of multiple protease “blades” assembled on a bead that roll across a peptide-coated surface.[347] Protease-catalysed peptide bond hydrolysis irreversibly converts substrate into product, dissipating chemical free energy and destroying previously visited binding sites.[347] Processivity is enhanced through multivalent surface engagement: multiple protease domains simultaneously interact with the substrate, reducing the probability of complete detachment, while successive cleavage events maintain forward bias.[347] More recently, DNA has been repurposed as a programmable track for protein-based walkers. Tumbleweed[214], an artificial motor composed of three ligand-responsive DNA-binding repressors, is an example of a protein motor that uses DNA as the track (Figure 26B). In Tumbleweed, energy for forward stepping is achieved through ligand-dependent modulation of DNA-binding affinities.[352] Sequential ligand pulses drive the system out of equilibrium, altering binding states in a controlled manner and rectifying thermal fluctuations without invoking a classical power stroke.[352] Directionality arises from the spatial arrangement and orientation of three DNA-binding sites combined with geometric constraints imposed by the

motor architecture, which restrict accessible stepping pathways.[214,352] Processivity is ensured by its three-legged design, which keeps at least one foot bound to the track during the stepping cycle.[352] Despite the diversity of nanomachines and DNA-based artificial motors, comparatively few artificial protein motors have been realised to date. However, new opportunities towards alternative strategies for artificial motor proteins are explored.[1,397]

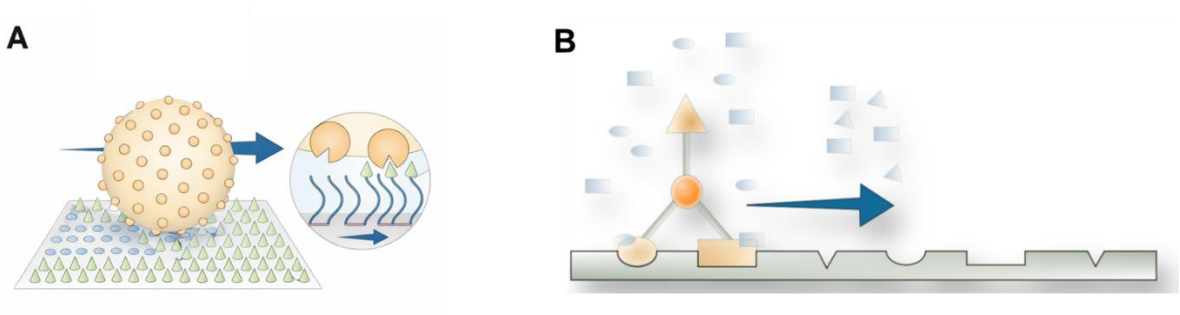


Figure 26. Examples of protein-based artificial molecular motors. A) The Lawnmower motor progresses along a peptide track through protease-mediated substrate cleavage, resulting in rolling-like motion. Modified from Korosec *et al.*[347] B) The Tumbleweed motor translocates along a DNA track through sequential ligand-induced binding and release of its repressor domains. Modified from Nilsson *et al.*[214]

Comparing the variety of artificial motors designed to date with natural motors, it is apparent that the efficiency of natural motors has not been reached. The efficiency of molecular motors is commonly characterised by two principal parameters: speed and processivity. Due to the precise mechanochemical coupling and strain-mediated coordination of the two feet, natural cytoskeletal motors achieve remarkable values for both. For DNA-based walkers, the processivity depends remarkably on the overall architecture. Bipedal DNA walkers[355,393,394,398–404] (Figure 27, orange circles) exhibit limited run lengths, as fewer interactions occur with the track at any given time, they are therefore lost more often. For multipedal DNA nanoparticle motors[395,405–411,411–415] (Figure 27, orange triangle), in which tens to millions of feet are tethered to a central hub, processivity is increased by extensive multivalent binding.[396] Having more track interaction domains present dramatically reduces the probability of complete dissociation, thereby enhancing effective processivity. But processivity is not only determined by the number of legs. The motor-track interface also plays an important role in the processivity. Protein-based rollers, such as the Lawnmower motor[347] (Figure 27, green triangle), exhibit lower processivity than multivalent DNA rollers (Figure 27, orange triangle) because tethering and energy-dissipating steps are intrinsically coupled. DNA rollers bind the track via DNA-hybridisation, performing strand exchange or other chemical reactions; single contacts are lost, resulting in a movement. For the Lawnmower (Figure 27, green triangle), the movement originates from protease-catalysed cleavage. The interaction with the substrate forms the interaction with the track. Due to the irreversible reaction, the connection with the track is lost after catalysis. While this results in movement, protease cleavage also increases the likelihood that no interaction site remains close as the track is consumed, reducing processivity compared to DNA rollers,

where the track remains present.[347,395,407] Compared with natural cytoskeletal motors, bipedal DNA walkers typically exhibit a much stronger reduction in processivity, because they lack intrinsic strain-mediated communication between the head domains and mechanochemical gating that coordinate alternating attachment and detachment.[345,346] Notably, the three-legged protein motor Tumbleweed (Figure 27, green circles) also operates without internal strain-regulated coupling; nevertheless, it achieves comparatively high processivity because detachment and rebinding of individual feet are orchestrated by externally applied, time-programmed ligand pulses.[214,352] This “clocked” chemical gating functionally mimics the coordination strategy of natural motors, maintaining continuous track engagement by ensuring that not all binding domains enter a low-affinity state simultaneously while replacing internal mechanochemical control with externally imposed ligand-dependent affinity modulation.[352]

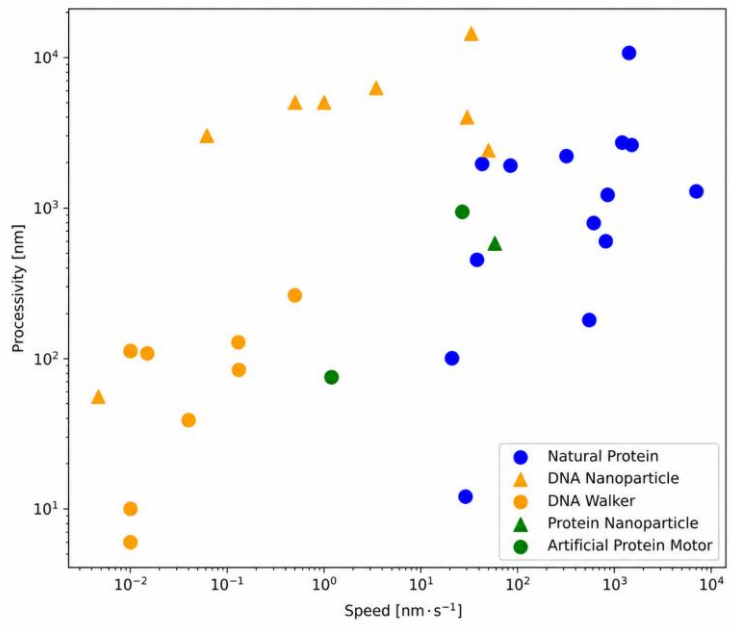


Figure 27. Comparison of processivity and speed across natural and artificial molecular motors. Log-log plot of motor processivity (nm) versus speed (nm·s^{-1}). Natural protein motors[335,416–425] (blue) cluster at high speeds and high processivities. DNA-based systems, including DNA walkers[355,393,394,398–402] (orange circles) and DNA nanoparticle motors[405–408,411–414] (orange triangles), generally exhibit lower speeds and/or reduced processivity. Protein nanoparticle systems[347] (green triangle) and artificial protein motors[214] (green circles) occupy an intermediate regime.

With respect to processivity, values comparable to those of natural motors are already achieved for specific architectures, whereas artificial motors have not yet reached the same speed (Figure 27). This is likely because most natural motors perform a power stroke, thereby increasing the speed to reach the next binding site.[282]

No power stroke is present in DNA-based motor proteins and existing protein motors.[214,345–347] DNA nanoparticle rollers typically achieve higher speed than bipedal DNA walkers because forward translocation emerges from many parallel binding-unbinding and (often) catalytic events within a multivalent contact patch, resulting in continuous rolling/crawling motion.[395,409] In contrast, bipedal DNA walkers advance through discrete, sequential stepping cycles in

which individual feet must repeatedly detach, diffusively search for the next foothold, and complete strand-exchange reactions. These serial, multi-step transitions introduce substantial dwell times and pause states, so net velocity is frequently limited by single-step reaction kinetics and intermittent reconfiguration.[411,426] Despite having three legs, Tumbleweed can achieve comparatively high stepping speed, because its stepping kinetics are governed by rapid ligand-controlled modulation of protein-DNA binding affinities rather than by multi-step strand-displacement reactions.[214,427] Thus, similar features that reduce long-term attachment stability compared with multivalent DNA systems also permit faster turnover between binding states and thereby higher stepping rates.

Although natural motor proteins still set the benchmark in both speed and processivity, recent advances show that engineered systems are increasingly approaching biological performance. In particular, the emerging class of protein-based walkers, even with a comparatively small number of realised designs, has already shown that walking architectures, rather than rolling mechanisms, can match or exceed the performance of many DNA-based walkers (Figure 27). As no power-stroke has been implemented for artificial protein walkers so far, substantial gains in the speed of artificial systems might arise from coupling chemical changes with conformational changes to achieve a power-stroke-like conformational transition. However, reducing off-path trapping would already increase the speed and processivity of novel artificial protein motors based on the Brownian ratchet.[1] A clear pathway toward synthetic protein motors with higher efficiency and more motor-like behaviour can therefore be outlined to narrow the gap to natural systems ultimately.

1.6 Motivation and Aim of this work

The rapidly advancing field of protein design enables the rational engineering of both natural proteins with novel functionalities and functional *de novo* proteins. These developments provide a unique opportunity to construct artificial protein-based molecular motors with programmable architectures and mechanisms. In this thesis, DNA was selected as the molecular track due to its exceptional programmability, as widely demonstrated by DNA walker systems. At the same time, the use of DNA as a track imposes strict requirements: motor motion must be tightly coupled to reversible, controllable, and sufficiently specific DNA binding and release. To meet these requirements, the lactose repressor (LacI) was chosen as the foot of the motor (Figure 28), as it intrinsically exhibits ligand-dependent DNA association and dissociation. This feature makes LacI particularly suitable for non-autonomous motor concepts driven by externally applied ligand pulses, analogous in principle to the Tumbleweed motor.[214] In addition, LacI constitutes an exceptionally well-characterised model repressor, providing a robust structural and mechanistic basis for rational modification of ligand binding, DNA affinity, and allosteric control.

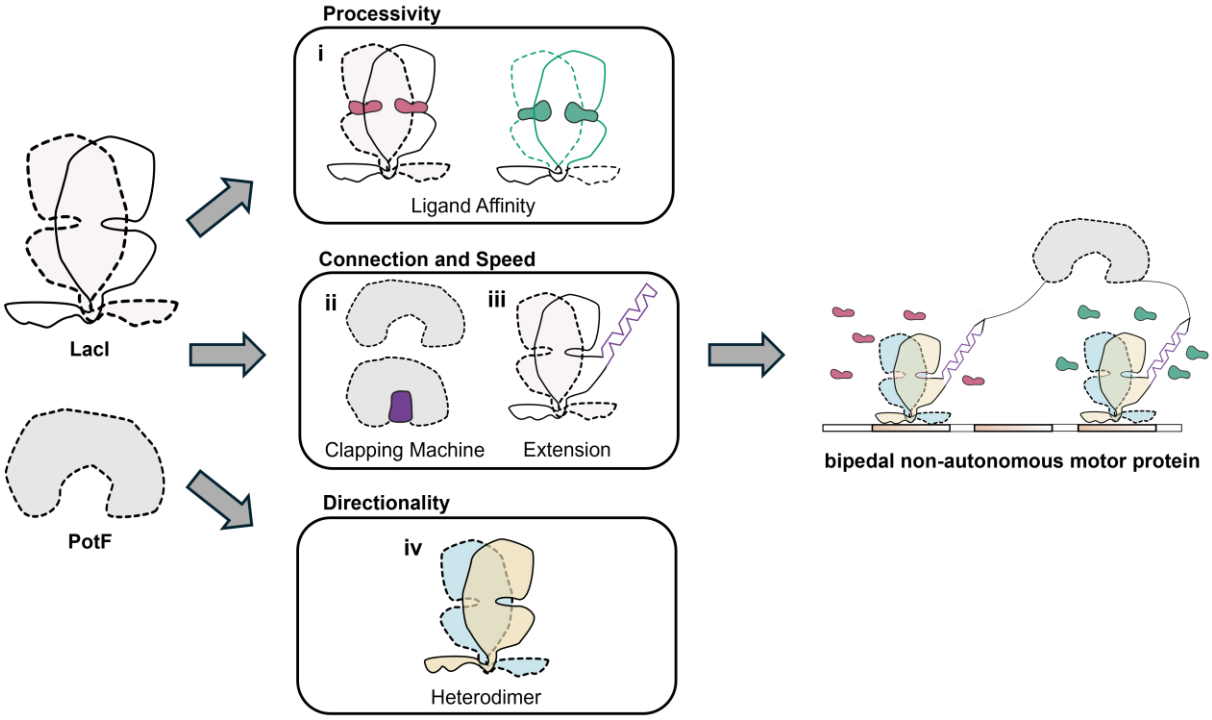


Figure 28. Schematic overview of the stepwise design strategy pursued in this thesis toward a bipedal artificial motor protein. Starting from the natural proteins LacI and PotF, a non-autonomous motor architecture was developed through four main design steps: (i) tuning the ligand-dependent DNA-binding response of LacI-derived feet to enable alternating DNA binding and release, (ii) connecting two feet through a central PotF-based actuator, referred to as the “clapping machine”, (iii) introducing rigid structural extensions to improve mechanical coupling and potentially enhance stepping efficiency, and (iv) converting the naturally homodimeric LacI scaffold into a heterodimeric architecture to provide the asymmetry required for directional stepping.

Only a limited number of protein-based synthetic walking motors have been reported to date, and existing designs are predominantly multilegged. Therefore, the primary aim of this thesis

is to construct a bipedal, non-autonomous motor protein that moves along a DNA track via a biased Brownian-ratchet walking mechanism. This requires the engineering of geometrical constraints, processivity and directional bias. Therefore, this thesis pursues three interconnected design aims (Figure 28). First, the aim is to achieve processivity by creating differential ligand-dependent behaviour in the two LacI-based feet, allowing ligand pulses to function as a selective gate that promotes alternating attachment (Figure 28, i). The second aim is to impose geometric constraints that enable the bipedal architecture and restrict configurational freedom, thereby biasing Brownian exploration towards forward-binding states. Main components toward the bipedal architecture and an elevated speed are the clapping machine (Figure 28, ii) and the structural extension attached to the effector binding domain of the foot (Figure 28, iii). The third aim is to build asymmetry into the motor at both global and local levels through a heterodimeric LacI design (Figure 28, iv) to enforce a strictly bipedal architecture and prevent uncontrolled polymerisation, and to bias asymmetric DNA engagement.

2 Results

2.1 Architecture and mechanism of a novel artificial motor protein

2.1.1 Structural design principles of a LacI-based bipedal motor protein

In analogy to natural motor proteins (Chapter 1.5.1), a functional artificial motor protein must fulfil a set of physical and architectural requirements: reversible track attachment, spatial coupling between multiple binding units, directional bias along an asymmetric energy landscape and sufficient processivity to prevent complete dissociation during repeated stepping cycles. In natural motor proteins, these requirements are implemented through three core architectural elements: a track-binding domain, an oligomerisation domain, and a coupling element. The track-binding domain mediates reversible interaction with the molecular track, the oligomerisation domain connects multiple binding units into one functional assembly, and the coupling element transmits structural and mechanical communication between the different feet. In the terminology of motor protein design, the track-binding domain is commonly called a foot. In contrast, their structural connection and oligomerisation domain are more broadly described as the leg (Figure 29, left side).[376] These main architectural building blocks are used to achieve a LacI-based bipedal motor protein (Figure 29, left side).

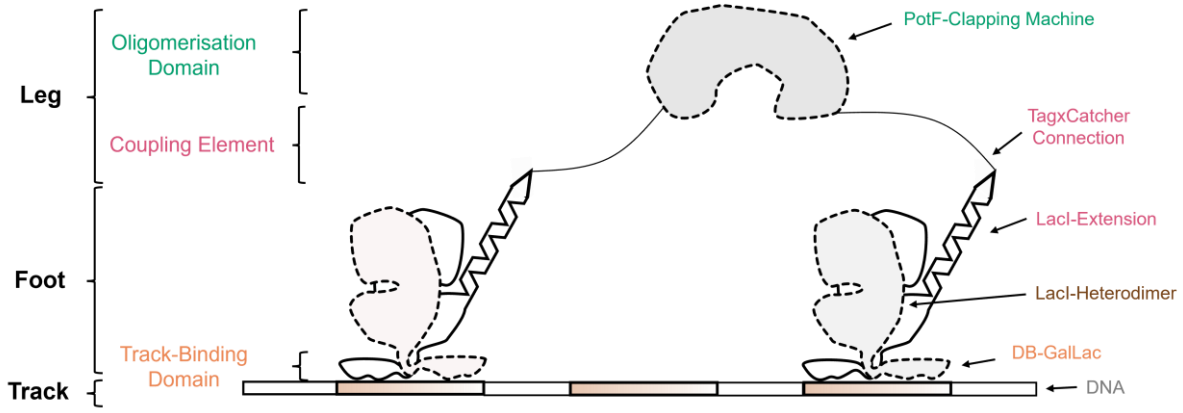


Figure 29. Schematic overview of the LacI-based bipedal motor and its core elements. The left panel illustrates the modular organisation of the artificial motor architecture. Two LacI-based DNA-binding feet are connected via coupling elements (pink) to a central oligomerisation domain (green), establishing a defined bipedal scaffold. The right panel summarises the design features implemented in this study to generate the LacI-based artificial motor architecture.

In the present study, the main design objective is an artificial bipedal motor protein that moves along a DNA track. The bipedal architecture is intended to support processivity by ensuring that at least one foot remains bound to the DNA track. In this way, complete dissociation from the track is disfavoured, while the second foot can transiently detach, explore neighbouring binding sites, and rebind. This process depends on the coordination between foot release, clapping machine movement, and rebinding. Therefore, the amount of ligands required for one step serves as a measure of the efficiency. In contrast to ATP-driven natural motor proteins, efficiency in a non-autonomous system is not defined by a fixed ratio of fuel molecules

hydrolysed per mechanical step. Minimising losses through flexible linkers, off-pathway binding, or complete detachment is crucial for generating an efficient motor.

LacI is particularly attractive as the core component of the foot, as it combines sequence-specific DNA binding with allosteric regulation of DNA affinity. These native properties directly define DNA as the most suitable track material, because defined operator sequences can be arranged along the desired path of motion. Thus, DNA is a suitable programmable track for the motor protein. As a first step, ligand-dependent changes in the DNA association provide a means to externally controlled attachment of individual feet in a non-autonomous stepping cycle (see Chapter 2.1.2.1). However, sequence-specific DNA binding alone is not sufficient to generate a bipedal motor architecture. To convert LacI from a DNA-binding protein into a functional component of a bipedal motor requires mechanical coupling between the feet, controlled oligomerisation, and heterodimerisation to attach the motor components specifically. Starting with a dimeric LacI, where the natural tetramerisation domain is removed, but which retains its intrinsic homodimeric organisation.[428] A first essential modification is the conversion of LacI into a heterodimeric architecture. This is required for both directional track engagement and bipedal organisation, as a homodimeric LacI scaffold would introduce all design features symmetrically into both subunits. Consequently, elements such as connector modules or structural extensions could not be selectively restricted to only one subunit (Figure 29, brown). The heterodimeric organisation also provides a route to directional bias. By engineering the two DNA-binding domains of one LacI dimer to recognise different operator half-sites, the motor-track complex no longer samples equivalent forward and backward binding registers. Instead, asymmetric DNA recognition defines a preferred binding orientation on the track. In the present design, one DNA-binding domain retains the specificity for the *lac* operator (DB-Lac). At the same time, the second subunit is modified to recognise the *gal* operator (DB-Gal) by altering three positions within the HTH motif. This results in the interaction with an asymmetric operator sequence by the DB-GalLac configuration of the HTH motif (Figure 29, orange). The leg of the motor consists of two functional parts: an oligomerisation domain and a coupling element (Figure 29, left). As an oligomerisation domain (Figure 29, green), a ligand-responsive clapping machine is utilised for the motor protein design. A clapping machine is defined here as a protein module that undergoes a ligand-induced conformational change. In the design pursued in this study, PotF is used as the clapping machine. Rather than acting as a classical power-stroke motor domain, PotF is intended to reshape the accessible conformational space of the detached foot. Ligand-induced opening and closing motions should change the distance between the detached foot and potential DNA-binding sites, reduce the diffusional search space, and increase the probability of rebinding at the next forward site rather than at the original position (see Chapter 2.1.2). Thus, the clapping machine is envisioned as the central actuator for generating biased Brownian motion.

The second part of the leg is the coupling element (Figure 29, pink), which connects the LacI-based foot to the clapping machine. In principle, such a connection could be achieved by direct genetic fusion of the protein domains, by chemical crosslinking, or by modular covalent ligation systems. A direct genetic fusion would require the relative order, linker length, attachment site, and orientation of all domains to be fixed within a single polypeptide chain. However, large multidomain fusion proteins may suffer from reduced expression, impaired folding or loss of domain function.[429] Chemical crosslinking could provide an alternative route to connect independently produced protein modules, but often lacks precise control over coupling stoichiometry, attachment site, and orientation unless additional site-specific chemistry is introduced.[430] This may result in heterogeneous products that are difficult to purify and mechanistically interpret. To overcome these limitations, a TagxCatcher connection was chosen as a modular covalent coupling strategy. This approach allows the LacI-based feet, the clapping machine, and the coupling elements to be produced and optimised as separate modules before assembly.

From a motor-design perspective, this modularity is important because foot identity, linker geometry, actuator properties, and connection topology can be varied independently. Thus, the individual contributions of these parameters to directionality, processivity, and stepping kinetics can be experimentally dissected. The covalent nature of the Tag×Catcher connection provides a defined and stable linkage, which is essential for maintaining mechanical communication between the clapping machine and the feet during the stepping cycle. At the same time, the coupling element must transmit conformational changes from the clapping machine to the feet as efficiently as possible.

In this sense, it functionally resembles natural coupling elements such as the kinesin neck linker[331] or the myosin lever arm[285]. A rigid linker is therefore considered advantageous because it constrains the relative orientation of the connected domains and reduces dissipative rotational and translational freedom, which is essential for implementing motor-, sensor-, or switch-like mechanisms.[431] In contrast, a highly flexible linker would be expected to absorb conformational changes and weaken directional communication between the clapping machine and the feet by allowing excessive rotational and translational freedom.[431] To address this requirement, LacI-specific structural extensions (Figure 29, pink) were designed in this work to establish a rigid connection site for the attachment of the coupling modules, which connect the foot to the clapping machine.

2.1.2 Conceptual stepping mechanism of a LacI-based bipedal motor protein

Nature has evolved a wide range of mechanistic principles and force-generating mechanisms to achieve directional motion (see Chapter 1.5.1). The diversity of natural motor proteins illustrates the broad design space available for constructing artificial motor systems. This study focuses on achieving an inchworm-type mode of motion, in which the two feet alternate between bound and unbound states without overtaking one another. In the proposed designs, processivity is enhanced by employing biased diffusion as the mechanochemical principle by focusing the designs to incorporate geometrical constraints that increase the probability of forward rebinding relative to backward rebinding.

The ultimate objective is to develop an autonomously operating motor protein. However, the rational design of an autonomous walker is highly challenging, as it requires the coordinated coupling of enzymatic reaction, conformational switching, inter-domain force transmission, and spatiotemporally controlled track interactions. Therefore, the present study focuses on the design of a non-autonomous, externally controlled motor system. The reduction in mechanistic complexity provides a tractable framework for evaluating whether the modular motor design comprising two feet and a clapping machine is fundamentally capable of producing directed, processive stepping behaviour. Furthermore, structural and functional parameters that require further optimisation for autonomous operation can be identified and addressed in a stepwise manner.

2.1.2.1 A non-autonomous inchworming motor mechanism

Based on the overall motor architecture described in Chapter 2.1.1, a conceptual mechanistic model for a non-autonomous inchworm-type motor was developed. This model describes how ligand-induced conformational changes in the central clapping machine are expected to be mechanically transmitted via rigid linker elements to generate a positional bias in the diffusional search of the detached foot. While track binding of the LacI-based foot remains regulated by specific ligands bound to the effector-binding domain, allowing DNA association and dissociation to be controlled allosterically. The proposed mechanistic cycle of this non-autonomous inchworm motor is outlined in Figure 30 with particular emphasis on the sequence and coordination of individual stepping events. From a physical point of view, the proposed cycle functions by reshaping the free-energy landscape of the motor-track complex through different ligand pulses, imposing non-equilibrium states that reshape it. The presence and absence of different ligands alter the relative stability of DNA-bound and DNA-unbound states, while the clapping machine modulates the accessible conformational space of the detached foot.

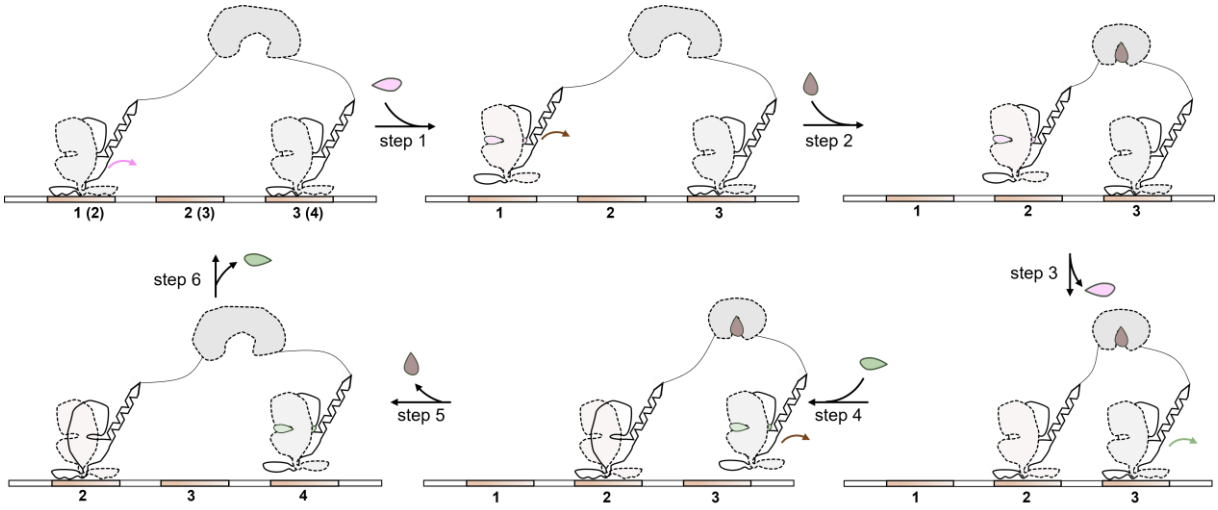


Figure 30. Proposed mechanism of the non-autonomous LacI-based walker controlled by sequential ligand pulses. The walker is proposed to operate through an inchworm-like stepping mechanism. Sequentially supplied ligands (green, brown, and pink) selectively address either the DNA-binding feet or the central clapping machine, thereby controlling the DNA-binding and conformational states of the motor components. The imposed sequence of ligand-binding events promotes alternating release, repositioning, and rebinding of the feet, allowing stepwise progression along the DNA track. Arrows indicate the directionality imposed by the ligand addition.

In the ligand-free state (Figure 30, start), both feet are bound simultaneously to their cognate DNA target sites, forming a bivalent minimum in the motor-track free-energy landscape. This state is expected to be strongly stabilised by avidity, because simultaneous dissociation of both feet is statistically less probable than dissociation of a single foot. This state represents the processive anchor of the motor, because simultaneous engagement of both feet strongly reduces the probability of complete dissociation from the DNA track. In the first step of the cycle, the ligand specific to the rear foot (pink) is added (Figure 30, step 1).

The presence of the ligand shifts the binding equilibrium of the rear foot, changing the rear foot from a high-affinity DNA-bound state towards a low-affinity state. Consequently, the dissociation rate is increased, and the probability of immediate rebinding to the same side is lowered (pink arrow). The front foot remains bound and defines the anchored reference point of the system. The one-foot-bound state is the most critical step throughout the cycle, as the track interaction relies only on the interaction of one foot with the track, while the remaining foot undergoes tethered diffusion, increasing the likelihood of losing contact between the foot and the track. The ligand interaction within the rear foot induces the conformational change of the effector binding domain. Since the foot is coupled to the motor's remaining counterparts by a rigid extension, the conformational change is translated into a displacement along the track (pink arrow). In the second step (Figure 30, step 2), the ligand for the clapping machine is added (brown). Closure of the clapping machine reduces the spatial separation between the detached rear foot and the next forward DNA-binding site (brown arrow), thereby changing the boundary conditions for the tethered diffusion of the detached rear foot. The closing motion

reduces the accessible search volume by reducing the effective distance to the forward binding site, thereby decreasing the entropic costs of binding in the forward direction. Therefore, the effective on rate for the binding site in the forward direction is higher than that of the original binding site. While forward rebinding, backward rebinding, and complete detachment are all thermally accessible, the different transition probabilities enriched by rigid coupling between the building blocks enhance the directionality of the system. Therefore, the forward pathway should become statistically more favourable than the backward stepping. The removal of the ligand from the rear foot (pink) in the third phase restores the high-affinity DNA-binding state (Figure 30, step 3), as the rear foot rebinds at the next binding site (Figure 30, binding site 2), thereby completing half of the stepping cycle and restoring the stable two foot bound state. The rear foot becomes the newly positioned trailing foot, since an inchworm-type mechanism, where the two feet are not intended to overtake each other, is followed rather than a hand-over-hand gait. The addition of the ligand specific for the front foot (green) shifts the foot into a low-affinity DNA-binding state and promotes its dissociation from the track (green arrow, Figure 30, step 4). The motor thereby reaches a second one-foot-bound intermediate step, where the rear foot remains attached to the DNA and serves as the processive anchor. Removal of the ligand from the clapping machine (brown) is then expected to reopen the clapping machine (Figure 30, step 5), which introduces a forward displacement bias during the diffusional search of the detached front foot (brown arrow). Reopening of the clapping machine is expected to shift the accessible conformational space of the detached front foot towards the next forward binding site (binding site 4). This makes rebinding to the forward register more likely than rebinding to the previous site. Removing the ligand specific for the front foot (green) restores high DNA affinity, allowing the front foot to rebind and complete one full translocation cycle (Figure 30, step 6). Thus, one full translocation step is generated by a sequence of externally applied ligand pulses. Although both feet may, in principle, be based on closely related binding modules, implementation of a non-autonomous motor requires orthogonal ligand-specificities of the feet to ensure selective and temporally resolved actuation.

The model defines a non-autonomous, externally clocked inchworm cycle. The aim of a motor is high efficiency, which requires specifically tailoring the directionality and processivity of a motor protein. The kinetic asymmetry between forward and backward rebinding introduced by the ligand-dependent modulation of the search space is crucial for the directionality of the walking. In the discussed mechanism, the directionality of the motor arises from kinetic asymmetry between forward and backward rebinding pathways, generated by the combined effects of asymmetric DNA recognition, selective foot release, and actuator-dependent modulation of the conformational search space.

Processivity depends on the competition between productive rebinding of the detached foot and complete dissociation of the remaining bound foot. The one-foot bound state resembles the most vulnerable state of processive motion, as each foot has a finite dissociation rate, meaning that complete detachment can occur if the remaining bound foot dissociates before the unbound foot productively rebinds. Minimising the time in the vulnerable state by increasing the stepping speed is crucial. The stepping speed, which can also be defined as the time the motor is in the one foot bound state, is expected to be limited by the time required for the external ligand exchange, as well as the time required for internal changes, like allosteric switching of DNA affinity, actuator opening or closure, tethered diffusion, and productive rebinding. The speed and the directionality of the system are envisioned to be increased by the rigidity of the coupling elements, as rigid connections should minimise the sampling space and reduce the time required to reach the forward binding site. This increases the likelihood of a successful step.

Although the proposed system is non-autonomous, it already embodies several design principles required for more advanced artificial motor proteins: externally driven non-equilibrium cycling, asymmetric track recognition, bipedal processive anchoring, actuator-dependent modulation of the energy landscape, and mechanically constrained coupling between actuator and feet. Therefore, the present system represents an important intermediate step towards persistent, directional, and eventually autonomous protein-based motion.

2.1.2.2 An autonomous motor inchworming mechanism

To convert the externally clocked motor described above into an autonomous system, several additional design requirements must be fulfilled. The most immediate prerequisite is the elimination of externally applied ligand pulses. Consequently, the reversible dissociation and rebinding of the feet to the DNA track must no longer be triggered by sequential ligand addition. Instead, controlled dissociation and association of the feet is envisioned to emerge directly from the mechanochemical cycle of the clapping machine. Autonomous motion requires that substrate turnover continuously resets the distribution of the conformational ensemble of the clapping machine, thereby disrupting the balance within the motor-track system and altering the overall position of the motor's core components.

A potential autonomous inchworming mechanism based on the proposed building blocks (Chapter 2.1.1) is depicted in Figure 31. The motor is envisaged to start from a state in which both feet are bound to the track. This bivalent state again represents a metastable minimum in the motor-track free-energy landscape. It provides the processive anchor from which one foot can be released without immediate loss of track association of at least one part of the motor protein.

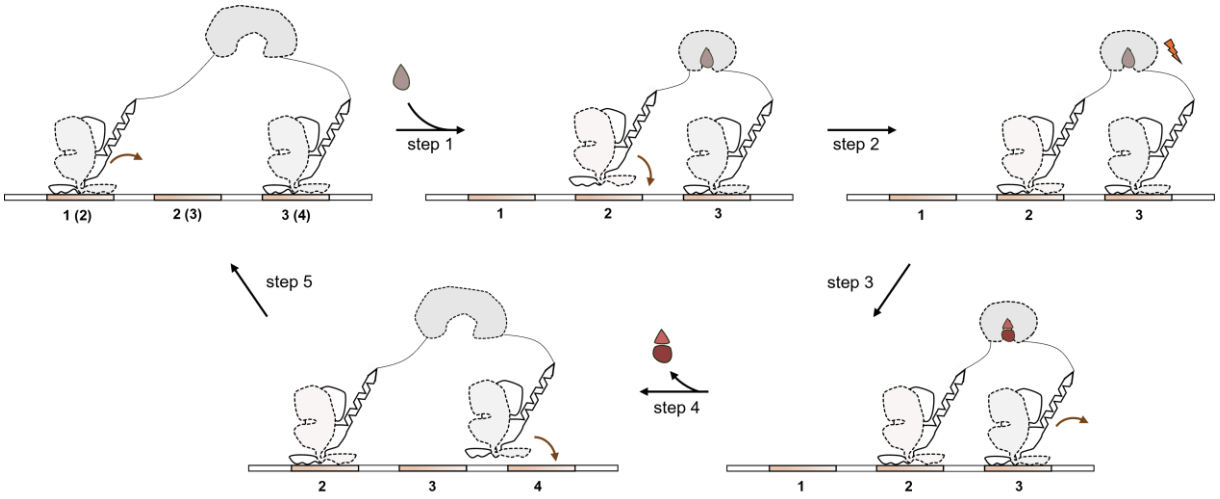


Figure 31. Proposed mechanism for an autonomous LacI-based walker. Substrate-induced conformational changes of the clapping machine are transmitted to the two DNA-binding feet through rigid connectors, thereby coupling catalytic turnover to alternating foot release and forward rebinding. Repeated substrate-to-product conversion cycles, shown as conversion of the substrate (brown drop) into products (orange triangle and red circle), are proposed to provide the energy input required for autonomous directional motion along the DNA track. Arrows indicate the proposed direction of movement.

The addition of the substrate, which interacts with the clapping machine, activates the system by inducing its closure (Figure 31, step 1). This conformational transition is transmitted through the rigid connector element to the rear foot, where it reduces DNA-binding affinity and causes forward displacement (brown arrow). The closure of the clapping machine is proposed to generate strain-dependent gating: mechanical strain destabilises the current rear-foot binding register while simultaneously shifting the probability distribution of the detached foot towards the next forward binding site. Subsequent rebinding of the rear foot at the next forward position (Figure 31, step 2) is proposed to result from strain-dependent modulation of the local binding kinetics. The strain generated by the conformational change of the clapping machine is expected to destabilise the DNA-bound state, thereby increasing the local dissociation rate and decreasing the probability of stable reassociation at the same site. Once the rear foot has detached, the same clapping machine conformation geometrically displaces the foot towards binding site 2. In this new register, the imposed strain is partially relaxed, which should lower the dissociation rate and increase the effective association probability for site 2. Productive forward stepping would therefore arise because clapping machine closure changes the relative on- and off-rates of the rear foot in a position-dependent manner: the original rear site becomes kinetically disfavoured, whereas the forward site becomes kinetically favoured. After catalytic turnover of the substrate (brown) to the product (orange and red) (Figure 31, step 3), and the dissociation of the products from the active site, the clapping machine returns to its open conformation (Figure 31, step 4). The reverse conformational change is proposed to be transmitted preferentially to the front foot due to asymmetry within the system, thereby destabilising its DNA-bound state and favouring its dissociation (brown arrow) and movement

towards the next binding site (binding site 4) (Figure 31, step 4). For autonomous motion, this reverse transition must not simply undo the preceding step. Instead, the motor must be designed such that reopening of the clapping machine again creates a new kinetic bias acting on the front foot. As in the first half-step, the combination of mechanically induced strain and its relaxation upon diffusion to the forward-bound binding site would bias reassociation toward that site (Figure 31, step 5). Following this mechanism, a motor would operate as a chemically driven Brownian ratchet. With biased Brownian motion originating from reshaping the conformational ensemble of the clapping machine, and the movements resulting from the change due to internal coupling between the feet and the clapping machine. This completes the second half-step and completes the directional inchworm-like translocation of the motor along the track. A critical prerequisite for autonomous stepping is the introduction of pronounced functional asymmetry into the motor architecture. One requirement identified for non-autonomous and autonomous motors is that simultaneous detachment of both feet from the DNA track must be avoided, as this abolishes processive forward motion. Rebinding of the detached foot must occur faster than the spontaneous dissociation of the remaining bound foot. Since finite off-rates can be defined for the bound foot, the time required for diffusion and binding the next binding site, and the residence time on the track are competing pathways for retaining the connection to the track. Furthermore, the motor must be designed so that one conformational transition of the clapping machine preferentially destabilises the DNA-bound state of the rear foot, while the complementary transition destabilises the front foot. This requires directional mechanosensitivity, meaning while both feet should respond to changes in the conformational state of the clapping machine, the actual conformational transition, from open to closed or from closed to open, is expected to affect each foot differently. Asymmetric extension of the PBP domain in only one monomer already introduces a first level of structural asymmetry within the foot design. However, for reliable autonomous stepping, a more pronounced asymmetry between the two feet, as well as an increased asymmetry within the clapping machine, is likely required. As the free energy and activation barriers need to vary between the system to enhance specific regulation through different conformational changes in the clapping machine. A further requirement for autonomy concerns the trigger of the conformational cycle itself. In the non-autonomous motor design, the clapping machine is closed by ligand binding. For autonomous motion, the externally imposed switching event must be replaced by an enzymatic reaction cycle. One possible strategy is to introduce an enzymatic function into the clapping machine, or to replace the current actuator with a naturally occurring protein domain that combines catalytic activity with a pronounced conformational transition between open and closed states. In this way, substrate binding and subsequent catalytic turnover could provide the driving force for repeated conformational cycling of the actuator, thereby coupling chemical energy consumption directly to mechanical stepping. Importantly,

catalytic turnover must do more than simply induce a conformational change. It must dissipate chemical free energy in a directional cycle, so that the motor does not relax back and forth between two equilibrium states. Only if substrate conversion and product release are kinetically coupled to distinct clapping-machine conformations can the system maintain a non-equilibrium flux throughout the stepping cycle. A crucial aspect of such a mechanism to work is the coordination of the individual substeps. Forward motion requires that DNA dissociation, diffusional exploration, and rebinding of the detached foot occur on a timescale compatible with the enzymatic cycle, which is the trigger to change the conformational ensemble of the clapping machine. After the shift towards the closed state of the clapping machine after substrate interaction, the released foot must detach and rebind to the next forward site before catalytic turnover restores the open state. If reopening occurs prematurely, the motor's efficiency will decrease because the catalytic turnover does not result in a defined change in the overall orientation of the clapping machine. Due to the rotational freedom of the clapping machine in the one-foot-bound state, a change in the conformation would not result in the required change of the DNA binding affinity and the position of the feet within the motor protein. For the forward movement to be preferred over backward transitions of off-track loss, competing rates of substrate binding, clapping machine closure and reopening, foot dissociation and association, tethered diffusion, catalytic turnover, and spontaneous detachment must be precisely controlled. The description of the mechanism highlights that various aspects of non-autonomous systems need to be improved to achieve an autonomous motor. As the different feet are no longer controlled by ligand binding, a pronounced asymmetry is required for specific responses to changes in the conformational ensemble of the clapping machine. The kinetic rates required for the individual steps must be coordinated to a larger extent, as no external stimuli can be varied to fit the kinetic parameters found within the system. Furthermore, an enzymatic reaction must be incorporated to ensure the independence of external control.

2.2 Generating asymmetry through asymmetric DNA binding with a heterodimeric LacI

2.2.1 *In vitro* analysis of engineered LacI heterodimer variants reveals impaired stability and insufficient heterodimer formation

The generation of a directional bipedal artificial motor protein requires asymmetric binding to the track (Chapter 2.1.1). When DNA is used as the track, and the repressor protein LacI serves as the track-binding part, the achievement of a heterodimeric repressor is essential to enable bipedal architecture and asymmetric track interaction.

Previous studies have demonstrated that modifying the dimer-dimer interface successfully generates heterodimeric LacI variants *in vivo*.[68] In the approach by Daber *et al.*[68], heterodimer formation relies on modifying specific interactions within the monomer-monomer interface in the C-terminal subdomain of the effector domain. One point mutation that disrupts native dimer formation is introduced into one subunit (subunit A), while compensatory mutations in the second subunit (subunit B) reconstitute the interaction network, forming a novel interface and a stable heterodimer (AB) (Figure 32).

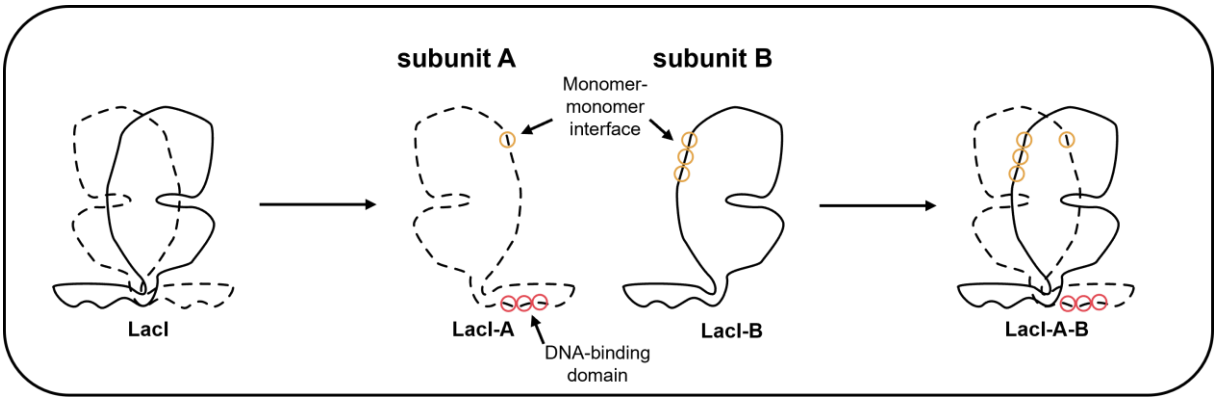


Figure 32. Schematic representation of LacI engineering towards heterodimeric variants. Starting from homodimeric LacI, subunits A (dotted line) and B (solid line) were modified by introducing one or more point mutations into the monomer-monomer interface (orange highlights). The DNA-binding domain of subunit A was additionally modified to enable recognition of an asymmetric operator sequence (red highlights). Together, these modifications were intended to favour the formation of LacI A-B heterodimer species.

In this thesis, the best variants previously described by Daber *et al.*[68], showing less than 5% homodimer formation and more than 100% heterodimer formation *in vivo*, were investigated. The introduction of mutations targeting the monomer-monomer interface alone is, however, insufficient to confer asymmetric track binding. To overcome this limitation, a second part of the approach described by Daber *et al.*[68] includes the introduction of asymmetry at the DNA-binding level. In their study, an asymmetric DNA-binding site was employed to enable selective recognition of heterodimeric LacI variants on correspondingly asymmetric operator sequences. This strategy was likewise implemented in the present work. Specifically, subunit A and subunit B differ in their DNA-binding domains, thereby altering their sequence specificity. Three key residues (positions 17, 18, and 22), which are well established as critical determinants of

specific operator recognition, were substituted in subunit A (red highlight, Figure 32) with the corresponding amino acids found in the GalR repressor (DBGal, Table 2). These modifications enable one monomer to selectively recognise the *galO* half-site within a hybrid operator sequence. The subunit (subunit B), which was not modified in its DNA-binding domain, still recognises the *lacO* half-site. This approach introduces functional asymmetry at the DNA-binding level. A complete overview of all constructs and their respective mutations in the monomer-monomer interface and DNA binding domain is provided in Table 2. In the chosen naming scheme, the designation A refers to the subunit carrying additional point mutations in the DNA-binding domain, whereas the number assigned to A or B indicates the respective mutation set within the monomer-monomer interface.

Table 2. Overview of LacI variants with modified monomer-monomer interface analysed for heterodimer formation. Mutations introduced to alter DNA-binding specificity and to modulate monomer-monomer interface interactions are indicated for each variant.

Name	Specification	Mutations within	
		DNA-binding domain	Monomer-monomer interface
LacI-A1	LacI-DBGal-A1	Y17T, Q18A, R22N	Y282A
LacI-B1	LacI-DBLac-B1	-	L251M, D278N, Y282F
LacI-A2	LacI-DBGal-A2	Y17T, Q18A, R22N	Y282S
LacI-B2	LacI-DBLac-B2	-	L251K, R255W, D278T, C281Q, Y282W
LacI-B3	LacI-DBGal-B3	-	L251K, R255K, D278S, C281Q, Y282W

The required variants were generated through iterative rounds of site-directed mutagenesis starting from LacI present in the pCDF-BAD vector. Consistent with the subunit A and B combinations identified by Daber *et al*.[68] to form heterodimers *in vivo*, heterodimer formation of LacI-A1-B1, LacI-A2-B2 and LacI-A2-B3 was investigated *in vitro*. To mimic *in vivo* conditions, subunit A and B were heterologously co-expressed in *Escherichia coli*.

First, I will focus on the potential heterodimer LacI-A1-B1. Both variants (LacI-A1 and LacI-B1) carried identical purification tags. Using the two-step purification protocol comprising nickel-immobilised metal affinity chromatography (Ni^{2+}-IMAC) and size exclusion chromatography (SEC), it is not possible to clearly prove heterodimer formation by the purification alone. Nonetheless, first indications of the correct oligomeric state were observed. Both subunits (A and B) were found in the soluble fraction, and compared to individual heterologous expression of the variants, the yield, particularly of LacI-B1, was increased during co-expression (SI Figure 93 (coexpression) and Figure 88-Figure 89 (individual)). In an ideal case, both proteins would coelute in equimolar amounts. Two distinct bands at 36 and 40 kDa during SDS-PAGE analysis of the elution fractions from Ni^{2+}-IMAC confirmed the presence of both proteins. A changing relative abundance of LacI-A1 (green) and LacI-B1 (pink) across the elution profile was observed (Figure 33A). However, the intensity of the bands did not change in parallel across

the collected fractions. LacI-B1 was enriched in earlier elution fractions, whereas LacI-A1 was distributed more uniformly across the elution fractions. Although the two proteins do not necessarily have to be expressed at equimolar levels, formation of a defined 1:1 heterodimer would be expected to result in comparable amounts of both subunits in the SDS-PAGE analysis, at least for some fractions. The SDS-PAGE results indicate that heterodimers were not the sole species present. Additional analysis is therefore required to resolve the oligomeric state of the variant LacI-A1-B1 *in vitro*.

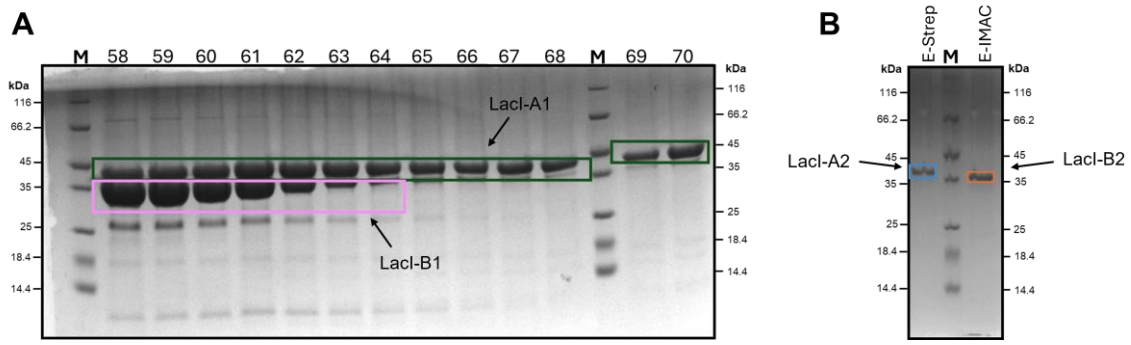


Figure 33. SDS-PAGE analysis of co-expressed LacI interface variants. A) Ni^{2+}-IMAC elution fractions after co-expression of LacI-A1 and LacI-B1 analysed with SDS-PAGE. Bands at 40 kDa and 36 kDa correspond to LacI-A1 (green box) and LacI-B1 (pink box), respectively, with LacI-B1 enriched in earlier fractions and LacI-A1 detected over a broader elution range. Fraction numbers correspond to the elution profile shown in SI Figure 93. B) SDS-PAGE analysis of fractions from different purification steps after co-expression of LacI-A2 and LacI-B2. The highlighted bands corresponding to LacI-A2 (37 kDa, blue box) and LacI-B2 (36 kDa, orange box) are separated by Ni^{2+}-IMAC and Strep-Tactin AC, suggesting that no stable co-purifying LacI-A2/B2 complex was formed under the tested conditions.

For the remaining potential heterodimeric pairs, purification was optimised to improve the analysis of heterodimer formation during the purification. Variant LacI-A2 carries a Strep-II tag, whereas LacI-B2 and LacI-B3 contain a His_6-tag, allowing the detection of potential heterodimer formation after heterologous co-expression in *Escherichia coli*. If protein-protein interactions occurred between the two variants, both variants would co-elute during both affinity chromatography steps. In the following, the use of both purification tags to analyse heterodimer formation is referred to as pull-down purification.

SDS-PAGE analysis highlights no simultaneous elution of subunit A and B for either LacI-A2-B2 or LacI-A2-B3 after coexpression (SI Figure 94 and Figure 95). Instead, the Strep-II-tagged LacI-A2 (Figure 33B, blue box) was found in the flow-through of the Ni^{2+}-IMAC and was recovered by purifying the flow-through with Strep-Tactin affinity chromatography (Strep-Tactin-AC) (Figure 33B, E-Strep). The variants LacI-B2 (Figure 33B, E-IMAC) and LacI-B3 were identified in the elution fractions of the Ni^{2+}-IMAC (SI Figure 95)

The pull-down purification experiments (Figure 33B, SI Figure 94 and Figure 95) clearly show that no heterodimer formation occurs. To further investigate potential issues with the variants that caused the unexpected behaviour, the variants (LacI-A2, LacI-B2, and LacI-B3) were

individually expressed heterologously in *Escherichia coli.* All variants were found in the soluble fraction. Purification by Ni^{2+}-IMAC or Strep-Tactin-AC followed by SEC yielded sufficiently pure proteins (SI Figure 90-Figure 92). The SEC chromatograms revealed that variants LacI-B2 and LacI-B3 exhibited a shoulder on the right side of the main peak (SI Figure 91 and Figure 92, red box). SDS-PAGE analysis across the peak fractions showed proteins of identical molecular weight (SI Figure 91 and Figure 92), indicating that the different elution profiles were caused by the presence of different oligomeric species. For subsequent analyses, fractions corresponding to the shoulder and the main peak were pooled before concentration to avoid experimental bias in deeper analysis of the variants' different oligomeric states using SEC-MALS.

The oligomeric states of the individually expressed and co-expressed variants were analysed by SEC-MALS (Figure 34). In contrast to using elution volume alone as an estimate of molecular weight, multi-angle light scattering (MALS) enables molecular-weight determination independent of hydrodynamic radius or interactions with the column matrix.

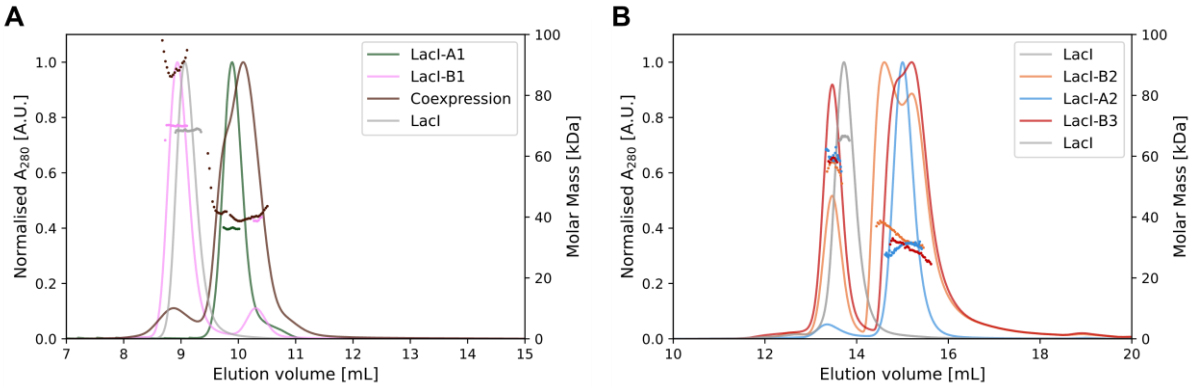


Figure 34. SEC-MALS analysis of variants with modified monomer-monomer interfaces. The normalised absorbance at 280 nm is plotted against elution volume, and the corresponding molecular weight is shown on the secondary y-axis. The wildtype LacI (LacI, grey, 36 kDa) is included as a reference. Measurements were performed in Buffer C using a protein concentration of 0.5 mg·mL^{-1}. A) Measurements for LacI-A1 (green, 40 kDa), LacI-B1 (pink, 36 kDa), and the co-expressed sample (brown, 77.2 kDa) were performed using a Superdex 75 Increase 10/300 GL column. B) Measurements for LacI-A2 (blue, 37 kDa), LacI-B2 (orange, 36 kDa), and LacI-B3 (red, 36 kDa) were performed using a Superdex 200 Increase 10/300 GL column.

SEC-MALS analysis of the co-expressed LacI-A1-B1 sample (brown, Figure 34A) showed two distinct peaks. Molecular weights of 93 kDa and 39 kDa are calculated for each peak. The peak eluting at 10 mL was assigned to monomeric species, likely corresponding to a mixture of LacI-A1 and LacI-B1. In contrast, the molecular weight determined for the earlier-eluting peak exceeded the expected mass of a dimeric species by around 20 kDa (SI Table 42). This overestimation is likely due, at least in part, to the low protein concentration in the corresponding peak, which limits the accuracy of the MALS measurement. Owing to this substantial deviation, the data do not permit an unambiguous distinction between between homo- and heterodimeric assemblies, in the co-expressed sample. Analysis of the individually

expressed subunits showed that LacI-A1 (green, Figure 34A) contains a single peak, for which the apparent molecular weight was consistent with a monomeric species. In contrast, LacI-B1 (pink, Figure 34A) displayed two peaks, with the major peak eluting at 9 mL and a minor peak at around 10.5 mL. Based on the calculated molecular weights of the respective peaks (70 kDa and 40.6 kDa), LacI-B1 was concluded to exist predominantly as a dimeric species, with a smaller fraction of monomeric protein. While these observations do not formally rule out heterodimer formation in the presence of LacI-A1, they indicate that LacI-B1 retains a substantial intrinsic propensity for homodimerisation. It is important to note that this behaviour may be accentuated under single-expression conditions, since subunit A1 is absent and therefore cannot compete for the monomer-monomer interface. A clear difference between co-expression and individual expression was observed. The co-expressed sample was predominantly monomeric, while the individually expressed LacI-B1 exhibited a substantially larger dimeric fraction. The discrepancy indicates that co-expression influences the oligomeric state of the sample. Even so, the co-expression data likewise failed to demonstrate efficient formation of a defined heterodimeric species. Collectively, the purification and SEC-MALS results indicate that the A1/B1 subunit combination does not provide the degree of selective heterodimer formation required for the intended motor-protein architecture.

For LacI-A2-B2 and LacI-A2-B3, the pull-down purification already showed that no heterodimer formation was present after coexpression. For the individually expressed variants, a trend similar to that observed for LacI-A1 and LacI-B1 was observed for the remaining variants (LacI-A2, LacI-B2, and LacI-B3). Individually expressed LacI-A2 (Figure 34B, blue) showed a predominantly monodisperse peak with a calculated molecular weight (36 kDa) corresponding to the size of a monomer (37 kDa). In contrast to LacI-A1, the Y282S substitution in LacI-A2 disrupted the monomer-monomer interface less effectively than the Y282A mutation introduced in LacI-A1. This is reflected by the presence of a small second peak at an elution volume of 13 mL, for which the apparent molecular weight is consistent with a dimeric species (SI Table 42). The B variants, LacI-B2 (Figure 34B, orange) and LacI-B3 (Figure 34B, red), showed two peaks corresponding to monomeric and dimeric species. However, unlike LacI-B1, in which the dimeric fraction predominated, LacI-B2 and LacI-B3 showed two peaks eluting at 13.5 mL and 15 mL, with the later-eluting species at 15 mL accounting for the major absorbance signal. The corresponding apparent molecular weights of around 73 kDa and 37 kDa identify these as dimeric and monomeric species. It is noteworthy that the second peaks of LacI-B2 and LacI-B3 were not monodisperse, and that the molecular weight decreased across the peaks. Comparison with the purification fractions indicated that this behaviour resulted from slight contamination by a 25 kDa large protein contamination (SI Figure 91-Figure 92). Nevertheless, it is evident that even the introduction of five mutations into subunits of B2 and B3, at positions 251, 255, 278, 281, and 282, still allows the formation of a homodimeric protein. Together with

the pull-down purification assay, it was shown that a novel interface that enables heterodimer formation with the chosen monomeric LacI variant (LacI-A2) was not established. Nevertheless, comparison of the B variants indicates that the interface mutations in LacI-B2 and LacI-B3 partially reduced the intrinsic homodimerisation propensity relative to LacI-B1. Whereas LacI-B1 was predominantly dimeric, LacI-B2 and LacI-B3 exhibited a more balanced distribution of monomeric and dimeric species, approaching an approximately 1:1 and 1:2 dimer to monomer ratio, respectively. Although this shift was insufficient to promote efficient heterodimer formation with LacI-A2, it suggests that the introduced mutations modulate the oligomeric equilibrium. The persistence of the homodimeric oligomerisation state in subunit B suggests that the specificity of the engineered interfaces remained insufficient, thereby representing a major bottleneck for the current variants.

Although further optimisation of the current variants with additional modifications to the interface remains possible to generate heterodimeric variants, caution is required, as even a limited number of mutations significantly affects the thermal stability. Thermal stability was analysed by circular dichroism spectroscopy, monitoring thermal unfolding at 222 nm to follow the unfolding of α-helices (Figure 35), thereby allowing determination of melting temperatures (T_{Melt}) (Table 3).

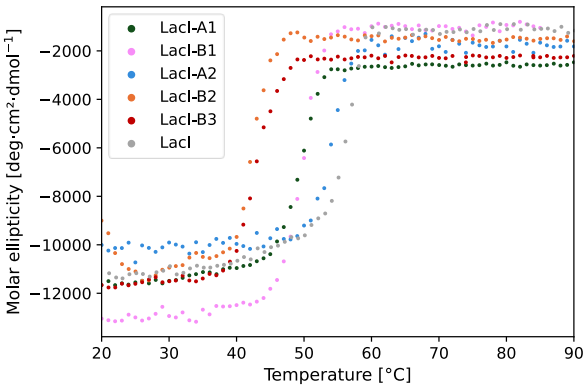


Figure 35. Temperature-dependent changes in molar ellipticity of the LacI variants with modified monomer-monomer interface CD thermal unfolding curves of LacI-A1 (green), LacI-B1 (pink), LacI-A2 (blue), LacI-B2 (orange), LacI-B3 (red), and wild-type LacI (grey) were recorded by monitoring the molar ellipticity at 222 nm. Measurements were performed in Buffer C using a protein concentration of 0.2 mg·mL^{-1}.

Table 3. Apparent melting temperature (T_{Melt}) of the LacI variants with varied monomer-monomer interface. Values were determined from CD-based thermal unfolding experiments as the transition midpoint of the temperature-dependent change in molar ellipticity at 222 nm.

Construct	T_{Melt} [°C]
LacI-A1	49.2
LacI-B1	49.6
LacI-A2	53.9
LacI-B2	42.2
LacI-B3	42.7
LacI	55.8

All variants still showed cooperative unfolding (Figure 35), but the extent of the effect on thermal stability varied (Table 3). The Y282S mutation (LacI-A2) had only a minor effect (2 °C) on stability. A substitution by alanine at position 282 (Y282A, LacI-A1) reduced the melting temperature by more than 6 °C. Variants LacI-B1, LacI-B2 and LacI-B3 showed an even stronger destabilisation, with decreases in the thermal stability by up to 13 °C. Comparing the influence of the thermal stability of LacI-A1 and LacI-A2 showed that the stability is not only

influenced by the position where the substitution is introduced, but also by the nature of the amino acid. Among the B variants, the degree of destabilisation appeared to increase with the number of mutations, since LacI-B1, containing three point mutations, was less destabilised than LacI-B2 and LacI-B3, which each harbour five point mutations. However, this is not generally valid; only one mutation can also exhibit the same destabilisation effect as three mutations, as seen for LacI-A1 and LacI-B1. Taken together, this data shows that the monomer-monomer interface can be modified, but only within a narrow stability window. Further optimisation towards heterodimer formation should therefore carefully balance interface redesign against preservation of the overall thermal stability of the LacI scaffold.

Despite the substantial impact of the interface mutations on thermostability, ligand binding remains largely unaffected. Ligand binding was monitored by changes in intrinsic tryptophan fluorescence. A tryptophan residue (W229) located in the binding pocket of LacI senses the environmental changes upon ligand binding, leading to a change in the emission spectrum after excitation at 285 nm.[432] The addition of the ligand IPTG induces a blue shift of the emission spectrum (Figure 36A and B, SI Figure 134).

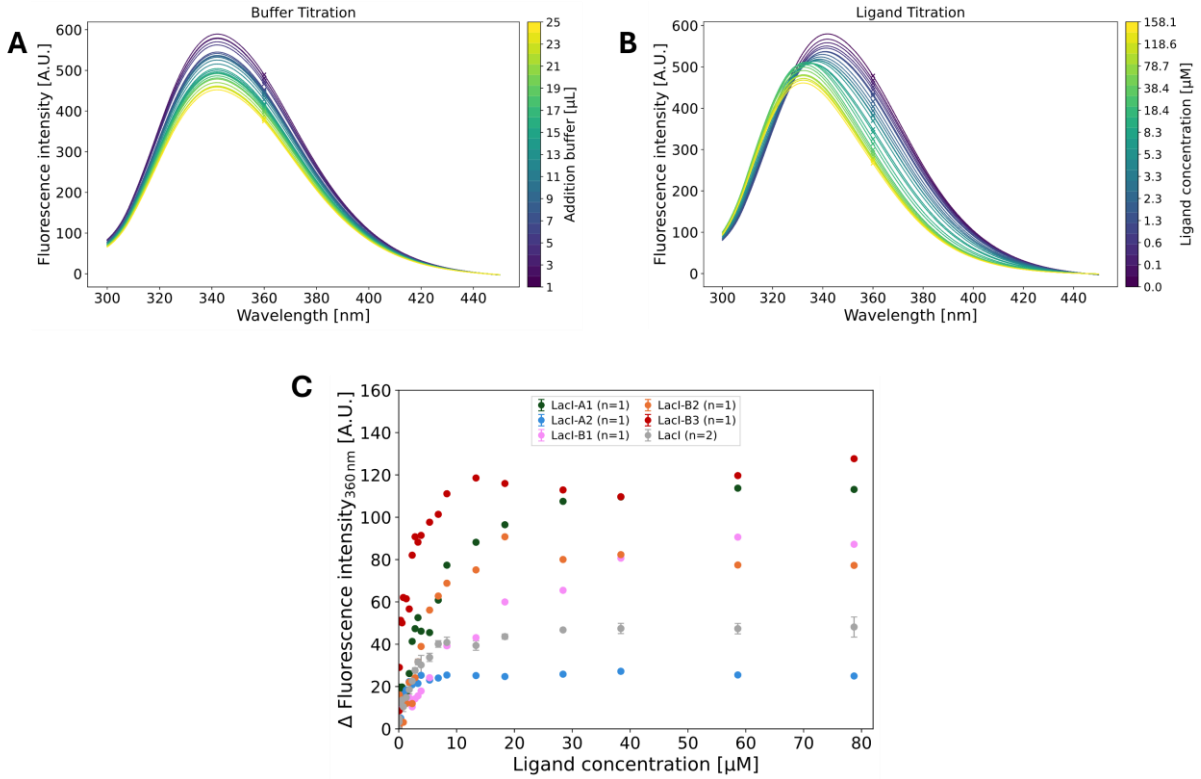


Figure 36. Fluorescence-based analysis of IPTG binding to LacI variants. A) Fluorescence emission spectra of LacI-A1 recorded after sequential buffer additions as a control. B) Fluorescence emission spectra of LacI-A1 recorded in the presence of increasing IPTG concentrations, as indicated by the colour scale. C) Change in fluorescence intensity at 360 nm plotted against IPTG concentration for LacI (grey), LacI-A1 (green), LacI-A2 (blue), LacI-B1 (pink), LacI-B2 (orange), and LacI-B3 (red). Measurements were performed in Buffer C using a protein concentration of 0.2 mg·mL^{-1}.

In contrast, a mutation of a critical residue for ligand interaction, arginine 197 to glycine, abolished the blue shift upon IPTG addition (SI Figure 133), confirming the specificity of the assay.[433] Although the number of biological replicates was insufficient for a reliable determination of binding affinity, a qualitative comparison of the change in fluorescence intensity at 360 nm suggests that ligand binding occurred in a concentration range comparable to wild-type LacI, as the pronounced increase in fluorescence was observed at similar ligand concentrations (Figure 36C). The variants differ in the total magnitude of their fluorescence change. Although an approximate concentration of 0.2 $mg \cdot mL^{-1}$ was used for all measurements, the exact protein concentration differed between variants. The fluorescence signal was not normalised to these concentration differences, resulting in absolute changes of the fluorescence intensity.

Analysing the proposed heterodimers LacI-A1-B1, LacI-A2-B2 and LacI-A2-B3 from Daber *et al.*[68] showed that while the variants were previously identified as heterodimers *in vivo*, the oligomeric state is not supported by *in vitro* experiments. However, the incorporated mutations at the monomer-monomer interface affect dimer formation and protein stability, whereas IPTG binding remains largely unaffected, consistent with the spatial separation of the regions responsible for ligand binding and dimer formation within LacI.

2.2.2 Assessment of a splitGFP-based BiFC system for analysis of LacI oligomerisation states

Protein-protein interaction-driven reconstitution of split reporter systems, such as splitGFP, represents a powerful approach for analysing heterodimer formation. Based on this concept, an experimental setup analogous to a bimolecular fluorescence complementation (BiFC) assay was designed. In this system, a bipartite splitGFP construct, generated by splitting GFP between β-strands 9 and 10, was fused to LacI subunits with the goal of analysing heterodimer assembly by fluorescence-based detection of reconstitution of the split reporter system (Figure 37). To determine whether this system is suitable for distinguishing between heterodimer, homodimer, and monomer formation, several control constructs were first examined *in vivo*. To assess whether spontaneous GFP reassembly could generate a fluorescent signal in the absence of dimerisation, the splitGFP system was fused to LacI monomers. Monomerisation was achieved by introducing the Y282A mutation. In a previous chapter (2.2.1), the variant LacI-A1, which was shown to be monomeric. However, in that context, the designation LacI-A1 referred not only to the Y282A substitution but also to additional mutations in the DNA-binding domain introduced for asymmetric DNA recognition. The additional mutations in the DNA-binding domain were not present in the construct A1-splitGFP used. For this variant, both subunits retained the native LacI DNA-binding domain but carried the Y282A mutation in the interface, as well as each one part of the splitGFP system (Figure 37A).

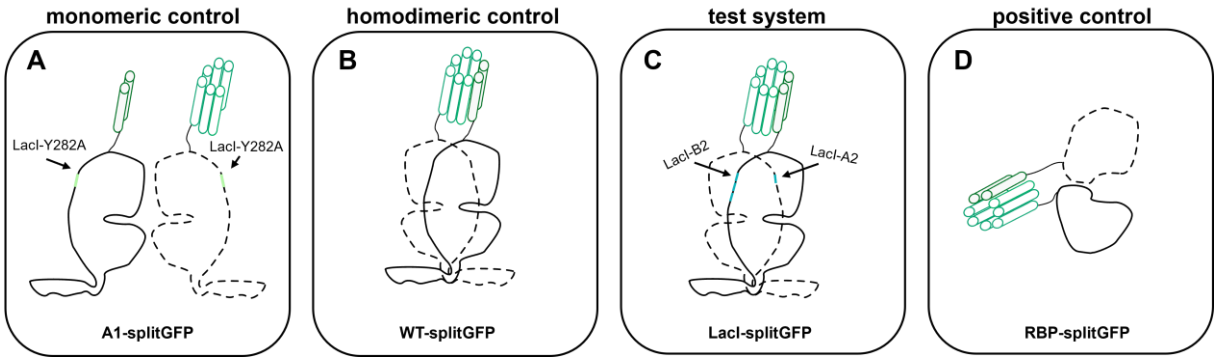


Figure 37. Control constructs used to evaluate splitGFP complementation as an *in vivo* readout for LacI heterodimer formation. A) A1-splitGFP, in which both splitGFP fragments were fused to the monomeric LacI-A1 variant, was used to test whether GFP reassembly occurs in the absence of LacI dimerisation. B) WT-splitGFP, in which the splitGFP system was fused to wild-type LacI, assessed whether a homodimeric LacI scaffold is sufficient to promote functional GFP reassembly. C) LacI-splitGFP, comprising LacI-A2 and LacI-B2, was included to compare *in vivo* splitGFP complementation with the heterodimerisation behaviour observed *in vitro*. D) RBP-splitGFP, an in-house heterodimeric construct based on ribose-binding protein, served as a positive control for heterodimer-dependent GFP complementation.

To test whether the expression of the split system attached to homodimeric LacI would promote the formation of a functional GFP, the split system was fused to the wild-type LacI (WT-splitGFP) (Figure 37B). As a functional benchmark, the splitGFP system was attached to LacI-A2 and LacI-B2 (LacI-splitGFP) (Figure 37C), which, in other *in vivo* assays, was identified as heterodimers,[68] although this assignment is not supported by the presented *in vitro* assays (see Chapter 1.1.1.4). This benchmark was intended to assess whether the BiFC assay might provide improved selectivity compared with the transcription-based assay used by Daber *et al.*[68] and whether the signal might reveal intermediate GFP maturation values between homodimers and heterodimers. In addition, a heterodimeric positive control was included. For this purpose, an in-house heterodimer based on ribose-binding protein RbsB (RBP-splitGFP) was used (Figure 37D). Michel *et al.*[434] showed that co-expression of separated N- and C-terminal subdomains of the ribose-binding protein reassembled into a stable and functional protein. This assembly would bring the splitGFP moieties into proximity, thereby generating a strong fluorescent signal. In contrast, for the other variants, fluorescence was expected to be absent (A1-splitGFP) or reduced because of statistical homodimeric interactions (WT-splitGFP). The different variants were generated by RE-based cloning into the pDUET^{M} vector, which contains two reading frames with the respective parts of the splitGFP positioned downstream of the selected restriction sites. Each subunit of the proposed constructs was incorporated into either the first or second expression cassette, enabling the specific fusion with splitGFP strand 1-9 or splitGFP strand 10-11 to the respective subunit.

Before using the BiFC assay in an *in vivo* screening to identify monomer-monomer interface mutations that promote heterodimer formation without homodimer formation, the controls described above were first evaluated following heterologous expression in *Escherichia coli*. To enable evaluation of heterodimer formation with the pull-down assay, different affinity tags were

introduced. One subunit (solid line, Figure 37), which was fused to GFP strands 10-11, carried a Strep-II tag, whereas the second subunit (dashed line, Figure 37), fused to GFP strands 1-9, carried a His_6-tag. Not all constructs exhibited favourable behaviour, as SDS-PAGE (Figure 38) and affinity chromatography revealed substantial differences among the constructs (SI Figure 96).

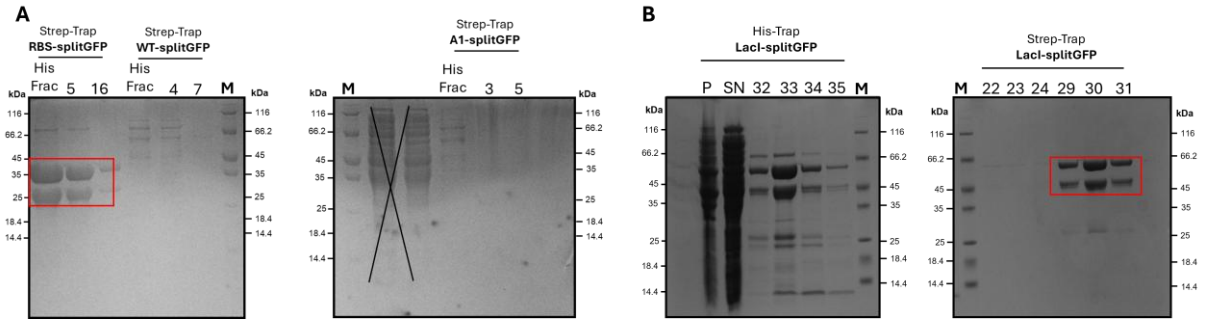


Figure 38. SDS-PAGE analyses of splitGFP fusion constructs. A) SDS-PAGE overview of elution fractions obtained after Strep-Tactin AC for RBP-splitGFP, WT-splitGFP, and LacI-splitGFP. The corresponding chromatograms are shown in Figure 96 B) SDS-PAGE analysis of LacI-splitGFP, highlighting elution fractions from Ni^{2+}-IMAC and Strep-Tactin affinity chromatography. The corresponding Ni^{2+}-IMAC chromatogram is shown in Figure 97. Red boxes highlight bands corresponding to successfully expressed splitGFP fusion proteins.

Neither the construct containing WT-splitGFP nor A1-splitGFP showed clear overexpression, and no soluble protein fraction was obtained for either construct, as evidenced by the absence of protein retained and recovered using Strep-Tactin AC (Figure 38A). All constructs were checked for sequence errors. However, no mutations or cloning artefacts were detected. In contrast, overexpression and purification of both the ribose-binding protein positive control (RBS-splitGFP) and the benchmark control (LacI-splitGFP) were successful, and sufficient amounts of protein were obtained in the soluble fraction after Strep-Tactin AC (Figure 38, red box). These differences in expression raised the question of whether soluble expression itself already reflects a degree of selection for heterodimer formation, as the monomeric and homodimeric controls did not show soluble expression, whereas the heterodimeric control and the benchmark construct, which had previously been classified as heterodimeric in another *in vivo* assay,[68] were both expressed and soluble. However, this conclusion could not be further validated, as no additional heterodimeric or homodimeric controls were included.

To evaluate whether splitGFP expression was accompanied by successful chromophore maturation, the GFP fluorescence of the RBS-splitGFP constructs and the LacI-splitGFP control was measured. The OD-normalised fluorescence differed between the constructs. Since the fluorescence values were normalised to cell density, differences in signal intensity are likely to reflect differences in splitGFP complementation efficiency and, consequently, in the formation or stabilisation of the corresponding heterodimeric LacI species (SI Figure 167). Notably, the benchmark control LacI-splitGFP still produced a fluorescence signal above the level expected in the complete absence of heterodimer formation. However, the basal level of

GFP maturation in the absence of a productive interaction could not be directly assessed, as the intended monomeric control A1-splitGFP showed no detectable expression. Nevertheless, the fluorescence signal suggests that the splitGFP assay may not only report on pre-existing LacI heterodimerisation but could also contribute to stabilising the complemented complex. Therefore, before using the assay to evaluate additional point mutations within the monomer-monomer interface, it is useful to assess whether splitGFP complementation itself influences heterodimer formation or stabilisation.

2.2.3 SplitGFP promotes LacI assembly but does not yield a stable heterodimeric complex

The results highlighted in Chapter 2.2.2 confirmed a pronounced interaction between both splitGFP parts of the LacI-splitGFP construct. This, however, prompted a further question regarding the role of splitGFP within the system: rather than serving solely as a passive reporter of heterodimer formation, splitGFP reassembly might itself stabilise or even promote heterodimeric complex formation. To address this possibility, subsequent analyses focused on the benchmark construct (LacI-splitGFP) to determine whether the reporter system could bias the oligomerisation equilibrium towards a heterodimeric assembly.

To test whether both splitGFP fusion partners remained associated during purification, a pull-down strategy using their distinct affinity tags was applied. In the first step, Ni^{2+}-IMAC was used to selectively purify LacI-B2-splitGFP1-9 via its His_6-tag. SDS-PAGE analysis of the elution fractions showed that not only the His-tagged construct, but also the complementary Strep-II-tagged construct LacI-A2-splitGFP10-11, was present in the eluate (Figure 38B, His-Trap). To confirm this co-purification, a second affinity purification step was performed using Strep-Tactin AC, which specifically selects for LacI-A2-splitGFP10-11. Again, SDS-PAGE analysis revealed the presence of both fusion partners in the elution fractions (Figure 38B, red box). These results demonstrate that the two constructs can be co-purified as a heteromeric assembly. However, further experiments were required to determine whether this assembly indeed corresponded to a defined heterodimer rather than another oligomerisation state.

To further test whether the splitGFP induced sufficiently stable heterodimer formation such that LacI-A2 and LacI-B2 might remain associated even after removal of the splitGFP trigger, the purification tags as well as the splitGFP moieties were removed by TEV protease. Separation of cleaved and uncleaved species was ensured by reverse Ni^{2+}-IMAC as an additional purification step (SI Figure 97). SEC revealed that, after cleavage of splitGFP, the sample no longer behaves as a single species but appears to distribute into different oligomeric states (Figure 39A). Fractions corresponding to the earlier-eluting peak were assigned as LacI-Dimer-GFPcleaved (purple box), whereas fractions from the later-eluting peak were assigned as LacI-Monomer-GFPcleaved (dark grey box).

To validate the heterodimeric oligomerisation state expected for LacI-splitGFP and to identify if the oligomeric state is retained after the cleavage of the splitGFP parts, SEC-MALS was performed as it gives a reliable estimate of the oligomeric state (Figure 39B).

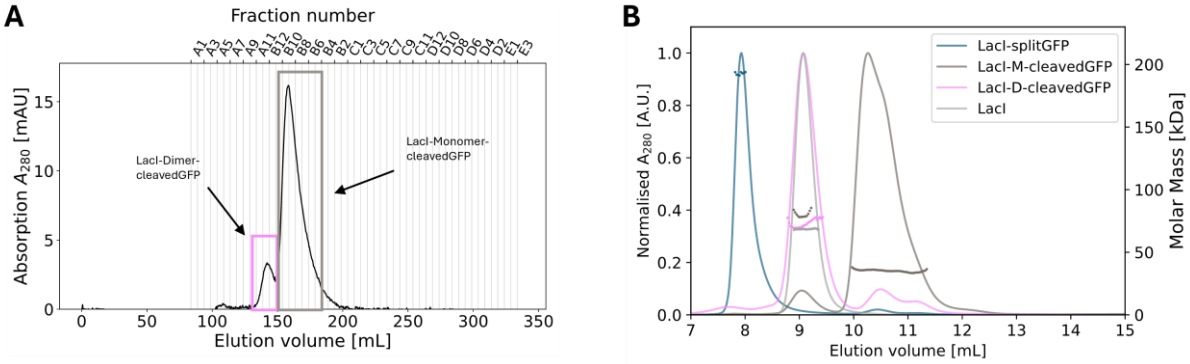


Figure 39. SEC and SEC-MALS analysis of cleaved and uncleaved LacI-splitGFP variants. A) Preparative SEC performed after proteolytic removal of the splitGFP domains from LacI-splitGFP using a HiLoad 26/600 Superdex 75 pg. Based on their elution volumes, the separated fractions were assigned to LacI-Dimer-cleavedGFP (pink) and LacI-Monomer-cleavedGFP (grey) species. B) SEC-MALS analysis performed using an analytical Superdex 200 Increase 10/300 GL column. Uncleaved LacI-splitGFP (blue), the corresponding cleaved fractions LacI-M-cleavedGFP (dark grey) and LacI-D-cleavedGFP (pink), and wild-type LacI (light grey) were analysed. The expected molecular weights are 106 kDa, 36 kDa, and 73 kDa for uncleaved LacI-splitGFP, LacI-M-cleavedGFP, and LacI-D-cleavedGFP, respectively. The normalised absorbance at 280 nm is plotted against elution volume, while the calculated molar mass is shown on the secondary axis. Measurements were performed in Buffer C using a protein concentration of 0.5 mg·mL^{-1}.

SEC-MALS data for uncleaved LacI-splitGFP (blue) showed a single major peak eluting close to the void volume. The corresponding molecular mass (SI Table 42) was substantially higher than the 106 kDa expected for a 1:1 heterodimeric LacI-A2/LacI-B2 splitGFP complex. Instead, an apparent molecular mass of around 190 kDa was determined. The exact composition of the parts resulting in a molecular weight of 190 kDa cannot be determined unambiguously from this data alone. However, the data suggest the formation of a higher-order assembly instead of a simple heterodimeric species. Based on the previous observation in Chapter 2.2.1, in which it was shown that LacI-B2 retains a pronounced tendency to form homodimers, one plausible interpretation is that splitGFP-mediated proximity enables interaction between LacI-A2 and LacI-B2, while LacI-B2 simultaneously forms homodimeric contacts. In such a scenario, a complex containing a LacI-A2 monomer and a LacI-B2 homodimer may be formed. The expected molecular weight of this construct would match the determined molecular weight. The two fractions of the splitGFP-cleaved sample were analysed regarding their oligomeric state (Figure 39B). The main peak of LacI-Monomer-cleavedGFP (Figure 39B, dark grey) displayed an apparent molecular mass of about 35.5 kDa, consistent with a monomeric species. In addition, a second, minor peak with an apparent molecular mass of 82.0 kDa was observed. Although this value is somewhat higher than the expected mass of a true dimer (~73 kDa), the low signal intensity of this peak may have reduced the accuracy of the mass

determination. Because this peak co-eluted with the dimeric peak observed for LacI-Dimer-cleavedGFP (Figure 39B, pink), it can be assigned to a dimeric species. A complementary pattern was observed for LacI-Dimer-cleavedGFP (Figure 39B, pink). In this case, the predominant species was dimeric, with an apparent molecular mass of 73.3 kDa, whereas only a minor monomeric contribution was detected. Two possible explanations are conceivable. First, the presence of both monomers and dimers in the peaks separated during purification may reflect a true monomer-dimer equilibrium. Alternatively, the apparent peak separation may arise from differences in the separation properties of the Superdex 75 and Superdex 200 column material. The purification was performed on a Superdex 75 column, while during SEC-MALS, a Superdex 200 column was used. The Superdex 200 column material provides improved separation in the relevant molecular-weight range, thereby separating only oligomeric species which were incompletely separated during the purification. A further limitation of the data is that no unambiguous assignment of the individual contributions of LacI-A2 and LacI-B2 to the observed monomeric and dimeric fractions of LacI-cleavedGFP is possible. However, when considering previous results (2.2.1), the peak assigned to the monomer (10.5 mL) is most plausibly interpreted as being a mixture of LacI-A2 as well as LacI-B2, whereas the peak assigned to the dimers (9 mL) appears to consist mostly of the dimeric species of LacI-B2. This interpretation is supported by thermal unfolding monitored by circular dichroism spectroscopy (Figure 40).

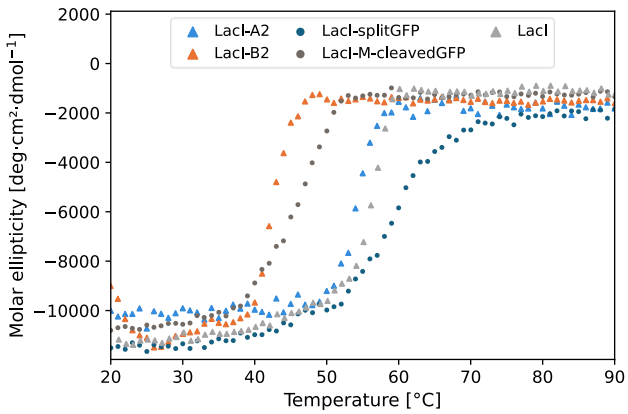


Figure 40. CD-based thermal unfolding analysis of LacI-splitGFP variants. Thermal unfolding of cleaved LacI-M-cleavedGFP (dark grey), uncleaved LacI-splitGFP (blue), and the reference variants wild-type LacI (light grey triangle), LacI-A2 (blue triangle), and LacI-B2 (orange triangle) was monitored by the temperature-dependent change in molar ellipticity at 222 nm. Measurements were performed in Buffer C using a protein concentration of 0.2 mg·mL^{-1}.

Table 4. Apparent melting temperatures (T_{Melt}) of LacI-splitGFP-derived fractions and reference LacI variants. Values were determined from CD-based thermal unfolding experiments as the transition midpoint of the temperature-dependent change in molar ellipticity at 222 nm.

Construct	T_{Melt} [°C]
LacI-M-cleavedGFP	45.4
LacI-splitGFP	58.6
LacI-A2	53.9
LacI-B2	42.2
LacI	55.8

The melting temperature of the monomeric fraction (Table 4, LacI-M-cleavedGFP) does not match the apparent thermal stability observed for either isolated LacI-A2 or isolated LacI-B2 but instead lies between the two values (Table 4). In addition, the unfolding transition of LacI-

M-cleavedGFP (Figure 40, dark grey) was less steep than unfolding curves of LacI-A2, LacI-B2 and LacI (Figure 40, blue, orange and grey), suggesting reduced cooperativity and thus supporting the presence of more than one species within this fraction. In contrast, the uncleaved variant (Figure 40, turquoise) exhibited increased thermal stability, with a melting temperature of 58.6 °C, even exceeding that of wild-type LacI (light grey). It seems that the mutations, which, in the isolated variants (LacI-A2 and LacI-B2), lower the melting temperature by 2 °C and 13 °C, respectively, do not have the same influence in the presence of splitGFP, or the intrinsic stability of splitGFP may counterbalance the destabilising effects introduced by the mutations.

Taken together, the SEC-MALS and thermal stability data indicate that the LacI-Monomer-cleavedGFP fraction contains a mixture of LacI-A2 and LacI-B2. By contrast, the observed separation of monomeric and dimeric species, together with the previous SEC-MALS results discussed in Chapter 2.2.1, suggests that the LacI-Dimer-cleavedGFP fraction consists predominantly of a homodimeric species. It was clearly shown that without removing the splitGFP, higher-order assemblies likely arise with elevated thermal stability.

To further characterise these assemblies, SPR experiments were performed to investigate binding to a symmetric and asymmetric DNA substrate. Interactions of the proteins were tested with a symmetric DNA sequence consisting only of *lacO* sequence and an asymmetric DNA sequence consisting of *lacO* and *galO* half sites. As explained in detail in Chapter 2.2.1, DNA-binding domains of the two subunits (LacI-A2 and LacI-B2) attached to the splitGFP system to form a potential LacI heterodimer vary in 17, 18, and 22, which are positions known to be crucial for the selective binding of DNA (see Chapter 1.1.1.2).

The LacI-Dimer-splitGFP variant, composed predominantly of LacI-B2 homodimers, is expected to interact mainly with symmetric DNA. LacI-splitGFP would bind only to asymmetric DNA if a precise heterodimer formation is present.

As shown in Figure 41, LacI-splitGFP displayed an interaction with asymmetric DNA (dark blue), suggesting that splitGFP complementation promoted the formation of an asymmetric DNA-binding species. At the same time, the sensorgrams showed tight interactions with the DNA surface, as the signal did not return to baseline after buffer flow for 10 min and addition of IPTG. Moreover, LacI-splitGFP also showed binding to symmetric DNA (Figure 41, blue). Compared with asymmetric DNA, dissociation from symmetric DNA was faster, and baseline recovery was observed after IPTG addition (Figure 41). These findings are in line with the atypical oligomeric state identified by SEC-MALS. Overall, the combined SEC-MALS and SPR data do not provide evidence for the formation of the intended heterodimeric species. Instead, they indicate a heterogeneous sample containing both homodimeric and heterodimeric assemblies. The cleaved GFP variant LacI-Dimer-cleavedGFP showed a minor association with symmetric DNA immobilised on the surface (SI Figure 156). However, the observed

response was substantially lower than that of LacI-splitGFP. Together with the observation that interaction with asymmetric DNA was present, this further supports the conclusion that LacI-B2 predominantly forms homodimers rather than the intended heterodimeric complex.

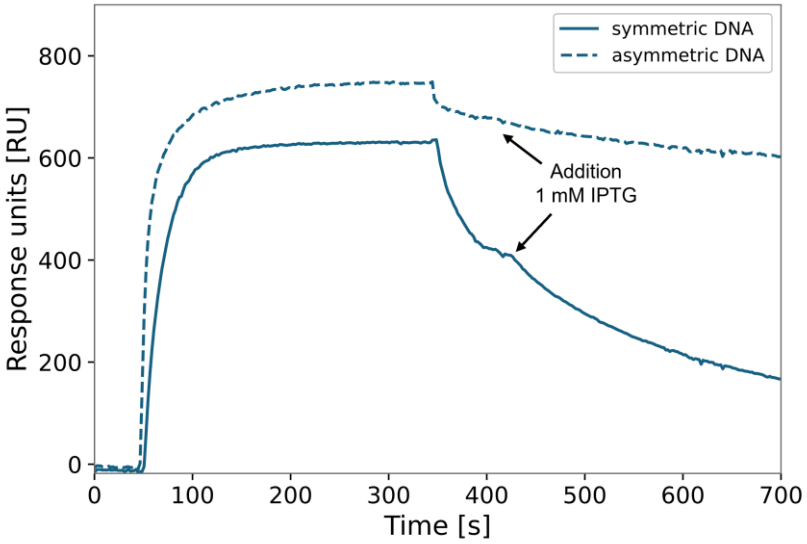


Figure 41. SPR sensograms of LacI-GFP binding to asymmetric and symmetric DNA. Binding to asymmetric DNA showed slower dissociation and incomplete baseline recovery, whereas interaction with symmetric DNA dissociated more rapidly and returned to baseline after IPTG addition. Measurement was performed by Vanessa Troßmann.

Taken together, the biochemical data suggest that the splitGFP fusion promotes the formation of higher-oligomeric states that retain the ability to bind DNA. The results further suggest that the problem already identified in Chapter 2.2.1, namely the residual ability of LacI-B2 to form homodimers, is likely not fully resolved in this variant. This remaining self-association propensity may underlie both the emergence of higher-order oligomeric states and the observed binding to both symmetric and asymmetric DNA. Removing the splitGFP system after expression did not enable the formation of a functional heterodimeric protein.

In principle, using an extension at the C-terminus as a trigger for heterodimerisation seems feasible; improvements to the LacI variants used as a starting point are needed to prevent homodimerisation of the individual parts (LacI-A2-GFP10-11 and LacI-B2-GFP1-9).

2.2.4 A SpyTag/SpyCatcher-based strategy enables stable and functional heterodimerisation of monomeric LacI variants

A major limitation of previous approaches was that, even after introducing five modifications to the monomer-monomer interface, designing a new heterodimeric interface was unsuccessful. However, the use of an external trigger to form a heterodimerisation module at the C-terminus appeared promising for guiding heterodimer formation. Current examinations seem to validate that the external trigger works best when attached to monomeric subunits (Chapter 2.2.3). One possible route would have been to continue with the splitGFP system; however, this system showed clear limitations in the expression of monomeric LacI variants (Chapter 2.2.2). Therefore, splitGFP was replaced by the SpyTag/SpyCatcher system. A monomeric LacI variant was generated based on the LacI-A2 construct (Chapter 2.2.1) (Step 1, Figure 43). SpyTag (LacI-A2-SpyTag) or SpyCatcher (LacI-A2-SpyCatcher) were attached at the C-terminus of LacI-A2 (Step 2, Figure 42). With the SpyTag/Catcher system attached, a covalent bond is formed between the Tag and Catcher subunit after ligation, which should enable the formation of a stable heterodimer. The variant LacI-A2, which was initially used to attach SpyTag and SpyCatcher, harbours additional mutations within the DNA-binding domain to recognise the *galO* operator DNA (Chapter 2.2.1). To enable asymmetric DNA binding, the DNA-binding domain of LacI-A2-SpyCatcher was modified to recognise a *lacO* operator half, resulting in LacI*-SpyCatcher (Step 3, Figure 42).

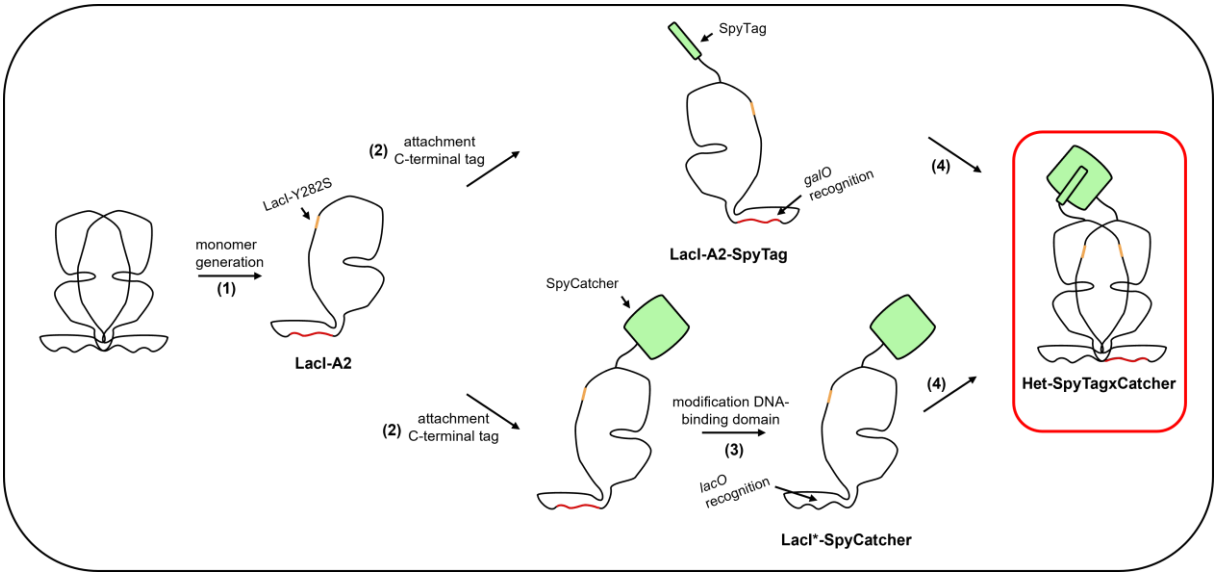


Figure 42. Design strategy for the generation of a heterodimeric LacI construct based on SpyTag/SpyCatcher coupling. Sequential engineering began with the monomerisation of LacI by mutating the monomer-monomer interface, yielding a monomeric LacI variant (1). SpyTag and SpyCatcher were then fused separately to the C-terminus to generate LacI-A2-SpyTag and LacI-A2-SpyCatcher (2). Asymmetric DNA recognition was introduced by modifying the DNA-binding domain of the SpyCatcher-bearing subunit, yielding LacI*-SpyCatcher (3). Finally, covalent SpyTag/SpyCatcher coupling links the two monomeric subunits to form the heterodimeric construct Het-SpyTag×Catcher (4).

In contrast to the previous approaches, no co-expression was performed. Instead, heterodimer formation was induced *in vitro* after cell lysis by mixing the corresponding soluble fractions of LacI-A2-SpyTag and LacI*-SpyCatcher. Het-SpyTag×Catcher was purified in a three-step protocol to enable the separation of heteromeric and monomeric subunits.

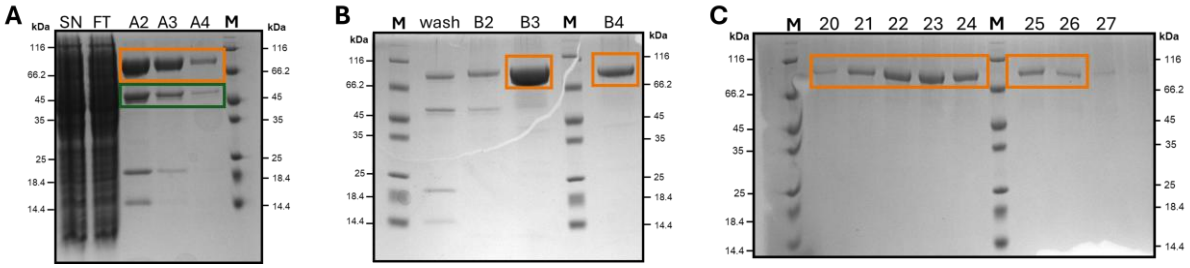


Figure 43. Three-step purification of Het-SpyTag×Catcher. For each purification step, SDS-PAGE is shown, while the corresponding chromatograms are found in SI Figure 98. Proteins of interest are highlighted by boxes coloured according to the corresponding variant. A) Het-SpyTag×Catcher (orange, expected MW: 87 kDa) and LacI*-SpyCatcher (green, expected MW: 48 kDa) are found in the elution, while unreacted LacI-A2-SpyTag (expected MW: 39 kDa) was removed. B) Ni^{2+}-IMAC separated Het-SpyTag×Catcher (orange) from LacI*-SpyCatcher. C) Final SEC isolated the Het-SpyTag×Catcher species (orange) and removed residual impurities.

In the first step, Strep-Tactin AC (Figure 43A) separated Het-SpyTag×Catcher (Figure 43A, orange box) together with LacI*-SpyCatcher (Figure 43A, green box) from unreacted LacI-A2-SpyTag. In a second step, Ni^{2+}-IMAC (Figure 43B) separates Het-SpyTagxCatcher (Figure 43B, orange box) from LacI*-SpyCatcher. This separation was possible because SpyTag-containing species carried a His_6 tag, whereas LacI*-SpyCatcher constructs carried a Strep-II tag. Consequently, the heterodimeric SpyTag×Catcher product contained both affinity tags and was therefore retained on the Ni^{2+}-IMAC column, while unreacted LacI*-SpyCatcher, lacking the His_6 tag, does not bind. SEC proved successful purification by the absence of additional protein contamination and showed that the variant elutes in a single oligomeric state (Figure 43C). As SpyTag and SpyCatcher form an irreversible peptide bond, the resulting heteromeric construct remains intact even under denaturing SDS-PAGE conditions. Therefore, a precise analysis of the ligation is possible by analysing the purification steps with SDS-PAGE. Het-SpyTagxCatcher was present throughout all purification steps and retained high purity. Even though LacI*-SpyCatcher remained after the Strep-Tactin AC (Figure 43A, green box), this does not provide insight into the efficiency of covalent bond formation between SpyTag and SpyCatcher. The reason is that the concentration of the protein of interest cannot be determined before mixing the lysates in the current setup; therefore, non-equimolar rates are used. This can lead to unreacted SpyCatcher species if it was produced in larger quantities during the heterologous expression, even in the presence of a highly effective ligation reaction. The heterodimeric product was obtained in a sufficiently high yield to enable its further use as a putative heterodimeric construct for subsequent incorporation into a designed motor protein. For this reason, neither the exact reaction efficiency nor possible approaches to further

improve it were examined in greater detail. Monomeric variants LacI-A2-SpyTag and LacI*-SpyCatcher were purified after heterologous expression in *Escherichia coli* with an affinity chromatography step depending on their purification tag and SEC (SI Figure 99 and Figure 100).

To validate the formation of a heterodimeric oligomerisation state from monomeric LacI variants (LacI-A2-SpyTag and LacI*-SpyCatcher), the samples were further examined by SEC-MALS (Figure 44.).

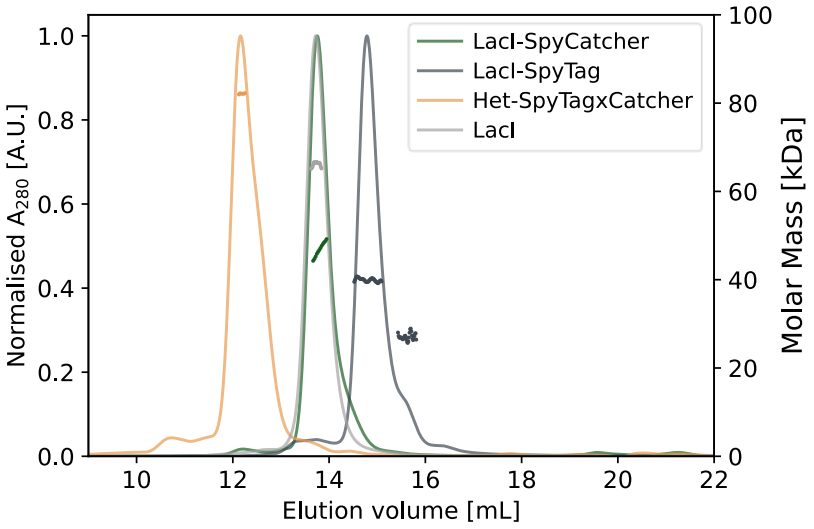


Figure 44. SEC-MALS characterisation of SpyTag- and SpyCatcher-fused LacI variants using an analytical Superdex 200 Increase 10/300 GL column. The normalised absorbance at 280 nm is shown as a function of elution volume, and the corresponding MALS-derived molar mass is plotted on the secondary axis. Het-SpyTag×Catcher (orange, expected MW: 87 kDa) elutes as a well-defined higher-molecular-weight species compared with the monomeric LacI-A2-SpyTag (dark grey, expected MW: 39 kDa) and LacI*-SpyCatcher (green, expected MW: 48 kDa) constructs. Measurements were performed in Buffer C using a protein concentration of 0.5 mg·mL^{-1}.

A distinct monodisperse peak was observed for Het-SpyTagxCatcher (orange). The calculated molecular weight of 88.7 kDa matched the expected value. A detailed summary of the determined molecular weights, as well as the expected molecular weights, is provided in Table 27 of the SI. However, a small secondary peak is present at an elution volume of 10.5 mL. For LacI-A2-SpyTag (grey), in addition to a major peak, a shoulder is present at an elution volume of around 15.5 mL. While the molecular weight determined for the major peak (39.7 kDa) is in good agreement with the molecular mass expected for a monomeric construct (SI Table 42), the lower molecular mass detected for the shoulder at around 15.5 mL might be a result of impurities, which were retained during the purification and are observed in the SDS-PAGE (SI Figure 100). For the main elution peak of LacI*-SpyCatcher (14 mL) the measured molecular mass of the main species corresponded well to that expected for a monomer (SI Table 27), but the molecular mass determined across the peak increased toward larger elution volumes. This behaviour was difficult to reconcile with ideal size-exclusion behaviour, in which smaller species would be expected to elute later than larger ones. At present, the reason for this

behaviour remains unclear. One possible explanation is an altered interaction of the construct with the column matrix. However, this is unlikely given the close similarity of this construct to LacI-A2-SpyTag and Het-SpyTagxCatcher. A more plausible explanation is that inaccuracies in the concentration determination led to an underestimation, resulting in a lower amount of LacI*-SpyCatcher being used for the measurements, which in turn reduced the accuracy of the molecular-weight determination.

Despite this limitation, the overall SEC-MALS results strongly support the conclusion that Het-SpyTag×Catcher forms a defined heterodimer from the largely monomeric LacI-A2-SpyTag and LacI*-SpyCatcher variants. At the same time, a small proportion of higher-order species was detected for all constructs. Given that such species were also observed for variants assumed to be monomeric, they most likely reflect incomplete monomerisation caused by the Y282S substitution (of the variant LacI-A2) at the monomer-monomer interface. However, their abundance remained sufficiently low to be regarded as negligible, and these species were furthermore separable by SEC during the purification steps used for motor protein assembly.

To assess whether the heterodimeric construct remained functional, ligand binding and DNA binding were examined by fluorescence titration and SPR, respectively. Neither the monomeric variants nor the heterodimeric construct showed a substantial alteration in IPTG binding behaviour. A characteristic blue shift of the fluorescence spectrum was observed for all variants (SI Figure 135). Analysing the change in fluorescence at 360 nm shows that all variants reach their plateaus at similar ligand concentrations (Figure 45), indicating that the apparent binding affinities are conserved across the variants. However, the maximal change in fluorescence intensity differs, most notably for the isolated LacI*-SpyCatcher (Figure 45, green). This is a result of lower concentrations used for measurement, but does not imply altered binding affinity. Quantitative evaluation of the fluorescence changes confirmed that the dissociation constants of all constructs fall within the range observed for wild-type LacI (Table 5). As previously shown, single-point mutations within the dimer interface do not affect IPTG binding (Chapter 2.2.1). Likewise, this was also shown for the addition of Tag/Catcher modules to the C-terminus, highlighting that IPTG binding is not influenced by variations required for heterodimer formation.

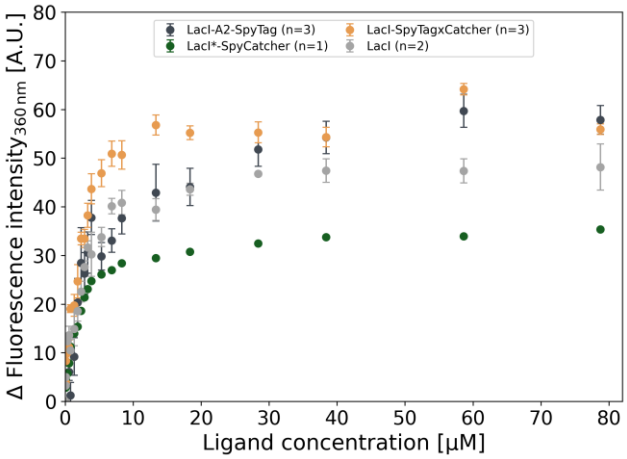


Figure 45. IPTG binding analysis of SpyTag/SpyCatcher-linked LacI constructs by fluorescence spectroscopy. IPTG-dependent changes in fluorescence intensity at 360 nm are plotted for Het-SpyTag×Catcher (orange), LacI-A2-SpyTag (dark grey), LacI*-SpyCatcher (green), and wild-type LacI (light grey). Data points represent mean values, and error bars indicate the standard deviation of the indicated number of replicates. The corresponding fluorescence emission spectra used to derive the changes in fluorescence values are shown in SI Figure 135. Measurements were performed in Buffer C using a protein concentration of 0.2 mg·mL^{-1}.

Table 5. Apparent dissociation constants (K_D) for IPTG binding to heterodimeric LacI (Het-SpyTagxCatcher) and its monomeric subunits (LacI-A2-SpyTag, LacI*-SpyCatcher), obtained by fitting IPTG-dependent change in the fluorescence at 360 nm to a one-site Langmuir binding isotherm.

Construct	K_D [μM]
LacI-A2-SpyTag (n=3)	2.7 ± 0.5
LacI*-SpyCatcher (n=1)	n.d.
Het-SpyTagxCatcher (n=3)	2.1 ± 0.2
LacI (n=2)	2.5 ± 1.2

In contrast to the largely uniform ligand-binding affinities, clear differences were observed in DNA binding. The monomeric proteins LacI-A2-SpyTag and LacI*-SpyCatcher (SI Figure 158) did not show specific binding to asymmetric operator DNA consisting of *lacO* and *galO* half sites. However, both variants showed a slight interaction with their corresponding symmetric operator DNA. Accordingly, LacI-A2-SpyTag, which carries three point mutations in the DNA-binding domain to confer *galO* recognition, showed residual binding to symmetric *galO* DNA, whereas LacI*-SpyCatcher interacted with symmetric *lacO* DNA. This behaviour is consistent with the small dimeric species observed by SEC-MALS, suggesting that monomerisation was incomplete. As the dimer is the minimal functional DNA-binding unit, monomeric species are not expected to show protein-DNA interactions, as the interactions required for a stable protein-DNA complex are not present.[435] The residual dimeric fraction can still bind the corresponding symmetric DNA substrates and therefore causes a weak increase in SPR response. However, the magnitude of this response remained low, indicating that only a small fraction of the protein was present in a DNA-binding dimeric state.

Owing to the distinct mutations in the DNA-binding domains of LacI-A2-SpyTag and LacI*-SpyCatcher, the resulting heterodimer is expected to preferentially recognise and bind an asymmetric operator sequence composed of *lacO* and *galO* half-sites. As expected, the heterodimeric construct Het-SpyTagxCatcher bound asymmetric DNA (Figure 46C), whereas no specific interaction with symmetric DNA was observed (Figure 46AB).

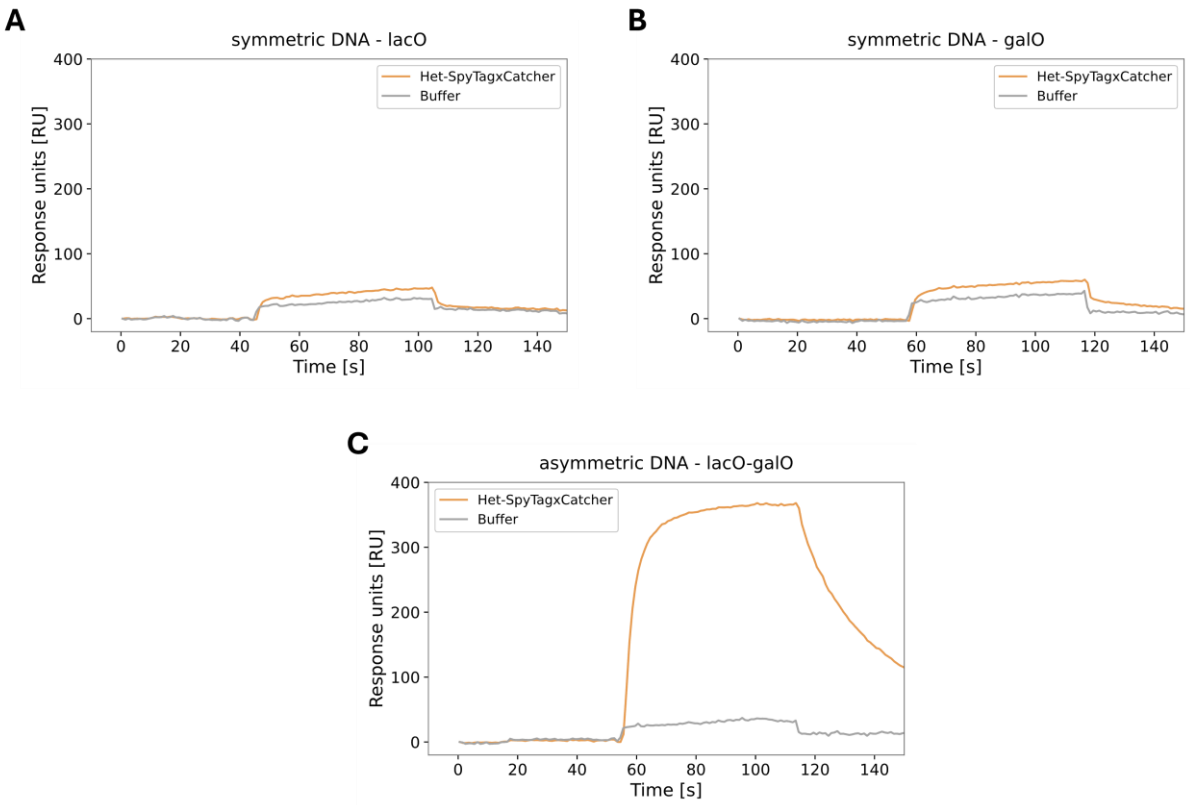


Figure 46. SPR analysis of Het-SpyTag×Catcher binding to different operator DNA sequences. Sensorgrams were recorded at room temperature with a constant flow rate of 30 µL·min⁻¹. Het-SpyTag×Catcher was injected at a concentration of 1 µM over immobilised double-stranded oligonucleotides containing different operator arrangements. Protein-containing sample curves are shown in orange, while buffer/reference curves are shown in grey. A) Symmetric *lacO* DNA. B) Symmetric *galO* DNA. C) Asymmetric *lacO-galO* DNA. SPR measurements were conducted by Vanessa Troßmann.

Analysing the interactions of LacI with asymmetric DNA using concentration-dependent kinetic analysis provides a brief understanding of the differences between the homodimeric and heterodimeric variants of LacI. The association and dissociation of different protein concentrations on chip-immobilised DNA revealed that the heterodimeric variant, upon binding of asymmetric DNA, showed a 24-fold slower association rate, whereas the dissociation rate was in a similar range (Table 6). This indicates that changes introduced to form a dimer strongly influence the association of the protein-DNA complex, while the stability of the complex once formed is hardly affected. These changes in association and dissociation kinetics translated into distinct overall DNA-binding affinities. The heterodimeric construct showed reduced affinity for asymmetric DNA, with a K_D of 127 nM, compared with the wild-type LacI, which bound symmetric DNA with a K_D of 9 nM for wild-type LacI on symmetric DNA (Table 6). However, in contrast to LacI bound to symmetric DNA, which did not return to baseline within 50 s, Het-SpyTagxCatcher was almost completely removed from the surface during the same time period.

Table 6. Kinetic parameters of DNA binding for the heterodimeric construct and wild-type LacI were determined by SPR. Association rate constants (k_{on}), dissociation rate constants (k_{off}), and apparent dissociation constants (K_D) were derived from fits of the corresponding SPR sensorgrams.

Parameter	Het-SpyTagxCatcher	LacI
DNA	asymmetric	symmetric
K_D [nM]	127 ± 8.7	9.5 ± 0.1
k_{on} [M^{-1} s^{-1}]	(2.65 ± 0.51) · 10^5	(2.60 ± 0.08) · 10^6
k_{off} [s^{-1}]	0.035 ± 0.008	0.026 ± 0.002

Using monomeric LacI variants (LacI-A2) and combining them with the split-protein SpyTag/SpyCatcher has proven an efficient strategy for forming a heterodimer that retains ligand and DNA binding. Attaching SpyTag/SpyCatcher to the C-terminus of LacI-A2 with varying DNA-binding motifs enabled the formation of a functional asymmetric DNA-binding species, but the resulting construct does not fully recapitulate the binding properties of the wild-type protein. Even though the DNA-binding affinity is decreased, the Het-SpyTagxCatcher variant enables use in an autonomous motor protein to achieve asymmetric binding to the track.

While the SEC-MALS, IPTG-binding, and DNA-binding data established that the SpyTag/SpyCatcher approach enables formation of a functional heterodimeric LacI construct, the attachment of additional fusion domains may still influence the structural stability of the resulting proteins. Because sufficient stability is an important prerequisite for further implementation of these constructs in motor protein design, their thermal unfolding behaviour was examined by circular dichroism spectroscopy (Figure 47).

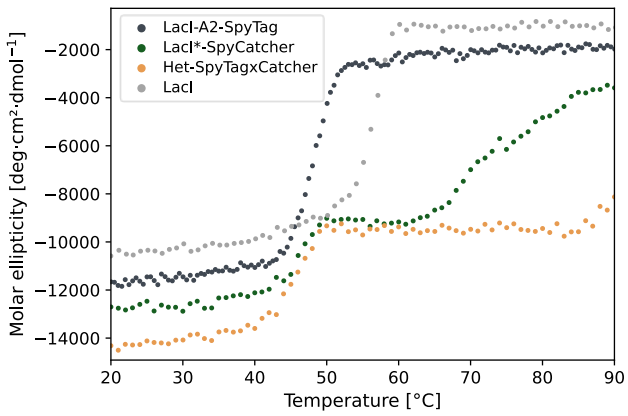


Figure 47. CD-based thermal unfolding analysis of SpyTag/SpyCatcher-linked LacI constructs.

Thermal unfolding of wild-type LacI (light grey), LacI*-SpyCatcher (green), LacI-A2-SpyTag (dark grey), and Het-SpyTag×Catcher (orange) was monitored by the temperature-dependent change in molar ellipticity at 222 nm. Measurements were performed in Buffer C using a protein concentration of 0.2 mg·mL^{-1}.

Table 7. Apparent melting temperatures (T_{Melt}) of SpyTag/SpyCatcher-linked LacI constructs.

Construct	T_{Melt} [°C]
LacI-A2-SpyTag	46.1
LacI*-SpyCatcher	47.5
Het-SpyTagxCatcher	48.1
LacI	55.8

Values were determined from CD-based thermal unfolding experiments as the transition midpoint of the temperature-dependent change in molar ellipticity at 222 nm.

Two clearly separated unfolding transitions were observed for the LacI*-SpyCatcher (Figure 47, green) fusion and the heterodimeric Het-SpyTag×Catcher construct (Figure 47, orange), indicating that the LacI and Spy-derived domains behave as largely independent unfolding units. This is feasible as the two domains are connected via a flexible linker. No second transition was observed for LacI-A2-SpyTag (Figure 47, dark grey), consistent with the fact that SpyTag comprises only a single β-strand, whose unfolding is not primarily followed by an analysis of the change in the molar ellipticity at 222 nm. The first transition, centred around 47 °C, was highly similar across all variants and thus can be assigned to unfolding of the LacI moiety. In comparison to LacI-A2, the thermal stability of the LacI moiety was decreased by 6 °C and is even 2 °C lower than what has been observed for LacI-A1 (Chapter 2.2.1). Therefore, even though the same mutations are present, the addition at the C-terminus seems to impair the stability of LacI, even though a second transition with a higher melting temperature was observed. The melting temperature of the second transition could not be determined accurately because unfolding remained incomplete up to 90 °C. Its onset occurred at a lower temperature (~70 °C) for the isolated SpyCatcher fusion than for the SpyTag×Catcher complex (~85 °C). This suggests that formation of the covalent Tag×Catcher complex stabilises the Spy-derived domain. The comparable apparent melting temperatures of LacI-A2-SpyTag and LacI*-SpyCatcher suggest that the differences between the two variants have only a limited impact

on thermal stability. This conclusion, however, relies on the assumption that additional mutations within the DNA-binding domain do not change the thermal stability. The resulting constructs maintain a level of thermal stability adequate for their intended use as core components in the design of an artificial motor protein.

2.3 Computational design of structural extension at responsive site in LacI

Connecting two asymmetric track-binding domains (feet) is a key requirement for the construction of a functional bipedal motor protein (Chapter 2.1). Various strategies can be used to establish a connection. To obtain a functional connector suitable for further applications in both non-autonomous and autonomous motor proteins, several design parameters must be considered. The central parameters are the secondary-structure composition, the attachment position, and the length. In principle, both feet could be connected by a flexible linker. In the present study, both the ligand-induced closing motion of LacI and the conformational change of the clapping machine are intended to be mechanically coupled to a displacement of the foot along the x-axis (2.1.2). A rigid linker appears more suitable for transmitting conformational changes between the foot and the clapping machine. Therefore, the aim was to design a rigid structural extension that translates the conformational changes of the effector-binding domain or clapping machine into a defined positional shift.

At first glance, the C-terminus seems to be an accessible site for linker attachment. However, the C-terminus was already repurposed to generate a heterodimeric LacI. Furthermore, the C-terminal lobe undergoes only minor conformational changes upon ligand binding.[28,30] Therefore, the C-terminus is less suitable as a responsive site for sensing and transmitting ligand-binding information towards the displacement in the x-direction of the foot.

For this reason, positions that undergo pronounced conformational changes between the open and closed states were envisaged as positions more suitable towards communication sensors. To this end, changes in the dihedral angle were analysed to identify positions within LacI that undergo substantial conformational rearrangements upon inducer binding (Figure 48A).

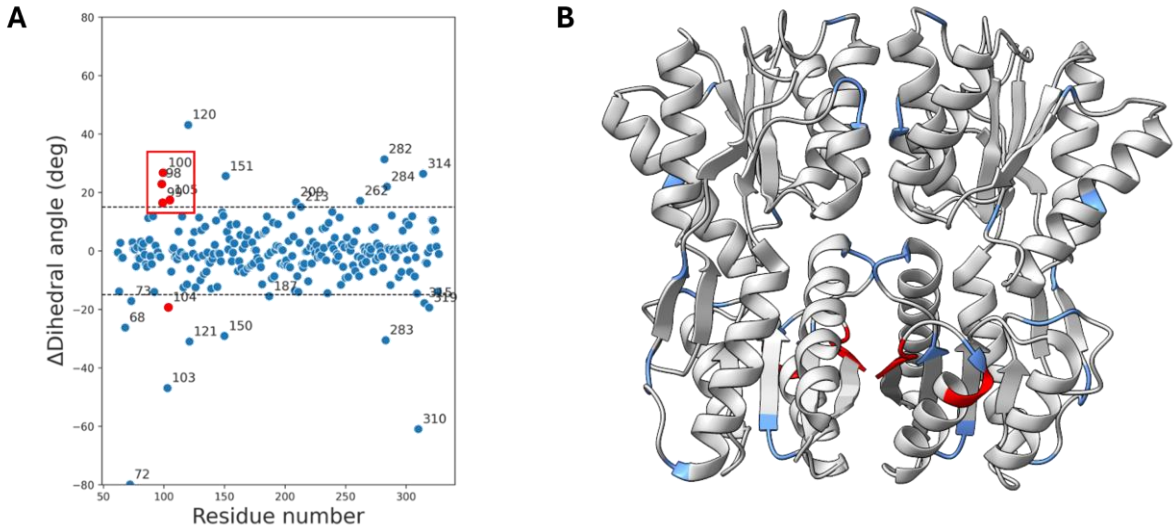


Figure 48. Identification of conformationally responsive regions in LacI based on backbone dihedral-angle changes. A) Residue-resolved dihedral-angle changes between the open and closed conformations of LacI. The red box highlights a contiguous region with pronounced conformational changes. B) Structural mapping of residues with pronounced dihedral-angle changes onto the structure of LacI (PDB:2P9H[30]). Residues with a change above 18° are highlighted in blue. Residues selected as attachment sites for connector elements are highlighted in red.

The cutoff was set to a change of the dihedral angle greater than 18°. The cutoff of 18° was chosen empirically to highlight residues showing the most pronounced conformational differences between the open and closed states. This threshold was sufficiently stringent to exclude minor local variations while still capturing contiguous regions of functionally relevant motion. The largest relative change of the dihedral angle is located in the lower PBP lobe between the second β-sheet and second α-helix of the periplasmic binding protein (residue 98-105, Figure 48, red highlights).

Several strategies for attaching an extension at this position were tested (Figure 49). Initially, a natural antiparallel coiled coil (Figure 49A, blue) was selected and connected to the target site using RF_{joint}. RF_{joint} hallucinates the region between two attachment regions to generate a structure and sequence for the connecting regions (Figure 49A, green). To further optimise the solubility, ProteinMPNN was applied to the designed structure parts (Figure 49A, orange). However, none of the five tested designs was found in the soluble fraction after heterologous expression in *Escherichia coli*.

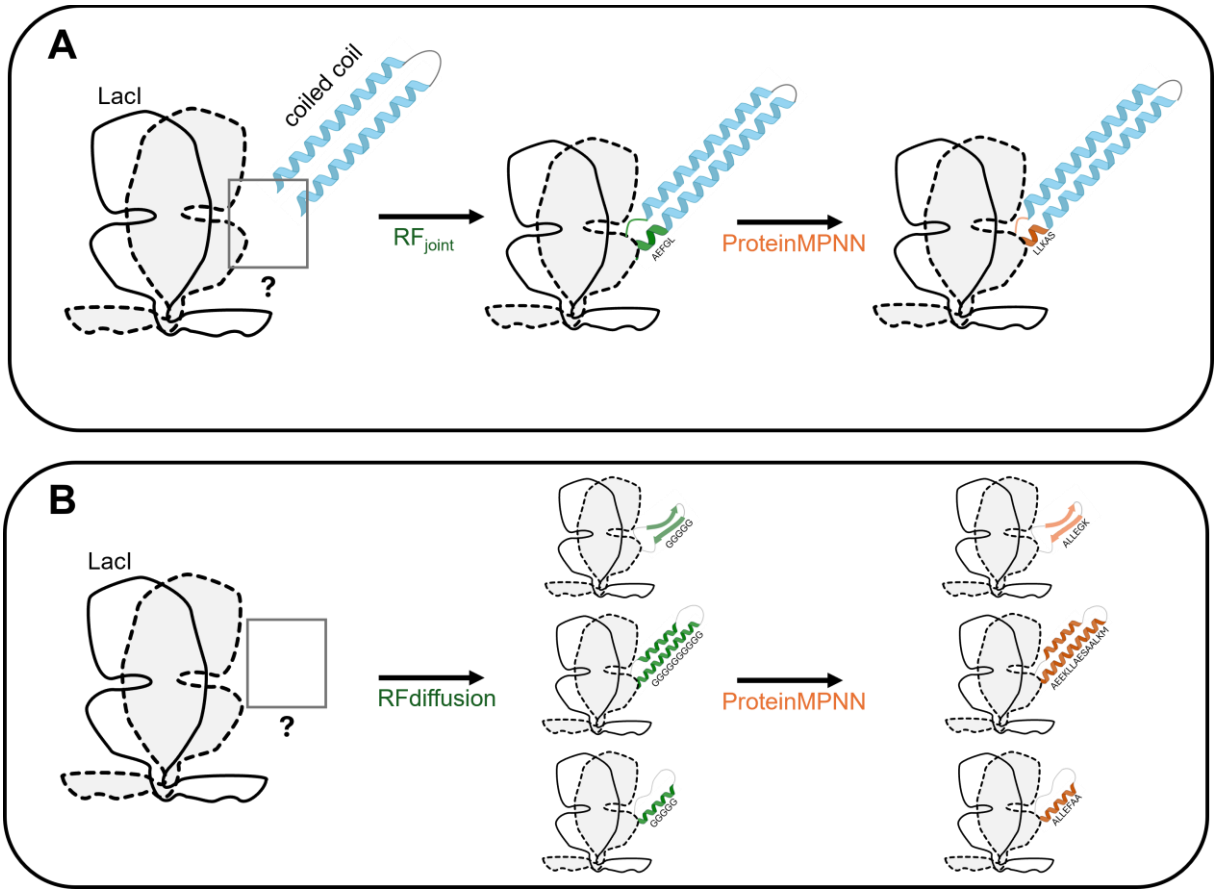


Figure 49. Overview of two computational design pipelines for engineering structural extension into LacI. Starting from the LacI scaffold, extension variants were generated using two complementary approaches. A) In the RF_{joint}-based pipeline, the structure forming the connection (green) between LacI and a coiled-coil fragment (blue) was modelled by inpainting, followed by ProteinMPNN-based sequence design of the inpainted regions (orange). B) In the RFdiffusion-based pipeline, structural extensions (green) were generated, and the sequence was subsequently improved using ProteinMPNN (orange).

To improve the structural integration of a secondary-structure element, RFdiffusion was subsequently employed to generate a novel structural element (Figure 49B, green). ProteinMPNN was used afterwards to optimise the sequence for improved expressibility (Figure 49B, orange). Since LacI contains a complex allosteric communication pathway that couples inducer binding to DNA binding affinity, no amino acids in the LacI core were altered. Only the extension sequence was redesigned using ProteinMPNN to preserve the native-like functions of the variants. A total of 250 designs were generated using RFdiffusion, with 5 different sequences per design generated with ProteinMPNN. Structures of the given sequences were predicted with AlphaFold2. In total, five different designs were selected based on their pLDDT score, secondary structure, and the length of the modelled extension. Only designs with a pLDDT value above 80 in the extension region were considered. Although protein-design tools often yield α-helical structures, this design approach yielded different secondary-structure elements. These included α-helical bundles, mixtures of loops and α-helices, loop-β-strand combinations, and β-sheet-containing structures. For all structural topologies, one design was selected except α-helical bundles, where two designs (LacI-Lev1 and LacI-Lev5) were selected (Figure 50). Linker lengths range from 19 to 37 amino acids. Although the attachment site was selected based on the conformational analysis, one additional design with an α-helical extension attached to the C-terminus (LacI-Lev2) was also included for comparison.

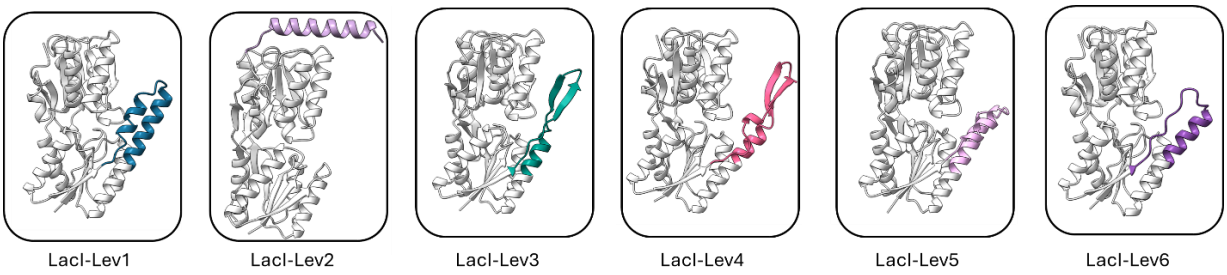


Figure 50. Final set of selected LacI extension designs generated using RFdiffusion and ProteinMPNN. The extensions were created with an RFdiffusion-based design strategy and subsequently sequence-optimised using ProteinMPNN. All designs shown passed the computational selection process and were selected for further characterisation. Structures shown are generated using AlphaFold2.

2.3.1 Biochemical and functional characterisation of LacI variants with designed extensions

Designs were ordered from Twist and incorporated into the second reading frame of the pDUET^{M} vector, with splitGFP present in both reading frames, using RE digestion. Therefore, the splitGFP strands 1-9 are attached to each monomer subunit of LacI-Lev design, while the splitGFP strands 10-11 are expressed without the presence of an additional protein fragment. Of the six designs highlighted in chapter 2.3, two (LacI-Lev1 and LacI-Lev2) exhibited good solubility following heterologous expression in *Escherichia coli* and were selected for further characterisation. While LacI-Lev1 is predicted with AlphaFold2 to form an α-helical bundle placed within the lower PBP lobe, LacI-Lev2 carries the extension at the C-terminus (Figure 50). The central question is whether both soluble designs retain the functional properties of LacI or whether the structural extensions interfere with its native behaviour. The constructs were purified using a three-step protocol comprising Ni^{2+}-IMAC, TEV-mediated cleavage of the His_6 tag and the C-terminal split-GFP(1-9) module introduced to improve solubility, followed by reverse Ni^{2+}-IMAC and final SEC. (SI Figure 101 and Figure 102). First, CD spectroscopy was used to assess whether the designs retain a LacI-like fold and to identify additional α-helical contents (Figure 51).

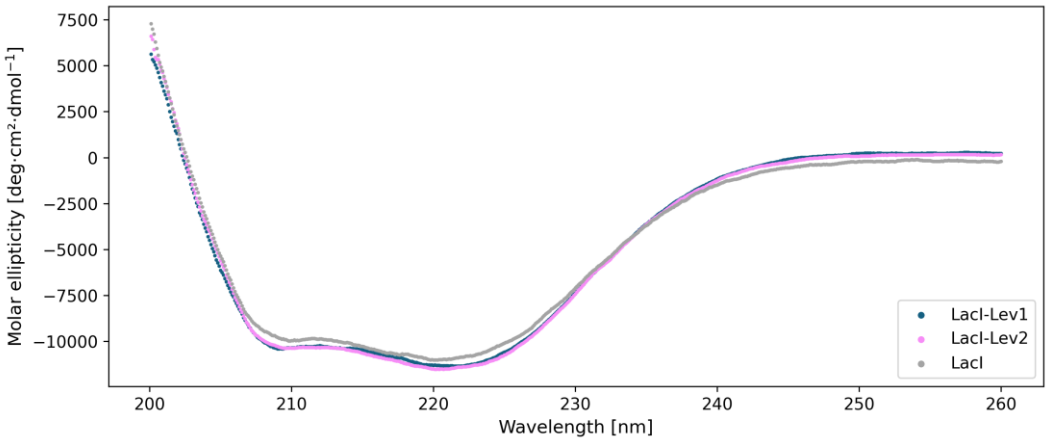


Figure 51. Far-UV CD spectra of LacI variants carrying designed extensions. Spectra of LacI-Lev1 (blue), LacI-Lev2 (pink), and wild-type LacI (grey) indicate that both extension variants retain an overall secondary-structure profile consistent with folded LacI (grey). Measurements were performed in Buffer C using a protein concentration of 0.2 mg·mL^{-1}.

The CD spectra of both variants are consistent with an α/β protein fold. Compared with wild-type LacI, the slightly increased signal intensity at 208 nm and 222 nm may indicate a modest increase in α-helical content. Although CD spectroscopy alone does not directly prove the presence of the designed structured extension, they provide a good indication that the overall LacI fold is retained.

Analysis of thermal stability, following the change in molar ellipticity at 222 nm, revealed pronounced changes due to the extensions, while the cooperative unfolding observed for wild-

type LacI is retained (Figure 52). The attachment site appears to influence protein stability to different extents. While LacI-Lev2 shows only a moderate decrease in thermal stability of 3 °C compared to wild-type LacI, insertion within the lower PBP lobe in LacI-Lev1 reduces the thermal stability by 9 °C (Table 8).

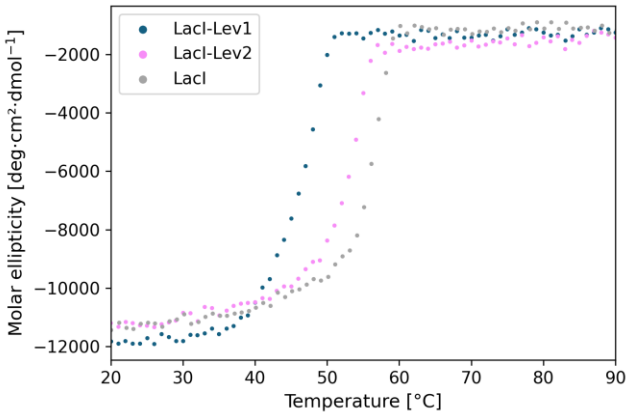


Figure 52. CD-based thermal unfolding analysis of LacI extension variants and wild-type LacI. Thermal unfolding of LacI-Lev1 (blue), LacI-Lev2 (pink), and wild-type LacI (grey) was monitored by the temperature-dependent change in molar ellipticity at 222 nm. Measurements were performed in Buffer C using a protein concentration of 0.2 mg·mL^{-1}.

Table 8. Apparent melting temperatures (T_{Melt}) of LacI extension variants and wild-type LacI. Values were derived from CD-based thermal unfolding measurements by determining the transition midpoint of the temperature-dependent change in molar ellipticity at 222 nm.

Construct	T_{Melt} [°C]
LacI-Lev1	46.2
LacI-Lev2	52.8
LacI	55.8

.

The LacI variant used as the starting construct for the extension design is a dimeric protein. Although no mutations were introduced directly into the dimer interface, insertion of the designed extensions into the lower PBP lobe could still indirectly affect dimer formation. Therefore, the oligomeric state of the LacI-Lev variants was analysed by SEC-MALS (Figure 53). The SEC profiles revealed clear differences between LacI-Lev1 (Figure 53, blue) and LacI-Lev2 (Figure 53, pink).

LacI-Lev2 displayed a largely monodisperse elution peak. In contrast, LacI-Lev1 showed a broader, non-monodisperse elution profile with a distinct shoulder at 10.5 mL. This shoulder is unlikely to originate primarily from contaminating protein species, as SDS-PAGE analysis after the two-step purification procedure did not reveal substantial impurities for LacI-Lev1 (SI Figure 101). Thus, the additional feature in the SEC profile is more likely to reflect heterogeneity within the LacI-Lev1 sample itself rather than incomplete purification. LacI-Lev2 eluted slightly earlier than wild-type LacI with a peak maximum at 9.0 mL, which is consistent with the expectation for a dimeric species with a 4 kDa higher molecular weight than wild-type LacI. LacI-Lev1, however, eluted later, with the main peak shifted to 9.5 mL and an additional shoulder at 10.5 mL. Based on the SEC alone, this latter elution would suggest a smaller apparent hydrodynamic size or a shift towards lower-molecular-weight species compared with LacI-Lev2 and wild-type LacI.

For LacI-Lev2, the experimentally determined molecular weight (72 kDa) was consistent with

the expected dimeric species (76 kDa), demonstrating that this variant retains the dimeric oligomeric state of LacI. A detailed summary of the determined and expected molecular weights is presented in Table 42 found in the supplementary information.

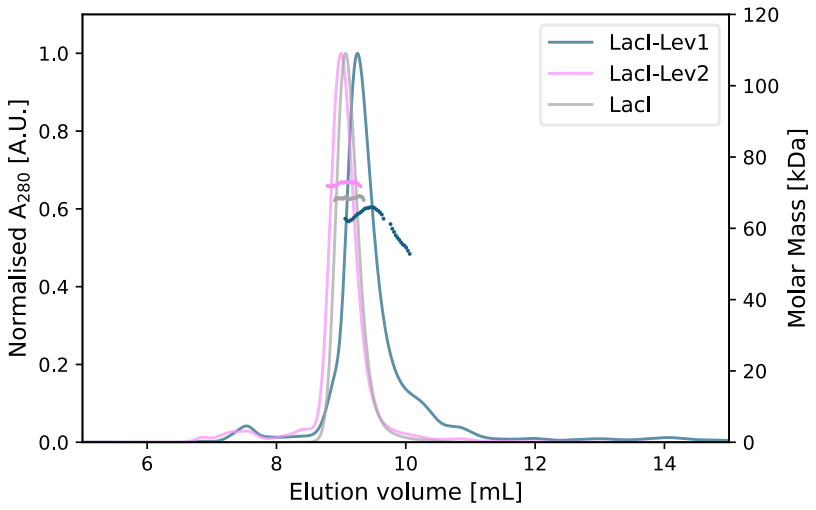


Figure 53. SEC-MALS characterisation of LacI extension variants and wild-type LacI. Measurements were performed using an analytical Superdex 75 Increase 10/300 GL column. Normalised absorbance at 280 nm is shown as a function of elution volume for LacI-Lev1 (blue, expected MW: 78 kDa), LacI-Lev2 (pink, expected MW: 76 kDa), and wild-type LacI (grey, expected MW: 72 kDa), with the corresponding MALS-derived molar mass plotted on the secondary axis. The analysis was used to evaluate the effect of the designed extensions on the oligomeric state of the constructs. Measurements were performed in Buffer C using a protein concentration of 0.5 mg·mL^{-1}.

In contrast, the apparent molecular weight determined for the main LacI-Lev1 peak was 66 kDa, clearly below the theoretical dimeric mass of 78 kDa. However, this value also does not correspond to a purely monomeric species, which would be expected at 39 kDa. LacI-Lev1, therefore, does not behave as a discrete monomeric species or as a stable, fully dimeric species. Rather, the intermediate apparent molecular weight suggests that LacI-Lev1 populates a monomer-dimer equilibrium in solution, in which the individual species are not resolved into separate chromatographic peaks under the applied SEC-MALS conditions. The shoulder at 10.5 mL exhibited an even lower apparent molecular weight than the main peak, but likewise did not fully correspond to a purely monomeric species. In addition, the MALS-derived molecular weight was not constant across the LacI-Lev1 elution peak. Instead, the apparent molecular weight increased across the peak (Figure 53, blue). Such behaviour is not expected for a single, homogeneous oligomeric species, for which the molar mass should remain largely uniform across the elution profile. The non-uniform mass distribution, suggests that LacI-Lev1 represents a heterogeneous ensemble, potentially involving overlapping species, partial dissociation, or interconversion between monomeric and dimeric states during chromatographic separation. The MALS-derived molecular weights confirmed the shift towards lower molecular weight for LacI-Lev1 and do not indicate a change in the hydrodynamic radius. Comparing LacI-Lev1 and LacI-Lev2 shows that the two extension designs affect LacI

oligomerisation differently. The extension introduced in LacI-Lev1 appears to indirectly destabilise dimer formation or promote a more dynamic monomer-dimer equilibrium, whereas the Lev2 insertion is compatible with stable LacI dimerisation.

Although the monomer-dimer equilibrium of LacI-Lev1 differs from wild-type, the more relevant functional properties for its application in motor-protein design are inducer and DNA binding. As for the motor protein, the Lev variants need to be modified into heterodimeric foot constructs, thereby requiring the generation of a monomeric construct (Chapter 2.4).

Since the extension of LacI-Lev1 is positioned close to the IPTG-binding site, it was essential to investigate possible effects on inducer binding by fluorescence titration. For both LacI-Lev1 and LacI-Lev2, a blue shift in the emission spectrum was observed with increasing IPTG concentration (SI Figure 136). The binding affinity was assessed by analysing the changes in fluorescence intensity at 360 nm using two replicates (Figure 53), thereby enabling determination of the apparent dissociation constant by fitting the curve to a Langmuir binding isotherm. The estimated binding affinities are in a similar range for all variants and remain comparable to wild-type LacI (Table 9).

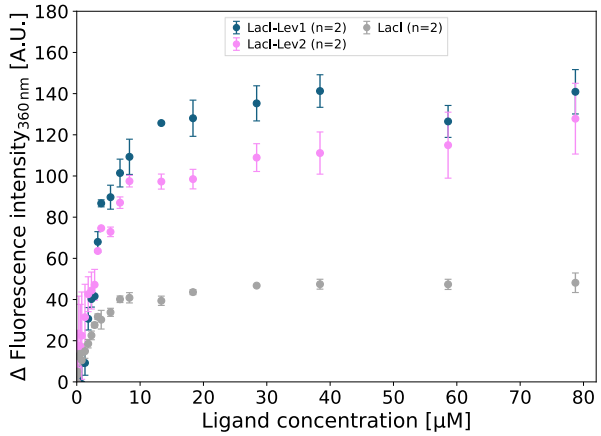


Figure 54. Fluorescence-based analysis of IPTG binding to LacI variants carrying designed extensions. IPTG-dependent changes in fluorescence intensity at 360 nm are plotted against IPTG concentration for LacI-Lev1 (blue), LacI-Lev2 (pink), and wild-type LacI (grey). Data points represent mean values, and error bars indicate the standard deviation of the indicated number of replicates. The underlying fluorescence emission spectra used to calculate the IPTG-dependent change of the fluorescence at 360 nm are shown in the SI Figure 136. Measurements were performed in Buffer C using a protein concentration of 0.2 mg·mL^{-1}.

Table 9. Apparent dissociation constants (K_D) for IPTG binding to LacI extension variants and wild-type LacI. Values were derived from fluorescence spectroscopy by fitting the IPTG-dependent change of fluorescence at 360 nm to a one-site Langmuir binding isotherm.

Construct	K_D [µM]
LacI-Lev1 (n=2)	3.8 ± 0.2
LacI-Lev2 (n=2)	4.2 ± 0.7
LacI (n=2)	2.5 ± 1.2

More substantial deviations from wild-type LacI are present for DNA-binding affinity, particularly for LacI-Lev1 (Figure 55, blue). At identical protein concentrations, the SPR sensorgrams revealed a faster and more pronounced dissociation of LacI-Lev1 from DNA than that of wild-type LacI and LacI-Lev2 (Figure 55, grey and pink), as well as a slower association shown by

the slower increase of the response units.

A detailed kinetic analysis, determined by measuring the association and dissociation constants at varying protein concentrations (SI Figure 156), showed that both LacI-Lev1 and LacI-Lev2 have distinct kinetic properties (Table 10).

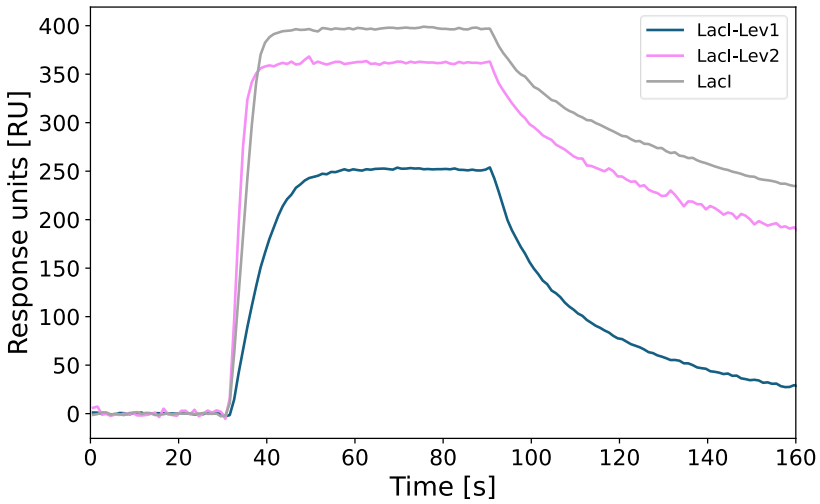


Figure 55. SPR sensorgrams of LacI structural extension variants binding to symmetric operator DNA. Sensorgrams were recorded at a constant flow rate of 30 µL·min^{-1} at room temperature for LacI-Lev1 (blue), LacI-Lev2 (pink) and LacI (grey). Protein variants were injected over immobilised double-stranded oligonucleotides containing the cognate symmetric operator site. Protein association was monitored for 60 s, followed by a dissociation phase of at least 60 s. SPR measurements were conducted by Vanessa Troßmann.

Compared to LacI, LacI-Lev2 behaved similarly. The dissociation rate does not differ, and the association rate only differs by twofold. Consequently, the overall affinity was only moderately reduced by 15 nM relative to LacI, resulting in a K_D of 24.7 nM.

In contrast, LacI-Lev1 showed a substantially altered kinetic profile. While the association rate was fivefold lower than that of the wild type, the dissociation rate was increased by about a factor of 6.5. This strongly accelerated dissociation dominated the overall binding behaviour and led to a pronounced loss in affinity, with the K_D rising to 697 nM. These findings indicate that the structural extension inserted into the lower PBP subdomain in LacI-Lev1 impairs the stability of the DNA-bound complex more severely than in LacI-Lev2.

Table 10. Kinetic parameters of DNA binding for the LacI extension variants and wild-type LacI determined by SPR. Association rate constants (k_{on}), dissociation rate constants (k_{off}), and apparent dissociation constants (K_D) were obtained by fitting the corresponding SPR sensorgrams.

Parameter	**LacI-Lev1**	**LacI-Lev2**	**LacI**
DNA	symmetric	symmetric	symmetric
K_D [nM]	697.0 ± 38.0	24.7 ± 1.3	9.5 ± 0.1
k_{on} [M^{-1} s^{-1}]	(2.35 ± 0.18) ·10^5	(1.10 ± 0.15)·10^6	(2.60 ± 0.08)·10^6
k_{off} [s^{-1}]	0.168 ± 0.016	0.027 ± 0.004	0.026 ± 0.002

Despite the overall more favourable behaviour of LacI-Lev2 with respect to thermal stability, oligomeric state, and DNA-binding affinity, LacI-Lev1 nevertheless provides important proof of principle for the intended motor-protein design. While the variant still retains the essential functional properties of LacI, insertion of the structured extension into the lower PBP subdomain already affects DNA binding. Rather than representing only a detrimental effect, this finding may be highly relevant to future design steps, as it demonstrates that this position may transmit structural changes to the protein's DNA-binding behaviour. Such coupling is expected to be crucial for an autonomous artificial motor protein, in which the extension should mediate inter-domain communication and facilitate regulated release from the DNA track. However, the structure of the extension itself has not yet been experimentally resolved, which remains a crucial aspect for further motor-protein design. Since the extension was specifically introduced to provide a structured, more rigid connection than an unstructured linker, structural validation is essential.

2.3.2 Structural validation of the designed LacI extension by crystallisation of a DNA-binding-domain-truncated variant

Solving the structure will provide important insights into the successful integration of the structural extension into LacI and into the extension's rigidity. However, crystallisation of the Lev variants was not successful. One possible explanation is the intrinsic flexibility of the DNA-binding domain. It was therefore hypothesised that removing this domain would increase the structural homogeneity and rigidity of the protein, thereby facilitating crystallisation.

Accordingly, the DNA-binding domain was removed from both LacI-Lev1 and LacI-Lev2 via RE cloning, yielding the truncated variants PBP1 and PBP2. Initial analysis of thermal stability following structural unfolding using CD spectroscopy revealed a shift in the thermal stability for both variants, while the unfolding cooperativity was not modified (SI Figure 145). However, the increases were modest, amounting to about 2 °C and 4 °C, respectively (Table 11A).

Table 11. Apparent melting temperatures (T_{melt}) of truncated PBP variants, corresponding full-length LacI extension constructs, and wild-type LacI. Values were determined from CD-based thermal unfolding experiments as transition midpoints of the temperature-dependent change in molar ellipticity at 222 nm (left) and by differential scanning calorimetry (right).

Construct	T_{Melt} [°C]
PBP1	48.1
LacI-Lev1	46.2
PBP2	56.9
LacI-Lev2	52.8
LacI	55.8

Construct	T_{Melt} [°C]
LacI-Lev1	47
	50
PBP1	50.5

DSC provides further insight into the effect of removing the DNA-binding domain on the protein's thermal stability. For LacI-Lev1 (Figure 56, blue), the DSC profile resolved two thermal

transitions: a smaller transition at 47 °C and a major transition at 50 °C (Figure 56, Table 11B). This observation suggests that different regions of the protein unfold independently. Analysis of PBP1 (Figure 56, green) supports this interpretation, as the lower-temperature shoulder disappears upon removal of the DNA-binding domain. Together, the disappearance of the lower-temperature transition in DSC and the increase in the thermal stability determined by CD spectroscopy support the conclusion that the DNA-binding domain contributes to the destabilisation of the full-length construct.

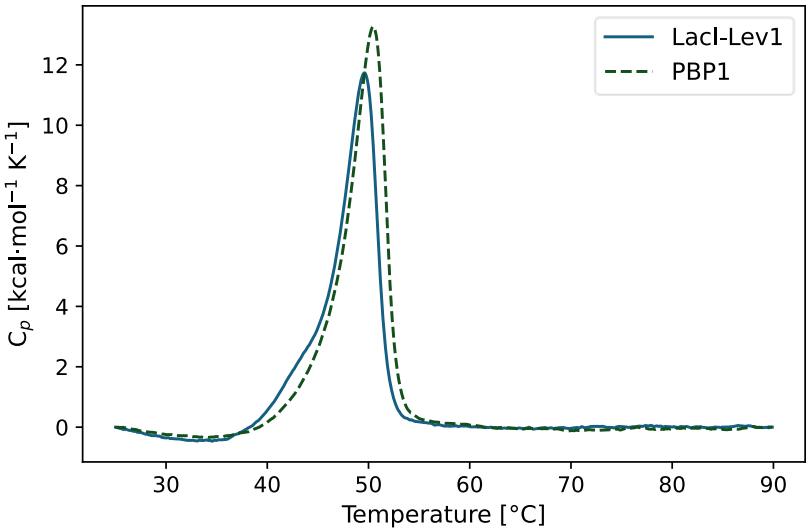


Figure 56. Differential scanning calorimetry profiles of LacI-Lev1 and PBP1. Excess heat capacity (C_p) is plotted against temperature for full-length LacI-Lev1 (blue solid line) and the corresponding truncated PBP1 (green dotted line) variant lacking the DNA-binding domain, allowing comparison of their thermal transition behaviour. Measurements were performed in Buffer C using a protein concentration of 0.5 mg·mL^{-1}.

Even though both LacI-Lev1 and LacI-Lev2 showed an increase in thermal stability upon removal of the DNA-binding domain, crystallisation was only possible for PBP1. A structural comparison of LacI-Lev1 and LacI-Lev2 further highlights an important difference. In LacI-Lev1, the extension is connected to LacI through a continuous α-helix. In LacI-Lev2, the main extension is likewise α-helical; however, its connection to LacI includes a short loop segment (Figure 50). Since crystallisation was only successful for PBP1, it is plausible that this loop introduces additional conformational flexibility in LacI-Lev2, thereby impairing crystallisation. The crystallographic details of PBP1 are listed in the SI Table 48. High structural conservation relative to wild-type LacI is observed. The effector domain of LacI and chain A of PBP1 only varies by around 0.6 Å (SI Figure 166). A cartoon representation comparing the AlphaFold2 prediction of PBP1 (blue) with the experimentally determined structure of PBP1 (green) is shown in Figure 57. The core of PBP1 agrees well with the prediction in both the N-terminal and the C-terminal PBP subdomains (Figure 57A). The elongation of the pre-existing α-helix in the lower PBP subdomain was successfully achieved. In contrast, the second shorter helix of the extension does not completely agree with the AlphaFold2 prediction and shows a change

in the relative position to the lower PBP lobe.

Analysis of the B-factors shows comparable B-factor values for the lower PBP and the helical extension, especially the α-helical extension of the existing α-helix (Figure 57B). Regarding the B-factor chain A (Figure 57B) and chain B (SI Figure 165), they were similar. A similar B-factor of the lower lobe and the extension indicates that the flexibility within the structural extension and the lower PBP lobe is similar. This provides strong structural support for the successful design of a helical extension that can move in accordance with the clapping motion of the lower PBP lobe due to similar rigidity.

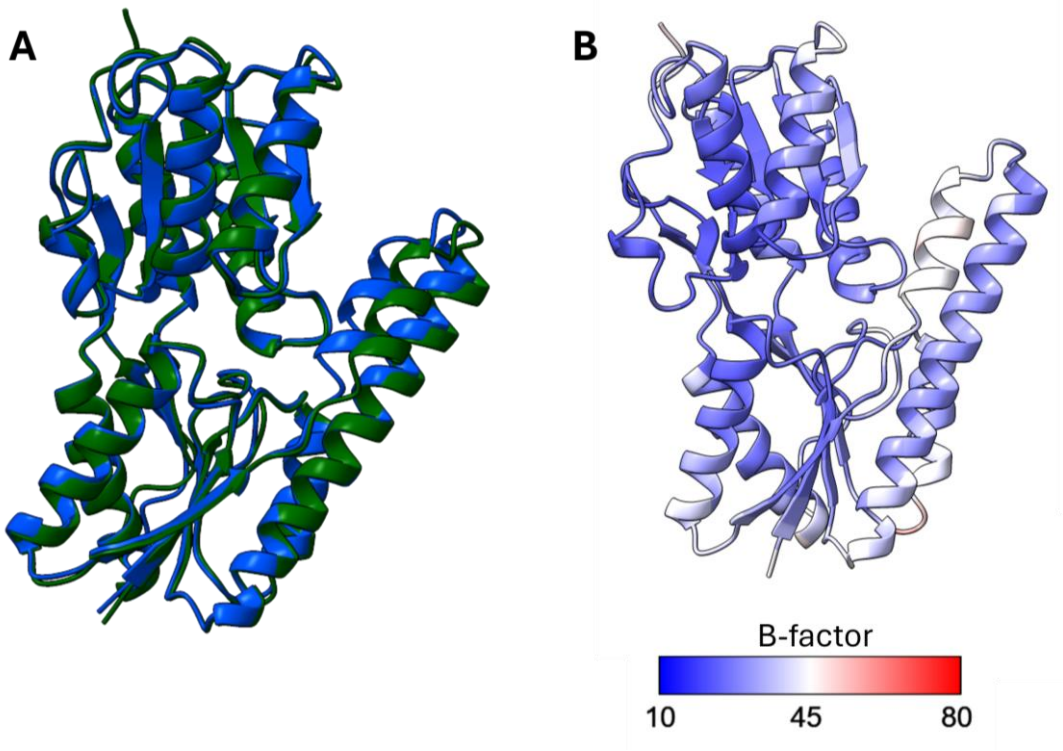


Figure 57. Comparison of the PBP1 AlphaFold2 model with the experimentally determined crystal structure. A) Superposition of the AlphaFold2 model (blue) and the PBP1 crystal structure (green). B) Crystallographic B-factor distribution mapped onto the PBP1 structure, highlighting regions with increased local disorder or flexibility.

Closer inspection of the structure revealed notable differences between chain A and chain B of the crystal structure. Although no ligand was intentionally added during crystallisation, IPTG was present in chain A, whereas no corresponding ligand density was observed in chain B (Figure 58A). Since IPTG had been used during recombinant expression for induction, it appears that a fraction of the ligand remained bound throughout purification. Despite purification by Ni^{2+}-IMAC, reverse Ni^{2+}-IMAC, and SEC, complete removal of IPTG was evidently not achieved.

No major differences in ligand-binding affinity were observed for PBP1 compared with LacI-Lev1 and wild-type LacI (SI Figure 137). More generally, this is consistent with observations made for other LacI-derived constructs analysed throughout this work. The ligand-binding affinity was hardly influenced by the incorporated variations or the expression system used.

Many constructs, including those employed for heterodimer generation, were produced using the pBAD expression system. However, no significant difference in ligand-binding behaviour could be detected between constructs expressed with L-arabinose or IPTG. The ligand-binding data alone do not allow any conclusion as to whether the ligand remained bound throughout purification for other variants analysed throughout this work, or if it is a problem specific to LacI-Lev1.

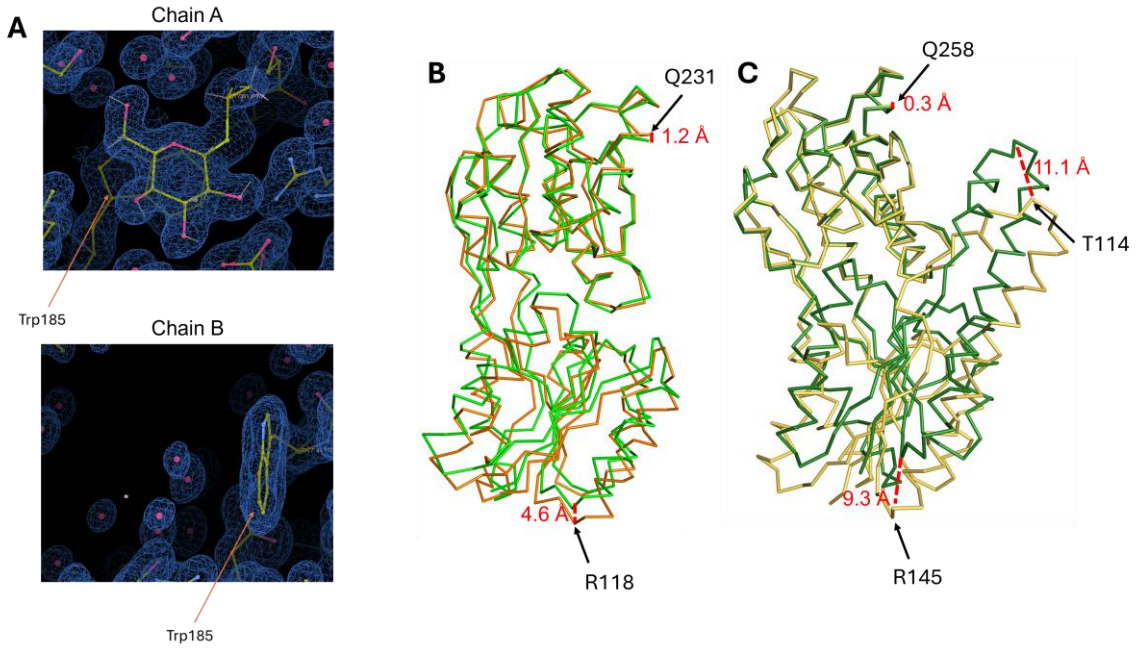


Figure 58. Structural comparison of ligand occupancy and conformational states in PBP1. A) Electron density maps of the ligand-binding pockets in chains A and B, consistent with IPTG bound in chain A but not in chain B. B) Superposition of the Cα traces of wild-type LacI in the closed IPTG-bound state (PDB: 2P9H[30], green) and the open DNA-bound state (PDB: 1LBG[28], orange). Selected residues and the corresponding inter-residue distances are shown to illustrate conformational differences between the open and closed states. C) Superposition of the Cα traces of the crystallographic PBP1 chains, showing chain A in a closed IPTG-bound conformation (green) and chain B in an open conformation without detectable ligand density (yellow). The same selected residues and distances are shown for chains A and B to compare the conformational states observed within the PBP1 crystal structure.

Although the presence of different ligand-bound states of chains A and B within the same asymmetric unit may itself influence the observed conformations, the structural differences between chain A and chain B, due to the presence or absence of IPTG, remain highly informative. In the original design concept, the extension was intended not only to form a rigid structural element but also to preserve a key functional requirement: the ability to transmit the conformational change associated with closure of the PBP domain.

In wild-type LacI, substantial differences in the Cα trace are observed between the open and closed conformation, with the major conformational rearrangements occurring predominantly in the lower PBP lobe (Figure 58B). This is exemplified by the displacement of selected residues upon IPTG binding. While residue Q231 in the upper lobe is shifted by only 1.2 Å, residue R118 in the lower PBP subdomain shows a larger displacement of 4.6 Å. A similar

pattern is observed when chain A (closed, IPTG-bound) and chain B (open, ligand-free) of the crystal structure are compared (Figure 58C). Due to the addition of the structural extension, the amino acid numbering is different. R118 of wild-type LacI corresponds to R145 in PBP1, while Q231 corresponds to Q258. Analysing the conformational changes occurring within PBP1 in the presence of the ligand shows a small change in the position of Q258, found in the N-terminal PBP subdomain, by 0.3 Å, while the residue chosen within the C-terminal PBP subdomain exhibits a change in position of 9.3 Å. The data shows that while the C-terminal subdomain hardly moves upon ligand interaction, the main player of the conformational change is the N-terminal subdomain. The greater displacement of the residue in the N-terminal subdomain of the designed construct compared to LacI may suggest that the helical extension influences the conformational change. Within the extension, residue T114 is displaced by 11.1 Å. This finding highlights that the motion of the lower PBP domain is propagated into the extension, indicating that the inserted structural element could serve as a mechanically coupled communication module in future autonomous motor-protein designs.

The structural and functional characterisation collectively identified LacI-Lev1 as the more promising scaffold for further motor-protein engineering. Solving the structure of LacI-Lev1 provided direct evidence that the designed extension is present, well-folded, and integrated into the lower PBP-like subdomain. At the same time, LacI-Lev1 retained the key functional properties required for its intended role as a motor-protein foot, as both inducer binding and DNA binding were still observed. Moreover, the differences between the Cα traces of chain A and chain B suggest that the extension follows the ligand-dependent closing motion of the PBP domain and may therefore provide a route for transmitting this conformational change to an attached mechanical element. In contrast, the extension of LacI-Lev2 is located at the C-terminus rather than within the moving PBP subdomain, and therefore, it is also less likely to be mechanically coupled to the ligand-induced closing motion as the main interaction happens within the N-terminal subdomain of the PBP. Furthermore, the structure was not solved. The results provided the rationale for prioritising LacI-Lev1 over LacI-Lev2 in subsequent experiments.

2.4 Design and assembly of the first foot of the proposed motor protein

Previous chapters demonstrated that both a successful heterodimer (2.2) and a structural extension (2.3) have been achieved. However, for the design of a motor-protein foot, these two features must be combined within a single construct. Furthermore, additional modifications are still required, as the current extension does not provide a terminus or functional element to attach the foot to the clapping machine. This chapter highlights the implementation of the missing features into LacI-Lev1 to generate a functional foot for an artificial motor protein.

2.4.1 Implementation and screening of the connector module for modular assembly

The first step toward developing a LacI-based motor-protein foot is designing a connection between the structural extension and the clapping machine. There are several possible strategies for designing a linking element between the Lev1 extension and PotF. To ensure high modularity in motor protein assembly, the Catcher/Tag system previously used for heterodimer design was repurposed. A variety of different Tag/Catcher families are available. In the case of LacI-Lev1, the connection to the clapping machine should be introduced into the loop region located between the two α-helices of the extension. For this purpose, a Tag/Catcher family, where two domains are separated after the first two β-strands, appeared particularly suitable. This topology has previously been reported to facilitate insertion into loop regions.[197] Based on these considerations, the DogTag/DogCatcher system was selected. DogCatcher was fused to the clapping machine, as described in Chapter 2.6, whereas DogTag was inserted into the Lev1 loop region. Different strategies were pursued to incorporate DogTag into LacI-Lev1 (Figure 59).

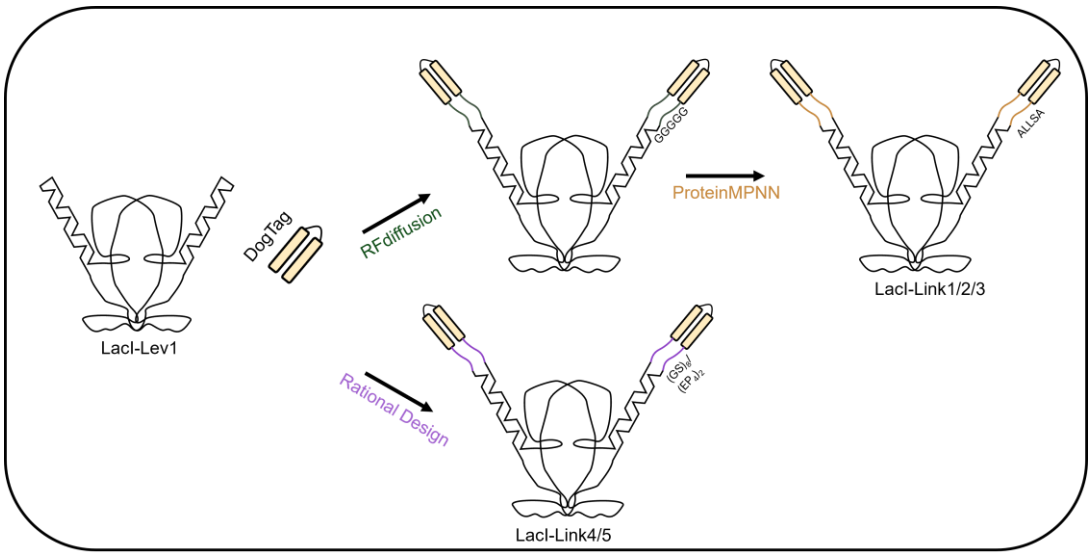


Figure 59. Schematic representation of DogTag integration into the LacI-Lev1 structural extension. Starting from the initial LacI-Lev1 scaffold, two design approaches were pursued. In the first approach, RFdiffusion was used to establish a structural connection between LacI-Lev1 and DogTag, followed by ProteinMPNN-based sequence design, yielding LacI-Link1, LacI-Link2, and LacI-Link3. In the second approach, $(GS)_4$ or $(EP_4)_2$ linkers were introduced by rational design between the DogTag termini and the insertion site within the Lev1 extension, yielding LacI-Link4 and LacI-Link5.

In one approach, the RFdiffusion/ProteinMPNN pipeline previously employed for extension design (Chapter 2.3) was used to generate both the structural connection and the sequence linking DogTag to the pre-existing structural extension of LacI-Lev1. Independent of the selected design length (5-15 residues), the resulting connecting regions were consistently predicted to remain unstructured after prediction with Chai-1 and AlphaFold2. Selection of the final variants was based on two criteria: first, minimal structural deviation from LacI-Lev1 in the PBP main body and the extension region, with an RMSD cut-off of 1 Å. Second, amino acid composition was assessed for all generated sequences, and priority was given to designs whose residue frequencies remained within the Gaussian mean distribution of the amino acid distribution observed across the linker sequence. This was done to minimise compositional bias and to exclude outliers with disproportionately high frequencies of particular amino acids. On the basis of these criteria, three variants with a linker length of 6-10 AS between the structural extension and DogTag were selected and designated LacI-Link1 to LacI-Link3. In the second approach, established linker sequences frequently used in protein design were introduced rationally between the split site of the extension and the N- and C-termini of DogTag. The resulting constructs were designated LacI-Link4 and LacI-Link5, representing insertion of a $(GS)_8$ linker and an $(EP_4)_2$ linker, respectively.

All constructs were ordered from Twist and incorporated into the pET21b(+) vector by RE cloning. Before investigating the interaction between DogTag and DogCatcher, the variants were first characterised with respect to their intrinsic structural integrity and functionality to identify the most suitable construct for subsequent motor-protein design. After heterologous expression in *Escherichia coli*, the soluble fractions of all proteins were purified by Strep-Tactin AC followed by SEC (SI Figure 105-Figure 109). Expression and purification were successful for all variants. CD spectroscopy indicated that all constructs retained the overall secondary-structure composition (α/β mixture) as observed for LacI-Lev1 (SI Figure 141). An increase in the molar ellipticity at 208 nm was shown for LacI-Link5 (purple, SI Figure 141), indicating a modest increase in α-helical content. Since DogTag forms a β-strand, the change in the molar ellipticity is more likely attributable to the $(EP_4)_2$ linker placed between the structural extension and DogTag.

The thermal stability was only marginally influenced by the addition of DogTag, as assessed by the change in molar ellipticity at 222 nm. LacI-Link1 (green) and LacI-Link4 (orange) display unfolding transitions with a degree of cooperativity comparable to that of LacI-Lev1 (Figure 60), and their apparent melting temperatures differ from that of LacI-Lev1 by no more than 1 °C (Table 12). LacI-Link2 (pink) exhibited a comparable melting temperature to LacI-Lev1, but its unfolding is incomplete, with a molar ellipticity of around -6000 $deg \cdot cm^2 \cdot dmol^{-1}$ at 90 °C. For LacI-Link5, an ellipticity of around -6000 $deg \cdot cm^2 \cdot dmol^{-1}$ remained at 90 °C. In the case of LacI-Link5, the $(EP_4)_2$ linker, which seems to form additional α-helices (SI Figure 141), may

promote the formation of a more rigid and thermally stable connection between the two protein moieties, which do not unfold at 90 °C. For LacI-Link2 such an interpretation is less straightforward, since the connection was predicted to be a loop. Nevertheless, it cannot be excluded that this connection likewise adopts additional α-helical structure, thereby enhancing thermal stability. LacI-Link3 (blue) and LacI-Link5 (purple) showed evidence of two thermal transitions, similar to the behaviour previously observed for LacI*-SpyCatcher (Chapter 2.2.4). In contrast, SpyCatcher contains α-helical elements, and their unfolding is expected to contribute strongly to the CD signal at 222 nm. DogTag, on the other hand, consists predominantly of β-sheet structure and therefore contributes less to changes observed at this wavelength.

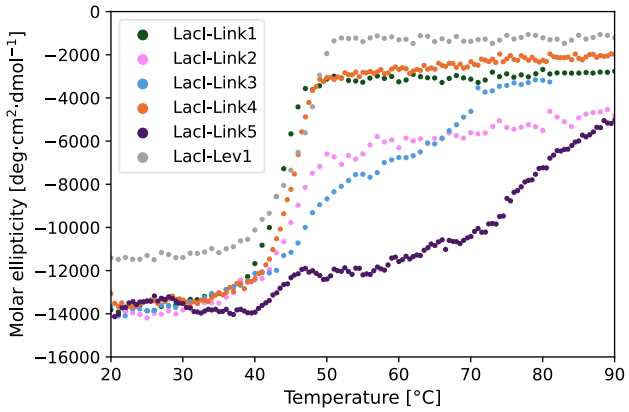


Figure 60. CD-based thermal unfolding analysis of LacI DogTag-linker variants and wild-type LacI. Thermal unfolding of LacI-Link1 (green), LacI-Link2 (pink), LacI-Link3 (blue), LacI-Link4 (orange), LacI-Link5 (purple), and wild-type LacI (grey) was monitored by the temperature-dependent change in molar ellipticity at 222 nm. Measurements were performed in Buffer C using a protein concentration of 0.2 mg·mL^{-1}.

Table 12. Apparent melting temperatures (T_{Melt}) of LacI DogTag-linker variants and the parent construct LacI-Lev1. Values were determined from CD-based thermal unfolding experiments as the transition midpoint of the temperature-dependent change in molar ellipticity at 222 nm.

Construct	T_{Melt} [°C]
LacI-Link1	43.1
LacI-Link2	45.5
LacI-Link3	50.0 73.0
LacI-Link4	45.3
LacI-Link5	45.0 >80
LacI-Lev1	46.2

Except for LacI-Link4, all LacI-Link variants were accessible to SEC-MALS analysis. LacI-Link4 aggregated during the measurement process, rendering analysis of the oligomeric state impossible. The remaining variants displayed markedly different elution profiles (SI Figure 153). LacI-Link1 (SI Figure 153, green) and LacI-Link2 (SI Figure 153, pink) showed largely monodisperse main peaks, indicating comparatively homogeneous samples under the conditions analysed. In contrast, LacI-Link3 (SI Figure 153, blue) exhibited a pronounced shoulder at 14 mL, suggesting the presence of an additional conformational population, as additional species at lower molecular weight were not present during the purification (SI Figure 107). LacI-Link5 (SI Figure 153, purple) showed the most heterogeneous elution behaviour. In addition to a peak near the void volume at 8.2 mL, which may indicate the presence of aggregated proteins, its main elution peak was substantially broader than those of the other variants and exhibited extended shoulders on both sides at 12 and 14 mL. Thus, LacI-Link5

did not behave as a single, well-defined species during SEC-MALS analysis. The heterogeneous profile observed for LacI-Link5 may at least partly originate from contaminating protein species that could not be fully separated from the target construct during the two-step purification procedure, in which additional protein bands remained visible after purification (SI Figure 109). Therefore, the broad elution profile of LacI-Link5 cannot be assigned exclusively to the target protein species. Nevertheless, the SEC-MALS data clearly indicate that this construct differs substantially from the AI-designed LacI-Link variants (LacI-Link1 to LacI-Link3). LacI-Link5 eluted markedly earlier than the other variants, with a main peak centred at 12.5 mL. LacI-Link3 eluted at around 13.3 mL, whereas LacI-Link1 and LacI-Link2 eluted at 13.8 mL. Notably, LacI-Link1 and LacI-Link2 eluted at around the same volume as wild-type LacI, despite their higher theoretical molecular weights if present as dimers. MALS analysis showed that the experimentally determined molecular weights were not consistent with fully dimeric LacI-Link species. For all LacI-Link variants, a dimeric state would be expected to correspond to an apparent molecular weight of 88 kDa. Since these constructs are derived from LacI-Lev1, which itself already displayed a shifted monomer-dimer equilibrium (Chapter 2.3.1), the absence of a clearly defined monomeric or dimeric state was expected. Consistent with this, none of the analysed LacI-Link variants reached the theoretical dimeric molecular weight. The AI-designed variants LacI-Link1, LacI-Link2, and LacI-Link3 showed experimentally determined molecular weights of 50 kDa, placing them much closer to the corresponding monomeric species than to the theoretical dimers. LacI-Link5, by contrast, displayed a higher apparent molecular weight of 65 kDa and therefore deviated by about 23 kDa from the expected dimeric state. Nevertheless, all Link variants exhibited behaviour that accelerates that of LacI-Lev1, for which the experimentally determined molecular weight was 13 kDa lower than the expected dimeric mass (SI Table 43). In comparison to LacI-Lev1, the LacI-Link variants appear to shift the monomer-dimer equilibrium further towards the monomeric state, with the strongest effect observed for LacI-Link1, LacI-Link2, and LacI-Link3. Overall, the LacI-Link variants do not form stable, homogeneous dimers under the conditions tested, but instead populate altered monomer-dimer equilibria. However, the precise oligomeric state of these variants is not the most critical property at this stage, since the monomer-monomer interface will later be deliberately disrupted by the introduction of the Y282S mutation during heterodimer generation (Chapter 2.4.2).

More relevant to the intended application are inducer binding and DNA binding, both of which were affected to different extents. To evaluate the influence of linker design on DNA binding, the binding behaviour of the different Linker variants was compared by SPR analysis (Figure 61). The analysed constructs comprised LacI-Link1 (green), LacI-Link2 (pink), LacI-Link3 (blue), LacI-Link4 (orange), and LacI-Link5 (purple). All sensorgrams of the variants showed a clear association with the DNA-coated surface, indicating that DNA binding was retained in

each case. However, differences were observed in both the association and dissociation phases, demonstrating that linker composition and length strongly affected the interaction.

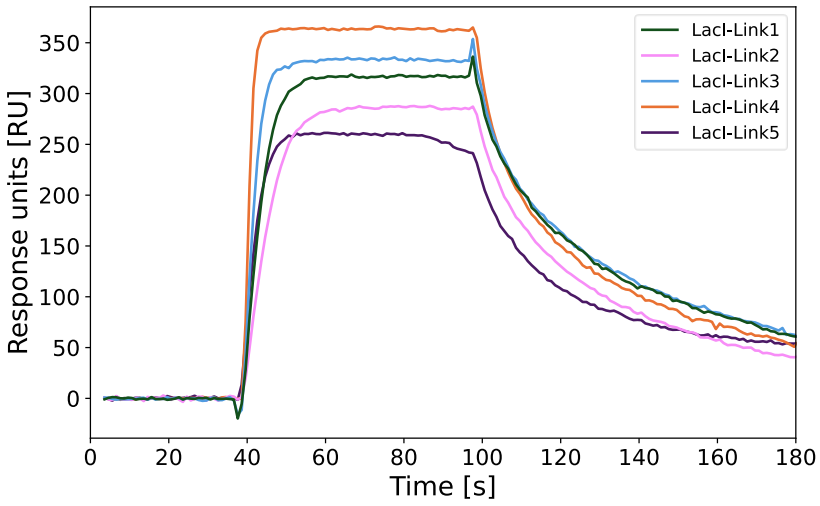


Figure 61. SPR analysis of linker variants' interaction with symmetric DNA. SPR sensorgrams for LacI-Link1 (green), LacI-Link2 (pink), LacI-Link3 (blue), LacI-Link4 (orange), LacI-Link5 (purple) were recorded at a constant flow rate of 30 µL·min^{-1} at RT with 500 µM protein concentration. Protein association was conducted for 60 s via injecting the appropriate protein solution, while protein dissociation was recorded for at least 60 s. SPR measurements were conducted by Vanessa Troßmann.

The highest response during association was reached by LacI-Link4. The lowest response levels are present for LacI-Link2 and LacI-Link5. The response during the association indicates the efficiency of complex formation, meaning that LacI-Link4 forms the DNA-protein complex more efficiently than the other variants. For the dissociation, LacI-Link5 exhibits the fastest response decay. A faster dissociation can be associated with the formation of a less stable protein-DNA complex. Overall, the data indicate that linker optimisation affected DNA binding, with varying effects on association efficiency and complex stability.

Overall, further extension of the LacI-Lev1 construct by adding DogTag did not abolish its functionality. However, the different variants observed different biochemical properties. Considering the biochemical properties, two designs were selected for further use in building the foot of the artificial motor protein (Chapter 2.4.2). Despite its favourable DNA-binding behaviour, LacI-Link4, in which DogTag was connected via a flexible $(GS)_4$-linker, was excluded from further consideration for use in the motor protein because SEC-MALS analysis revealed aggregation. In contrast, the more rigid linker design of LacI-Link5 appeared to interfere most strongly with DNA binding and was therefore not pursued further. This decision was further supported by its thermal unfolding. LacI-Link3 was also excluded from subsequent design steps due to its thermal unfolding and the impact of the extension on the oligomerisation. Consequently, LacI-Link1 and LacI-Link2 were selected for further investigation. Similar to what has been observed for the structural extension (LacI-Lev1) without DogTag, LacI-Link1 and LacI-Link2 exhibit modified DNA affinity compared to wild-type

LacI. Even though the exact values determined by concentration-dependent analysis of the on and off kinetics varied slightly between the different variants, k_{on} remains in a similar range to LacI-Lev1, while k_{off} seems to be slightly improved, with only a limited decrease in the complex stability compared to LacI (SI Table 47 and Figure 160). IPTG binding of LacI-Link1 and LacI-Link2 appeared to remain largely unchanged, as indicated by a comparable blue shift and a similar concentration-dependent fluorescence response across the variants (Supplementary Figure 138).

2.4.2 Heterodimerisation of extension variants for motor-protein foot assembly

As previously shown for LacI, heterodimer formation is achieved by introducing a point mutation at position 282 and by fusing SpyTag or SpyCatcher to the C-termini of subunit A and subunit B (2.2.4). This approach was applied to LacI-Link1 and LacI-Link2, the designs selected (Chapter 2.4.1). To generate a heterodimeric variant that could subsequently be used as a foot in the motor protein assembly, subunits A and B were designed through distinct engineering steps (Figure 62). In the linker constructs, the C-terminus was extended with SpyTag, and the previously established mutation set in LacI-A2 (Table 2) was introduced (Figure 62, subunit A), including mutations in both the DNA-binding domain (Figure 62, red) and the monomer-monomer interface (Figure 62, orange). As a result, the construct used as subunit A is designed to remain monomeric while enabling its DNA-binding domain to recognise the *galO* half-site of the asymmetric DNA substrate. Corresponding variants are named LacI-A2-Link1-SpyTag and LacI-A2-Link2-SpyTag. To realise a bipedal motor architecture in which only one side of the heterodimer is connected to the clapping machine, the linker module was introduced exclusively on one subunit. For this purpose, subunit B was derived from LacI. Subunit B was equipped with a C-terminal SpyCatcher as the coupling element for heterodimer formation (Figure 62, subunit B). To suppress self-association, the monomerising point mutation (Y282S), previously incorporated into LacI-A2, was transferred to subunit B. In contrast to subunit A, the mutations in the DNA-binding domain present in LacI-A2 were deliberately omitted, as native *lacO* recognition had to be retained to enable binding to asymmetric DNA in combination with the complementary *galO*-recognising subunit. As a result, subunit B represents the *lacO*-recognising subunit of the heterodimer and corresponds to the same subunit previously employed for heterodimer formation in Chapter 2.2.4 (LacI*-SpyCatcher). Combining the subunit A and subunit B are expected to enable the formation of the heterodimeric species Het-Link1-SpyTag×Catcher or Het-Link2-SpyTag×Catcher (Figure 62, red box) by the spontaneous isopeptide bond formation.

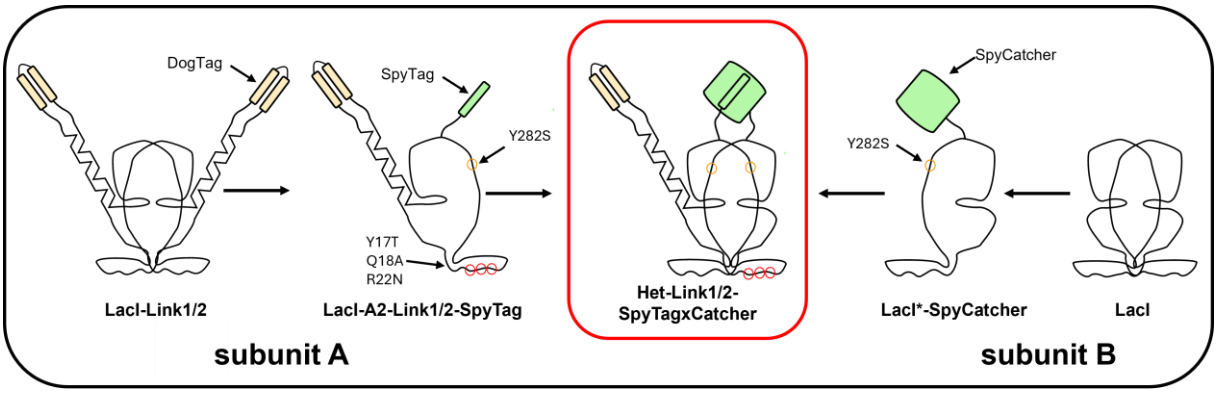


Figure 62. Schematic representation of the modular assembly of a LacI heterodimer. Starting from LacI-Link1 or LacI-Link2 and LacI, SpyTag and SpyCatcher (green) were fused to the respective C-termini. In both subunits, the Y282S mutation (orange) was introduced to weaken the native monomer-monomer interface. In subunit A, the additional mutations Y17T, Q18A, and R22N were incorporated to confer recognition of the *galO* half-site (red). A connection of subunits A and B through covalent SpyTag/SpyCatcher coupling yields the heterodimeric variants Het-Link1-SpyTag×Catcher and Het-Link2-SpyTag×Catcher, highlighted by the red box.

The two variants of subunit A were ordered in the pET21(b)+ vector from Genscript, and subunit B was generated by SDM and RE-based cloning, as highlighted in Chapter 2.2.4. Subunit A and subunit B were produced by heterologous expression in *Escherichia coli*. To characterise the individual subunits, subunit A was purified by Ni^{2+}-IMAC followed by SEC (SI Figure 110 and Figure 111), whereas subunit B was purified by Strep-Tactin AC followed by SEC (SI Figure 99). For the purification of the heterodimeric assemblies, cell lysates containing subunit A and subunit B were combined directly after cell lysis to allow complex formation and were subsequently subjected to sequential purification by Strep-Tactin AC, Ni^{2+}-IMAC, and SEC (SI Figure 112 and Figure 113). After overnight ligation, covalent heterodimer formation was first assessed by denaturing SDS-PAGE (Figure 63A). Since SpyTag/SpyCatcher ligation results in the formation of an irreversible isopeptide bond, the two LacI-derived monomers remain covalently connected under denaturing SDS-PAGE conditions. For both Het-Link1-SpyTag×Catcher and Het-Link2-SpyTag×Catcher, a distinct band at a higher molecular weight than Het-SpyTag×Catcher was observed. The shift to a higher apparent molecular weight confirms heterodimerisation and supports the presence of the additional structural extension in both linker-containing variants. The constructs were further analysed by SEC-MALS to assess whether the ligated species also behaved as defined complexes in solution (Figure 63B).

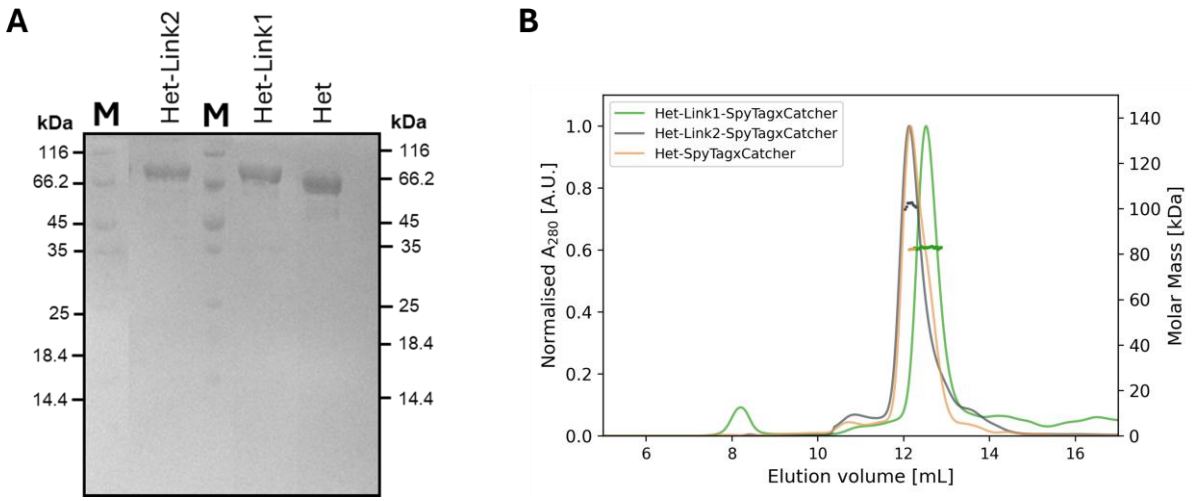


Figure 63. Experimental validation of heterodimer formation and oligomeric state. A) SDS-PAGE analysis of the purified heterodimeric constructs. The shift to a higher apparent molecular weight compared with the parent Het-SpyTag×Catcher construct is consistent with incorporation of the structural extension and DogTag. B) SEC-MALS analysis of Het-Link1-SpyTag×Catcher (green, expected MW: 94 kDa), Het-Link2-SpyTag×Catcher (dark grey, expected MW: 92 kDa), and Het-SpyTag×Catcher (orange, expected MW: 87 kDa), performed using an analytical Superdex 200 10/300 GL column. Absorbance at 280 nm is plotted against elution volume, and the MALS-derived molar mass is shown on the secondary axis. Measurements were performed in Buffer C using a protein concentration of 0.5 mg·mL^{-1}.

All variants show a mainly monodisperse peak; however, Het-Link2-SpyTagxCatcher shows a more pronounced tailing. Together with the main peak at around 12-13 mL, all variants have a smaller peak eluting at around 11 mL. As discussed for Het-SpyTag×Catcher, this minor species may arise from residual dimer formation mediated by the remaining LacI dimer interface, indicating that the Y282S mutation does not fully suppress native LacI-type association in all molecules (Chapter 2.2.4). The main elution volumes differed between the heterodimeric variants. This behaviour is unexpected for an ideal size-based separation. Since SDS-PAGE showed that both linker-containing heterodimers migrate at a higher apparent molecular weight than Het-SpyTag×Catcher, the additional extension would be expected to decrease the elution volume of both Het-Link variants. The MALS-derived molecular weights further distinguished the two linker-containing heterodimers (SI Table 43). For Het-Link2-SpyTag×Catcher, the experimentally determined molecular weight is 97 kDa, while for Het-Link1-SpyTagxCatcher the apparent molecular weight is only 85 kDa. Compared to the expected molecular weight, the difference between the two variants is surprising, as SDS-PAGE (Figure 63A) showed a high-molecular-weight band at a similar height. With a size of 85 kDa, Het-Link1-SpyTagxCatcher has the same apparent molecular weight as the heterodimer without the linker extension, even though it differs in its position on the SDS-PAGE. Nevertheless, SDS-PAGE and SEC-MALS indicate that the presence of DogTag within the structural element does not impair heterodimerisation with SpyTagxCatcher. Additionally, based on their higher theoretical molecular weights, both extended heterodimers would be

expected to elute earlier than Het-SpyTag×Catcher. Instead, Het-Link2-SpyTag×Catcher eluted at a similar volume, and Het-Link1-SpyTag×Catcher eluted later. A conclusion can be drawn regarding the hydrodynamic radius of the properties, taking into account the varied elution behaviour. Het-Link1-SpyTagxCatcher and Het-Link2-SpyTagxCatcher seem to exhibit a smaller hydrodynamic radius than Het-SpyTagxCatcher.

To assess whether SpyTag/Catcher incorporation at the C-terminus and required point mutations in the monomer-monomer interface and DNA binding domain of subunit A and B affected the overall structural stability of the resulting constructs, thermal unfolding measurements were performed. The addition of SpyTag, as well as the introduction of the A2 point mutation, did not further alter the melting temperature of LacI-A2-Link1-SpyTag and LacI-A2-Link2-SpyTag in comparison to LacI-Link1 and LacI-Link2 (SI Table 40). This contrasts with earlier observations, where introducing the mutations associated with A2 decreased the melting temperature by 2 °C (Chapter 2.2.1). In LacI-Link1 and LacI-Link2, the melting temperature had already been reduced by the linker extension to a value of 44 °C, which is lower than the decrease caused by the mutations associated with LacI-A2 alone (54 °C) (Chapter 2.4.1); therefore, further destabilisation effects might not be present due to the additional point mutations. Similar to wild-type heterodimer Het-SpyTag×SpyCatcher (Chapter 2.2.4), formation of the heterodimer also introduced a second unfolding event starting at around 85 °C for Het-Link1-SpyTagxCatcher (SI Figure 146, green) and Het-Link2-SpyTagxCatcher (SI Figure 146, grey).

To further assess whether the introduced modifications affected the functional properties of the constructs, inducer binding was also examined. The IPTG binding was retained and comparable to that of the wild type, with an affinity of 2.5 µM (SI Figure 139). These results highlight that the presence of the linker does not substantially impair ligand binding, even when the extension is present on only one side of the dimer.

In contrast, a pronounced impairment was observed for DNA binding, as neither Het-Link1-SpyTag×SpyCatcher nor Het-Link2-SpyTag×SpyCatcher showed significant interaction with symmetric or asymmetrical DNA (SI Figure 161). No binding signal was detected even at elevated DNA and protein concentrations, indicating that combining the features that generate a heterodimeric version of LacI with the addition of a structural element leads to the loss of DNA-binding capability.

These results show that the modular design of a heterodimeric foot equipped with only one extension and connector for attachment to the clapping machine is feasible in regard to the assembly of the foot. However, the concomitant loss in DNA-binding affinity impairs the use of these heterodimeric variants as a functional foot for the bipedal motor protein.

2.5 Design of a second foot for a bipedal motor protein

2.5.1 Identification and initial characterisation of the scaffold for the second foot

Having two feet is a defining feature of a bipedal motor protein; however, the functional requirements imposed on these feet depend strongly on the underlying motor concept. In a non-autonomous walker, in which stepping is driven by externally applied ligand pulses as described in Chapter 2.1.2, ligand responsiveness of the individual feet becomes a central design criterion. A major objective of the present study was to identify a second LacI-derived foot that provides orthogonal ligand recognition while maintaining specific DNA binding and functional allosteric coupling between ligand binding and DNA affinity. With the aim of reconstituting the proposed non-autonomous motor protein (Chapter 2.1.2.2)

Previous work has shown that this is, in principle, feasible. Several studies[94,436] have reported modifications to the LacI ligand-binding pocket that alter effector recognition, including variants with altered specificity for alternative sugar ligands.[94] However, these studies also indicate that complete orthogonality to IPTG is difficult to achieve, and that many engineered variants retain residual responsiveness to the native inducer.[94,436] In a non-autonomous motor, such residual cross-reactivity would be highly problematic because selective ligand pulses should only change the DNA binding affinity of one foot construct from the track to achieve processivity. One of the most promising candidates to recognise an alternative ligand, described in literature, is a LacI variant carrying the mutations Q291H, T276L, and S279G, here referred to as LacI-Gentio, which was reported to respond to gentiobiose but not to IPTG *in vivo*.[94] To assess whether this specificity is also retained *in vitro*, this variant was generated by SDM. After heterologous expression in *Escherichia coli*, purification by Ni^{2+}-IMAC and SEC (SI Figure 114), and subsequent biochemical characterisation followed.

To test the specific interaction of gentiobiose and IPTG to LacI-Gentio and LacI, DSC experiments were performed to evaluate the influence of the different ligands on the thermal stabilisation of analysed variants (Figure 64). The use of DSC as an indirect readout for ligand binding relies on the preferential stabilisation of the native protein conformation upon complex formation. The thermal energy required to unfold a protein is influenced by the shift of the equilibrium between folded and unfolded states toward the folded state in the presence of a ligand. This is typically observed as a positive shift in the transition midpoint temperature and can also result in a change in the calorimetric unfolding enthalpy.[437–439] Accordingly, ligand-dependent changes in the DSC thermograms provide qualitative evidence for ligand binding, even though the experiment does not directly yield binding affinities under the conditions applied. All DSC-derived thermodynamic parameters are summarised in SI Table 41. No clear shift in melting temperature was observed when gentiobiose was added to LacI-Gentio (Figure 64A). Although changes in the unfolding enthalpy (ΔH) of around 30 $kcal \cdot mol^{-1}$ suggest that

the ligand may influence the thermodynamic unfolding behaviour of the protein, the measurements do not provide unambiguous evidence for productive ligand binding with conformational stabilisation under the tested conditions.

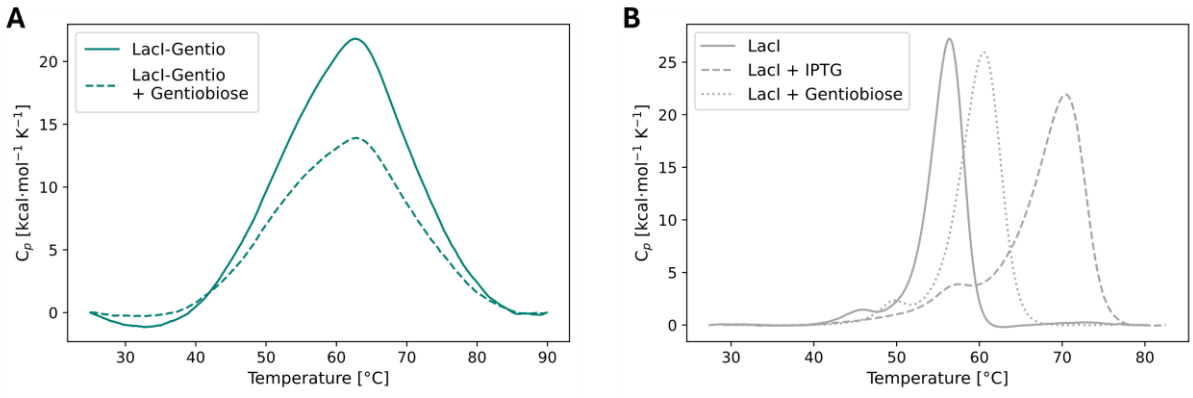


Figure 64. DSC analysis of LacI-Genti and LacI in the presence of different ligands. A) DSC profiles of LacI-Genti in the absence (solid line) and presence of gentiobiose (dashed line). B) DSC profiles of LacI in the absence of ligand (solid line) and in the presence of IPTG (dotted line) or gentiobiose (dashed line). Measurements were performed in Buffer C using a protein concentration of 0.5 mg·mL^{-1}, and ligands were added to a final concentration of 1 mM.

The DSC measurements with LacI showed that the addition of gentiobiose resulted in a thermal shift of 4 °C (Figure 64B). In comparison, IPTG led to a pronounced increase in the melting temperature of LacI by 14 °C (Figure 64B), consistent with strong ligand binding and stabilisation of the folded state. This difference was also reflected in the calorimetric enthalpy, which was markedly lower in the presence of gentiobiose (170 kcal·mol^{-1}) than with IPTG (211 kcal·mol^{-1}), indicating that gentiobiose not only stabilises LacI to a lesser extent but also induces weaker overall energetic coupling upon binding. Although the stabilising effects of gentiobiose are markedly weaker than the IPTG-induced effect, this result nevertheless indicates that wild-type LacI interacts with gentiobiose. Whether gentiobiose is an inducer, a neutral ligand or an anti-inducer for LacI is not analysed in detail. DSC data does not support the use of the LacI-Gentio/gentiobiose pair as a robust basis for selective foot actuation in a non-autonomous motor, as orthogonality is not achieved when LacI interacts with gentiobiose, but LacI-Gentio also lacks measurable gentiobiose interaction.

Using a ligand that is not selected for in nature indicates that a selective response in a given design is difficult to achieve with a small amount of point mutations in the ligand binding domain. A more promising route might be to replace the LacI ligand-binding domain with domains from related transcription factors, while retaining the LacI DNA-binding module. Previous studies have demonstrated that the ligand-binding domain of LacI can be exchanged with regulatory domains from other members of the LacI/GalR family to rewire transcriptional input control or to construct logic-gated regulators.[44,440] For the present study, five chimeric repressors were selected for initial *in vitro* evaluation. Variants were selected based on their

analysed properties *in vivo*.[44,440] These constructs combine the LacI DNA-binding domain with ligand-binding domains derived from RbsR, GalS, PurR, TreR and FruR. The nomenclature used here follows that of Meinhardt *et al.*[25]: L denotes the LacI-derived DNA-binding domain, Lh indicates that the hinge helix also originates from LacI, and the final letter denotes the source of the ligand-binding domain. Among the five variants tested, two constructs, LLhT and LLhF, derived from TreR and FruR, respectively, have been found in the soluble fraction after heterologous expression in *Escherichia coli*. However, their biochemical behaviour differed substantially. Purification of LLhF proved to be problematic. During Strep-Tactin AC as well as SEC, no clean separation could be achieved between a protein species of the expected size (SI Figure 116) and a second slightly smaller species (SI Figure 116). This observation might indicate proteolytic degradation, incomplete translation, or persistent co-purification of a protein that forms protein-protein interactions with LLhF, all of which complicate its use as a defined motor component. In contrast, purification of LLhT was successful, yielding a biochemically homogeneous sample (SI Figure 115). The two constructs also differ markedly in their thermal properties. LLhF displayed a very high thermal stability, with a melting temperature above 80 °C, whereas LLhT unfolded at 54 °C (SI Figure 147, Table 40). The thermal unfolding behaviour of LLhT is highly comparable to that of wild-type LacI, indicating that replacement of the LacI effector domain with the TreR effector domain has little effect on the overall thermal stability of the chimeric protein. At first glance, the higher stability of LLhF might appear advantageous, particularly because subsequent introduction of mutations to generate foot construct like mutations within the monomer-monomer interface (Chapter 2.2.1), linker elements or structural extensions (Chapter 2.3.1) often leads to a substantial reduction in protein stability. Nevertheless, thermal stability alone is not sufficient as a selection criterion for a motor component. In the present context, biochemical homogeneity, defined oligomeric state and absence of undesired intermolecular interactions are equally, if not more important, because they directly affect the predictability and functionality of the motor design.

SEC-MALS measurements were performed to assess the oligomeric state and solution behaviour of the LLhT and LLhF variants (Figure 65). The SEC elution profiles revealed clear differences between the two constructs. LLhT displayed a sharp and largely monodisperse peak at 13.7 mL (Figure 65, pink), closely resembling the elution behaviour of LacI. In contrast, the main peak of LLhF occurred in a similar elution volume range to LLhT, but the peak was broader and showed a distinct shoulder at around 13.0 mL (Figure 65, blue). For LLhT, the experimentally determined molecular weight was 66 kDa, which is below the theoretical dimeric mass of 78 kDa. However, a comparable deviation was also observed for LacI under the same experimental conditions, suggesting that LLhT retains broadly LacI-like solution behaviour despite its slightly reduced apparent molecular weight (SI Table 44). In contrast, LLhF showed an apparent molecular weight 20 kDa higher than expected for the

corresponding dimeric species (SI Table 44). The broad peak indicates that LLhF does not behave as a homogenous sample, while the elevated molecular weight indicates that LLhF does not form a dimer. Considering the purification data (SI Figure 116), it is most likely that LLhF forms a PPI with a coeluting species.

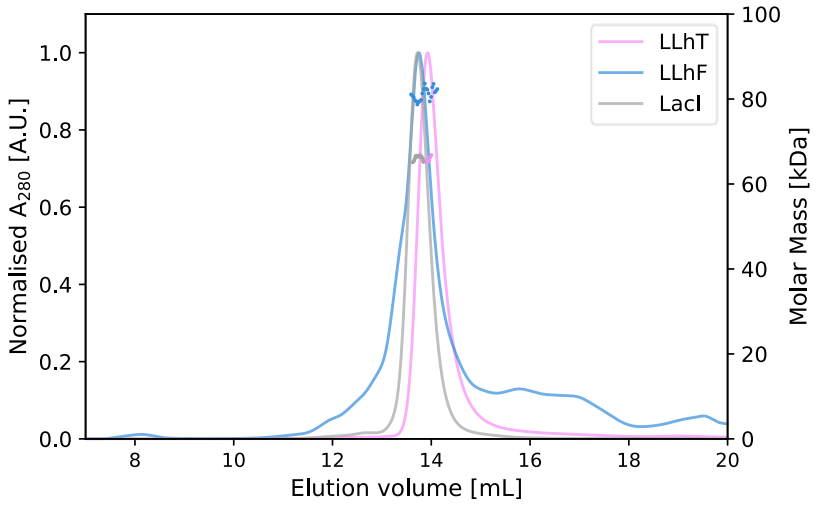


Figure 65. SEC-MALS characterisation of LLhT, LLhF, and LacI. Measurements were performed using an analytical Superdex 200 Increase 10/300 GL column. Normalised absorbance at 280 nm is shown as a function of elution volume for LLhT (pink, expected MW: 72 kDa), LLhF (blue, expected MW: 80 kDa), and LacI (grey, expected MW: 72 kDa), with the corresponding MALS-derived molar mass plotted on the secondary axis. Measurements were performed in Buffer C using a protein concentration of 0.5 mg·mL^{-1}.

Due to the obtained biochemical information, LLhT was selected for further analysis. Even though LLhF is thermodynamically more stable, remaining PPI and imprecise oligomerisation states render it less suitable as the main building block for the foot of the motor protein. To further assess whether LLhT, in combination with its natural ligand T6P, fulfils the specificity requirements for use as an orthogonally addressable second foot, DSC experiments were performed to analyse the effects of T6P and IPTG on the thermal stability of LLhT and LacI (Figure 66, SI Table 41). Ligand-free LLhT (Figure 66A, solid line) had a peak maximum at 58 °C and a total unfolding enthalpy of 114 kcal·mol^{-1}. Addition of IPTG (Figure 66A, dashed line) did not produce a detectable increase in melting temperature and reduced the total unfolding enthalpy to 73.5 kcal·mol^{-1}, arguing against a stabilising interaction. In contrast, T6P (Figure 66A, dotted line) shifted the unfolding transition by 2 °C to higher temperature and increased the total unfolding enthalpy to 293 kcal·mol^{-1}. Thus, although the thermal shift was moderate, the pronounced increase in unfolding enthalpy strongly supports an interaction of T6P with LLhT.

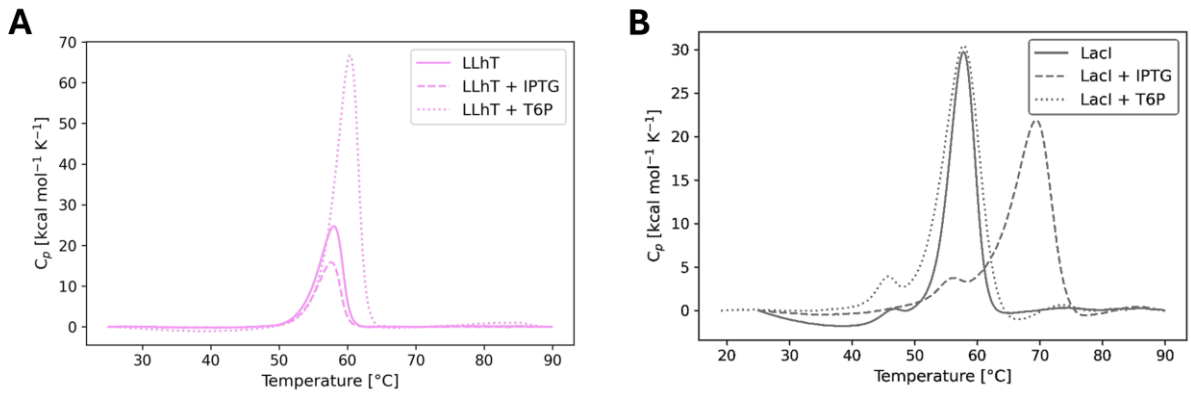


Figure 66. DSC analysis of LLhT and LacI in the presence of different ligands. A) DSC profiles of LLhT recorded in the absence of ligand (solid line) and in the presence of IPTG (dashed line) or T6P (dotted line). B) DSC profiles of LacI recorded in the absence of ligand (solid line) and in the presence of IPTG (dashed line) or T6P (dotted line). Measurements were performed in Buffer C using a protein concentration of 0.5 mg·mL^{-1}. Ligands were added to a final concentration of 1 mM.

DSC analysis of wild-type LacI (Figure 66B) showed a strong stabilising effect upon IPTG addition (dashed line, Figure 66B). The main unfolding transition was shifted from 58 °C to about 70 °C, indicating pronounced ligand-induced stabilisation. However, no increase in the melting temperature is observed after the addition of T6P to LacI, also the shift in the unfolding enthalpy can be disregarded (SI Table 41).

To further evaluate whether LLhT and LacI can serve as orthogonally controlled DNA-binding modules, their ligand-dependent allosteric responses to T6P and IPTG were analysed by SPR (Figure 67). The binding of the proteins to immobilised DNA was followed qualitatively in the presence of the respective ligands. LLhT remained competent for DNA binding in the presence of IPTG (Figure 67A, pink), whereas addition of T6P abolished detectable DNA interaction (Figure 67A, blue). Conversely, LacI retained DNA binding in the presence of T6P (Figure 67B, blue), but DNA interaction was lost upon addition of IPTG (Figure 67B, pink). Thus, LLhT and LacI display reciprocal and ligand-specific allosteric responses. This data supports the suitability of the two proteins as orthogonally regulated feet for a non-autonomous motor architecture, in which alternating ligand pulses could selectively switch the DNA-binding competence of each foot.

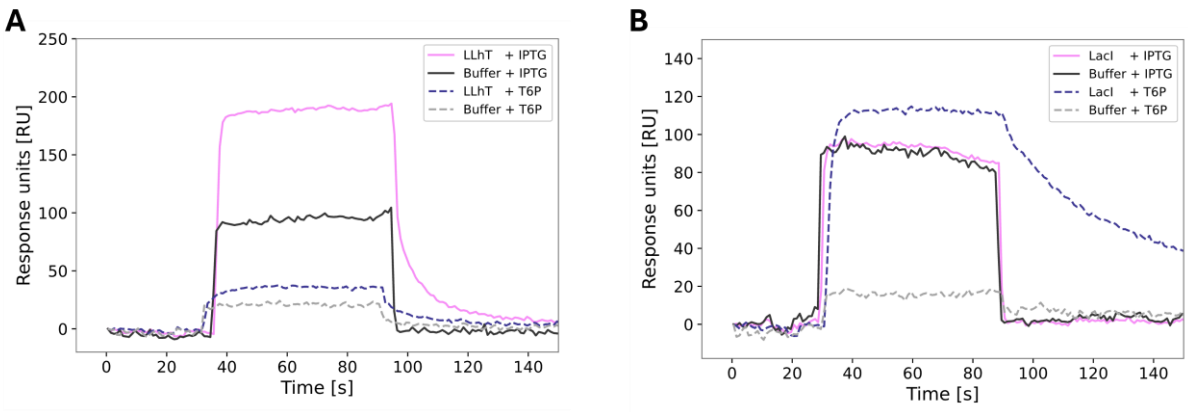


Figure 67. SPR analysis of LacI and LLhT binding to symmetric operator DNA in the presence of different ligands. Sensorgrams were recorded at room temperature with a constant flow rate of 30 µL·min⁻¹. Double-stranded oligonucleotides containing the cognate symmetric operator sequence were immobilised on the sensor surface, and protein variants were injected at a concentration of 1 µM. The respective ligand was supplemented at 1 mM to both the running buffer/reference flow cell and the injected protein solution. Association and dissociation were measured for 60 s. Sensorgrams recorded in the presence of IPTG are shown as solid lines, whereas those recorded in the presence of T6P are shown as dotted lines. Buffer/reference curves are displayed in grey, while protein-containing sample curves are shown in pink for IPTG and blue for T6P. A) LLhT. B) LacI. SPR measurements were conducted by Vanessa Troßmann.

Although the LLhT construct differs from wild-type LacI only in the exchanged reporter domain, this modification nevertheless had a measurable impact on overall DNA-binding behaviour (Table 13). Kinetics of protein-DNA interactions were obtained by concentration-dependent SPR measurements (SI Figure 162). The observed reduced affinity to 83 nM is predominantly caused by a 7-fold increase in the dissociation rate, indicating reduced stability of the protein-DNA complex. No impairment regarding the association kinetics is present.

Table 13. Kinetic parameters of DNA binding for the LLhT and wild-type LacI were determined by SPR. Association rate constants (k_{on}), dissociation rate constants (k_{off}), and apparent dissociation constants (K_D) were derived from fits of the corresponding SPR sensorgrams (SI Figure 162).

Parameter	**LLhT**	**LacI**
DNA	symmetric	symmetric
K_D [nM]	83 ± 12.7	9.5 ± 0.1
k_{on} [M^{-1} s^{-1}]	(2.14 ± 0.42)·10^6	(2.60 ± 0.08)·10^6
k_{off} [s^{-1}]	0.185 ± 0.069	0.026 ± 0.002

These results show that substituting the effector domain affects the construct's overall DNA-binding properties even when the DNA-binding domain itself remains unchanged. Mainly due to the more interactions between the effector binding domain, the hinge helix, and the DNA binding domain, which form a stabilising interaction network within LacI. Nevertheless, LLhT is a promising candidate for use as the second foot in the non-autonomous motor design. In combination with LacI, LLhT establishes a pair of ligand-responsive DNA-binding constructs that can be addressed by IPTG and T6P.

2.5.2 Extension design and initial functional characterisation of extended LLhT variants

As previously outlined in Chapter 2.3, structural attachment of an extension to the PBP domain of a member of the LacI/GalR family is, in principle, feasible. Based on this finding, the same general design strategy was applied to the chimera LLhT to introduce an extension that could serve as a connecting element between the foot and the clapping machine in a bipedal motor architecture.

As discussed for LacI, the key design parameters for such an extension are its attachment site, length, and secondary structure composition (Chapter 2.3). For LacI, the position chosen for extension attachment was guided by structural analysis of conformational changes, in particular by identifying regions that exhibited pronounced changes in dihedral angles between the open and closed states (Chapter 2.3). In the case of LLhT, however, such a comparative analysis was not possible because structural information for both conformational states of LLhT or its parental protein TreR are not available. Therefore, the same attachment position that had previously proven functional in LacI was used for LLhT.

It is noteworthy that, after screening 250 designs generated computationally using RFdiffusion in combination with ProteinMPNN (Figure 68), none of the constructs with a defined secondary structure element as the structural extension that passed the pLDDT cut-off of 80 contained an extension with a similar length as LacI-Lev designs. Even forcing the diffusion of longer extensions was not successful, as only the proportion of loop regions increased and not the length of the secondary-structure element of the extension.

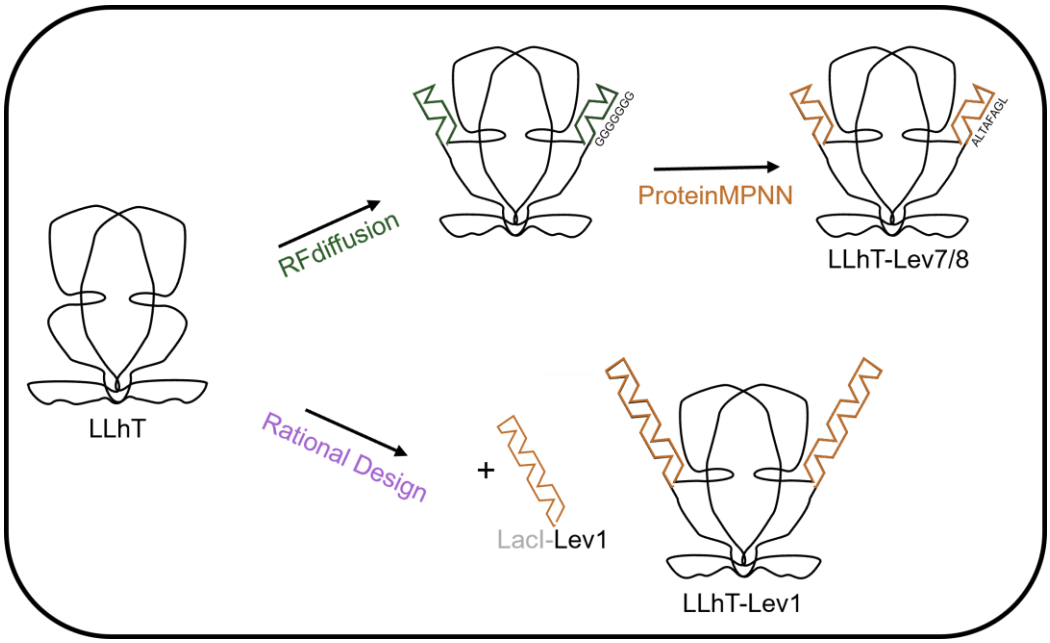


Figure 68. Schematic representation of two different methods to generate structural extension into the lower PBP domain of LLhT. In the first approach, RFdiffusion was used to generate new structural extensions, followed by sequence design with ProteinMPNN, yielding the variants LLhT-Lev7 and LLhT-Lev8. In the rational-design approach, the structural extension previously designed for LacI (Lev1) was transferred to LLhT, yielding LLhT-Lev1.

In contrast to LacI, for which RFdiffusion generated a broad range of structural architectures that satisfied the applied cut-off criteria, all LLhT extension designs were predicted to consist predominantly of α-helices. After further selecting the ProteinMPNN-generated sequences, two designs with extension lengths of 11 and 17 amino acids, termed LLhT-Lev7 and LLhT-Lev8, were selected for experimental testing (Figure 68). To increase the length of the structural extension, a third construct was generated in which the extension previously designed and characterised for LacI-Lev1 was rationally transferred to the PBP domain of LLhT. This construct was termed LLhT-Lev1 (Figure 68). All predictions for the extensions were helices. The prediction of LLhT-Lev1 (Figure 69) displays a pronounced structural similarity of the structural extension to LacI-Lev1, suggesting that the extension geometry was largely retained in the LLhT background. In contrast, LLhT-Lev7 (Figure 69) and LLhT-Lev8 (Figure 69) were predicted to consist of a single α-helical segment followed by an extended loop region as the structural extension.

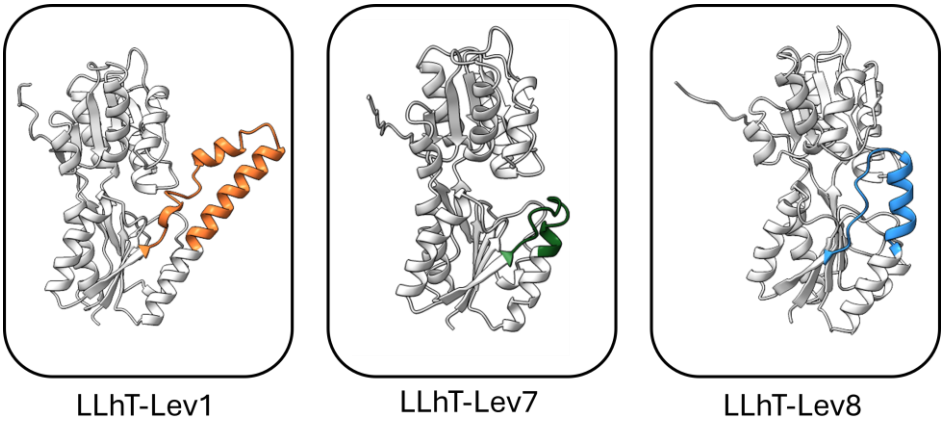


Figure 69. Final set of selected LLhT extension designs generated using RFdiffusion and ProteinMPNN. LLhT-Lev7 and LLhT-Lev8 were generated using an RFdiffusion-based design strategy and subsequently sequence-optimised with ProteinMPNN, whereas LLhT-Lev1 was generated by rational design. The structures shown were predicted using AlphaFold2.

Of the three constructs, LLhT-Lev1 and LLhT-Lev7 were found in the soluble fraction after heterologous expression in *Escherichia coli*. Purification of LLhT-Lev7 by Ni^{2+}-IMAC followed by SEC achieved a protein with decent yield and good purity (SI Figure 118). In contrast, LLhT-Lev1 showed conspicuous behaviour during purification. Even after SEC, high-molecular-weight contaminants remained present in the sample (SI Figure 117). This was likely linked to the behaviour of the target protein itself, which eluted in the void volume, approximately 40 mL earlier than LLhT-Lev7 and considerably earlier than expected based on its theoretical molecular weight (74 kDa). Such an elution profile is indicative of either aggregation or the formation of higher-order oligomeric assemblies. This qualitative observation was further supported by SEC-MALS analysis (Figure 70).

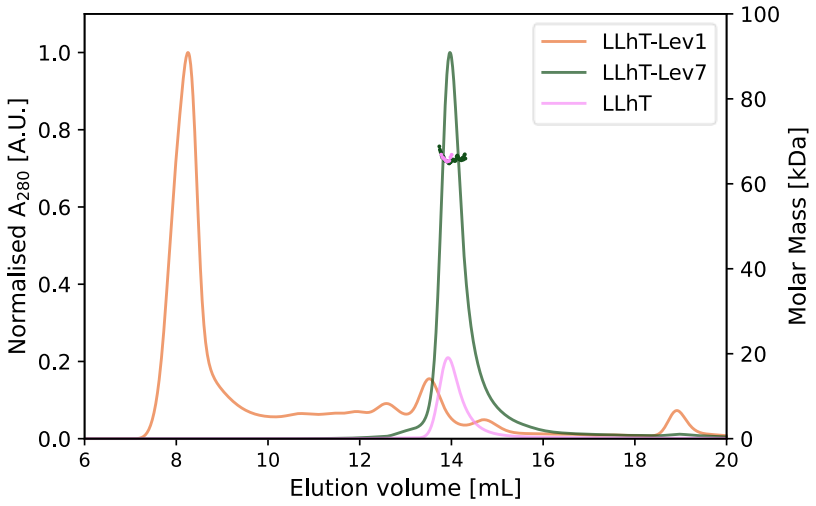


Figure 70. SEC-MALS characterisation of LLhT and LLhT extension variants. Measurements were performed using an analytical Superdex 200 Increase 10/300 GL column. Normalised absorbance at 280 nm is shown as a function of elution volume for LLhT (grey, expected MW: 78 kDa), LLhT-Lev1 (green, expected MW: 78 kDa), and LLhT-Lev7 (orange, expected MW: 74 kDa), with the corresponding MALS-derived molar mass plotted on the secondary axis. Measurements were performed in Buffer C using a protein concentration of 0.5 mg·mL^{-1}.

Even when analysed on a Superdex S200 column, the majority of LLhT-Lev1 (Figure 70, orange) eluted in the void volume. This suggests that the introduced extension strongly affects the solution-state behaviour of the construct, resulting either in aggregation or in the formation of higher-order oligomeric species. CD measurements further indicated slight deviations in the secondary structure content of LLhT-Lev1 compared with LLhT and LLhT-Lev7 (SI Figure 143). In particular, the increased signal around 208 nm suggests an altered secondary structure composition, indicating that the fold of this variant may be affected. In contrast, LLhT-Lev7 (Figure 70, green) exhibited the expected oligomeric state, with a monodisperse peak at around 14 mL and an experimentally determined molecular weight of 66.4 kDa, indicating dimeric behaviour. This shows that the structural extension Lev7 within LLhT does not influence the monomer-dimer equilibrium, which is in contrast to LacI-Lev1 (Chapter 2.3.1). This may indicate that the shorter extension length of the LLhT-Lev7 variant may impose a smaller perturbation. Table 44 in the SI provides an overview of the experimentally determined and calculated molecular weights.

The effect of the extension on thermal stability was likewise examined. For LLhT-Lev7, which retained a dimeric oligomerisation state, the introduction of the extension resulted in a decrease in melting temperature of 14 °C relative to parental LLhT (SI Table 40). This destabilisation exceeds the 9 °C decrease previously observed upon the addition of the structural extension Lev1 to LacI (Chapter 2.3.1). But LLhT-Lev7 is not completely unfolded at 90 °C (SI Figure 148, green). After an initial unfolding, some secondary structure elements remain. LLhT-Lev1 (SI Figure 148, orange) displayed unusual thermal behaviour: the molar

ellipticity increased linearly with increasing temperature; no cooperative unfolding similar to the parental protein, as well as no complete unfolding, was observed. Owing to its unexpected molecular weight and thermal unfolding behaviour, LLhT-Lev1 was excluded from further functional characterisation.

LLhT-Lev7 was analysed further to determine whether attachment of the extension interfered with the two properties essential for its intended function as a motor foot, namely ligand recognition and DNA binding. Ligand binding was first assessed qualitatively by DSC (SI Figure 151). Comparison of the thermograms in the absence and presence of T6P revealed only a minor shift in thermal stability (0.5 °C), while the total area remained unchanged. (SI Figure 151 and Table 41). As no clear conclusion could be drawn regarding whether ligand binding is retained in LLhT-Lev7, SPR measurements were performed to analyse not only DNA binding but also ligand-dependent changes in binding behaviour.

The SPR sensorgrams showed that LLhT-Lev7 still binds DNA in the presence of T6P (Figure 71), as the increasing response upon the addition of the protein LLhT-Lev7 remains.

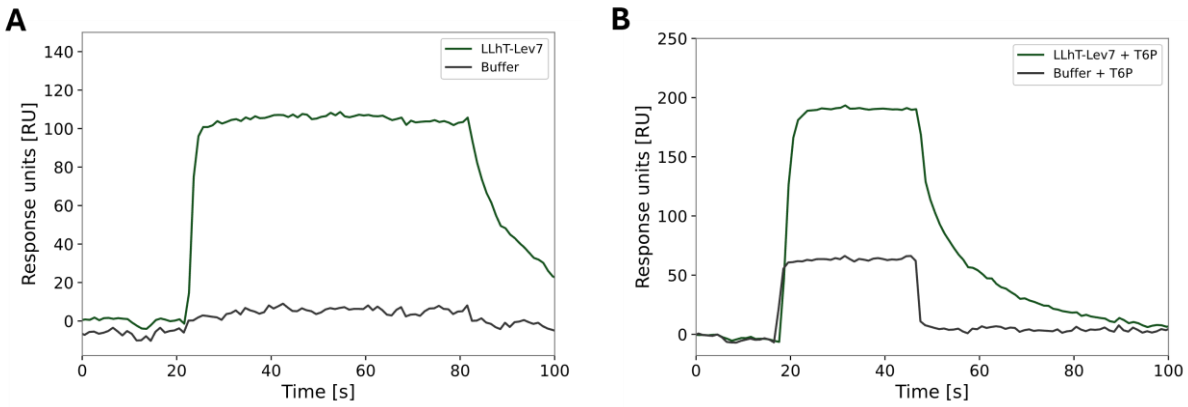


Figure 71. SPR analysis of LLhT-Lev7 binding to symmetric operator DNA in the absence and presence of trehalose-6-P. Sensorgrams were recorded at room temperature with a constant flow rate of 30 µL·min⁻¹. Double-stranded oligonucleotides containing the cognate symmetric operator sequence were immobilised on the sensor surface, and protein variants were injected at a concentration of 1 µM. Association was monitored for 60 s, followed by a dissociation phase of at least 60 s. A) Measurement in the absence of ligand. B) Measurement in the presence of trehalose-6-P. The respective ligand was supplemented at 1 mM to both the running buffer/reference flow cell and the injected protein solution. SPR measurements were conducted by Vanessa Troßmann.

Together with the DSC data, which do not provide clear evidence of ligand binding, this result suggests that LLhT-Lev7 has likely lost its ability to bind T6P, and not only its allosterically controlled DNA binding. However, based on the SPR data alone, an alternative interpretation would be that LLhT-Lev7 has not necessarily lost ligand binding, but its allosteric response to ligand binding. In either case, whether due to loss of T6P binding or loss of allosteric coupling, this variant appears less suitable for application in a non-autonomous motor protein.

Previous experiments demonstrated that different modifications within LacI have a substantial impact on DNA binding. One of them is the introduction of the structural extension Lev1 into LacI (Chapter 2.3.1), which reduced DNA-binding affinity by 73-fold. Since the LLhT parent

construct already shows an around 8-fold lower affinity than wild-type LacI, it was still of interest to determine how strongly the addition of a structural extension would affect DNA binding in this background. This analysis is especially relevant for comparing different structural extensions and their influence on DNA-binding affinity.

Analysing the impact of the structural extension incorporated into LLhT regarding DNA binding showed that, in contrast to LacI-Lev1, the extension Lev7 had only a limited impact on the DNA-binding affinity (Table 14). For a detailed comparison, it is worth noting that the parental LLhT variant already exhibits a nine-fold lower affinity for DNA than wild-type LacI. Introducing the structural extension caused an additional two-fold reduction in affinity. By comparison, insertion of an extension into LacI resulted in a 73-fold decrease in affinity. This difference was primarily reflected in the kinetics of association. Whereas LacI-Lev1 displayed an 11-fold lower association rate than LacI, the association rate of LLhT-Lev7 was reduced only 3.3-fold relative to LLhT. Highlighting that the extension interfered much less strongly with productive DNA association in the LLhT background. Additionally, the presence of the structural element lowered the dissociation constant in relation to the parental protein LLhT, therefore, leading to the formation of a more stable complex.

Table 14. Kinetic parameters of DNA binding for the LLhT-Lev7 with LLhT in comparison to LacI-Lev1 and wild-type LacI were determined by SPR. Association rate constants (k_{on}), dissociation rate constants (k_{off}), and apparent dissociation constants (K_D) were derived from fits of the corresponding SPR sensorgrams.

Parameter	LLhT-Lev7	LLhT	LacI-Lev1	LacI
DNA	symmetric	symmetric	symmetric	symmetric
K_D [nM]	150 ± 27.5	83 ± 12.7	697 ± 38.0	9.5 ± 0.1
k_{on} [M^{-1} s^{-1}]	$(7.40 \pm 1.70)\cdot 10^5$	$(2.14 \pm 0.42)\cdot 10^6$	$(2.35 \pm 0.18)\cdot 10^5$	$(2.50 \pm 0.08)\cdot 10^6$
k_{off} [s^{-1}]	0.110 ± 0.054	0.185 ± 0.069	0.168 ± 0.016	0.026 ± 0.002

The results show that the transfer of the extension design concept from LacI to LLhT is feasible. But the outcome depends strongly on the specific geometry of the extension. Whereas LLhT-Lev1 exhibited severe defects in oligomeric state, LLhT-Lev7 retained the oligomeric state and DNA-binding capability. However, structural extension significantly affected thermal stability and ligand binding.

2.5.3 Engineering of the connector element for the second foot

As shown in Chapter 2.4.1, the presence of an extension alone is not sufficient to establish a fully functional motor foot interacting with a clapping machine. In addition to the structural extension, a dedicated connector element is required to enable coupling of the foot to the central actuator. Because the present design aims at a selectively addressable non-autonomous walker composed of one IPTG-responsive foot and one T6P-responsive foot, a connector system different from DogTag/DogCatcher needs to be considered to enable

specific connections of one foot responding to IPTG and one to T6P. Since SpyTag/SpyCatcher had previously shown favourable binding specificity and good coupling efficiency in the context of heterodimeric LacI assembly, this interaction pair was repurposed here for attachment of the foot to the clapping machine rather than for heterodimer formation (Figure 72). In contrast to the β-sheet DogTag, which was specifically developed for integration into loop regions, SpyTag comprises only a single β-strand and is therefore not well suited for insertion within a structure. For this reason, SpyCatcher was selected for fusion to the LLhT-derived extension. In previous designs involving DogTag, an already characterised extension (LacI-Lev) had been used as the basis for connector attachment (Chapter 2.4.1). In the present case, a different strategy was pursued in the first design round. Rather than first generating an extension and subsequently adding a connector module, the linker region between LLhT and SpyCatcher was designed in a single computational step. Selection of the *in silico* designs was again based on pLDDT scores above 80, and on sequences whose amino acid content distribution fell within the Gaussian distribution of each defined sequence characteristics as applied for the design of the LacI-Linker extensions (Chapter 2.4.1). Two variants, termed LLhT-Link6 and LLhT-Link7, were selected for experimental analysis. With linker lengths of 16 and 35 amino acids, respectively, the designed connecting regions were slightly longer than the extension present in LLhT-Lev7.

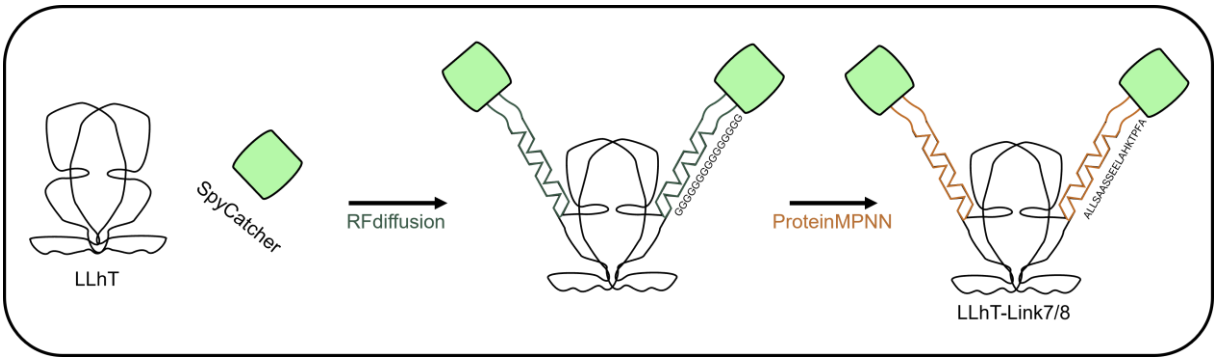


Figure 72. Schematic representation of SpyCatcher integration into the PBP domain of LLhT Starting from the LLhT scaffold, successive designs were generated by RFdiffusion-based design of a connection between LLhT and SpyCatcher, followed by ProteinMPNN-based sequence design.

Both variants are found in the soluble fraction after heterologous expression in *Escherichia coli*. However, similar to the behaviour observed for LLhT-Lev1, both constructs eluted predominantly in the void volume during SEC, indicating highly unfavourable solution behaviour (SI Figure 119, Figure 120).

This observation was confirmed by SEC-MALS measurements, which showed that the majority of both samples eluted in the void volume (<9 mL) and exhibited a broad, multimodal elution profile rather than a single well-defined peak at an elution volume around 11 mL. Those results are consistent with aggregation or formation of higher-order oligomeric species (SI Figure 154).

At the same time, CD spectroscopy indicated that both variants retained an overall α/β-type secondary-structure composition (SI Figure 144). Nevertheless, similar to LLhT-Lev1, which also eluted in the void volume (Chapter 2.5.2), both variants exhibited a more pronounced peak around 208 nm than LLhT. This suggests that the secondary-structure composition may be altered. Thus, although the constructs were not unfolded globally, they failed to adopt a well-defined, biochemically tractable oligomeric state. Neither LLhT-Link6 nor LLhT-Link7 were considered suitable for further functional analysis or subsequent motor design.

These results show that direct one-step design of the LLhT-SpyCatcher connection did not yield constructs which are suitable for use as a foot for the motor protein. As only a limited number of designs were examined, it cannot be concluded that simultaneous design of the extension and the SpyCatcher connection is generally not feasible. Given that similar abnormalities were also observed for LLhT-Lev1, it appears plausible that the selected insertion site constitutes a critical factor influencing the behaviour of these constructs.

2.5.4 Engineering of a monomeric LLhT precursor for heterodimer formation

As shown in Chapter 2.2.4, the generation of a heterodimeric repressor is achieved by suppressing the native homodimerisation to allow controlled heterodimer formation exclusively through orthogonal coupling modules. While interface-disrupting mutations had already been identified for LacI (see Chapter 1.1.1.4), no corresponding set of mutations was available for TreR-derived constructs. A step towards the generation of a heterodimeric LLhT variant was to identify mutations that sufficiently weaken the native monomer-monomer interface to favour formation of monomeric subunits before reconstitution through Tag/Catcher-mediated association.

Comparison of the monomer-monomer interfaces across members of the LacI/GalR protein family reveals relatively low sequence conservation, despite substantial conservation at the structural level.[106] In other words, although the specific amino acid identities differ, the overall architecture of the interface and several key interunit contacts are preserved. Structural conservation allows mapping of interface positions between homologues; for example, residue Y282 in LacI corresponds to residue L272 in TreR.[106] Since single-point mutations at the LacI dimer interface have been shown to produce predominantly monomeric variants,[68,106,441–443] an analogous strategy was applied to LLhT. Amino acids L244, L268, F271, and L272, which are located within the monomer-monomer interface, were individually mutated to alanine. In addition, residue 272 was also considered in the context of a serine substitution, analogous to the interface-disrupting Y282S mutation described for LacI (LacI-A2) (Chapter 2.2.1). All single-point variants exhibited expression and purification behaviour comparable to that of parental LLhT (SI Figure 121-Figure 125). However, the initial analytical SEC did not identify any single mutation that sufficiently destabilised the interface to generate a predominantly

monomeric species (Figure 73A). Compared with the wild-type LLhT, all alanine variants elute at a similar elution volume of around 17.5 mL. This suggests that the overall oligomeric state was retained even in the presence of mutations at the monomer-monomer interface. Unlike LacI, the LLhT scaffold tolerates single-interface mutations and requires a stronger perturbation to disrupt the native dimer. On this basis, a more extensive mutational strategy was pursued. In total, eight structurally conserved residues were identified as contributors to conserved intersubunit interactions (residues 217, 218, 221, 244, 248, 268, 271, 272).[106] Based on these positions, two multiple-mutant variants were generated by modifying four or six out of these eight positions: LLhT-4Mut, in which amino acids L244, L268, F271 and L272 were replaced by alanine, and LLhT-6Mut, in which residues M217, K218, L244, L268, F271 and L272 were all mutated to alanine. Despite the increased mutational load, no major reduction of the soluble fraction after heterologous expression in *Escherichia coli* was observed (SI Figure 126-Figure 127). Analytical SEC provided the first evidence that the introduction of four or six mutations was sufficient to disrupt the monomer-monomer interface, as both variants eluted 3 mL later than wild-type LLhT (Figure 73B).

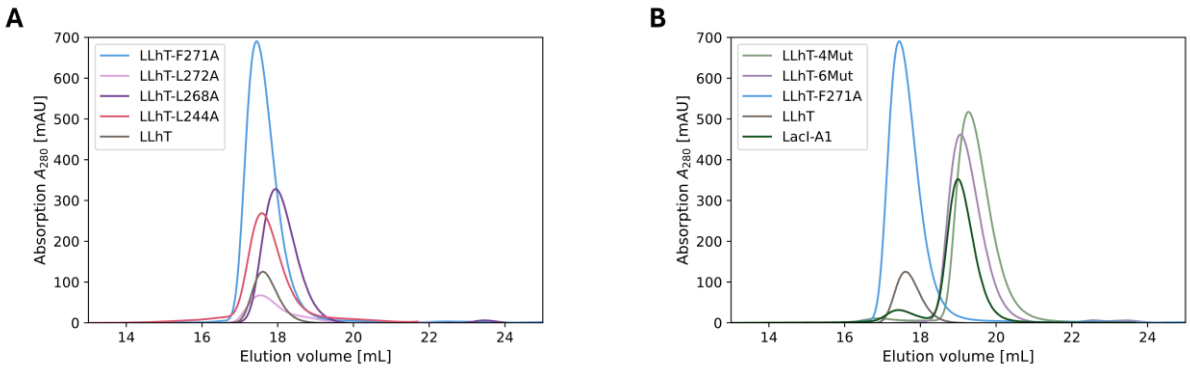


Figure 73. Analytical size-exclusion chromatography analysis of LLhT interface variants. Measurements were performed using a Superdex 200 10/300 GL column. A) Elution profiles of LLhT variants carrying single amino acid substitutions at the monomer-monomer interface compared with LLhT. B) Elution profiles of selected multi-mutant LLhT variants and the corresponding LacI-based reference construct LLhT (dimer) and LacI-A1 (monomer). Measurements were performed in buffer C using a protein concentration of 0.5 mg·mL^{-1}.

Comparisons with wild-type LLhT as well as a monomeric LacI variant showed that the main peak of LLhT-4Mut and LLhT-6Mut eluted at a similar elution volume, indicating monomeric species. Notably, LLhT-4Mut shows a small secondary peak overlapping with the elution position of wild-type LLhT, suggesting the presence of a residual population of dimeric species. LLhT-6Mut was therefore selected for further characterisation.

SEC-MALS confirmed the disturbance of the monomer-monomer interface by introducing six point mutations. LLhT-6Mut eluted as a monodisperse peak (SI Figure 155), and the molecular weight determined for this species was 34.5 kDa, in good agreement with the expected mass of a monomer (36 kDa) (SI Table 44). As observed previously for LacI (Chapter 2.2.1), the introduction of interface mutations also affected the thermal stability of LLhT. In the case of

LLhT-6Mut, the melting temperature was reduced by 10 °C relative to the parental construct (SI Figure 149, Table 40). This destabilisation exceeds that observed for single interface mutations in LacI but remains less pronounced than the effect reported for five point mutations in the interface of LacI (Chapter 2.2.1). These results indicate that the C-terminal PBP subdomain of TreR, and thus of the LLhT chimera, makes a substantial contribution to overall protein stability.

Despite its reduced intrinsic thermal stability, LLhT-6Mut remained responsive to T6P in DSC measurements (SI Figure 152). Upon ligand addition, the unfolding transition shifted by 1 °C to a higher temperature (SI Table 41). In parallel, the total unfolding enthalpy increased from 131 to 177 kcal·mol^{-1}. Although both effects were less pronounced than for LLhT (Chapter 2.5.1), they were more pronounced than for the unresponsive variant LLhT-Lev6 (2.5.2). These changes indicate that LLhT-6Mut still retains ligand-binding capability, albeit with a reduced stabilising response.

In contrast to LacI, monomerisation of LLhT could not be achieved by a single point mutation at structurally conserved positions; however, the combined introduction of multiple interface mutations was sufficient to disrupt the native subunit interaction. The resulting monomeric variant remained soluble and ligand-responsive, providing a favourable basis for subsequent heterodimer construction.

2.5.5 Combination of design elements for the generation of a heterodimeric LLhT-based foot variant

The generation of a heterodimeric foot capable of connecting to the central clapping machine requires the combination of several design features. The native oligomeric state of the repressor must be shifted towards the monomeric state to suppress uncontrolled homodimer formation. Chapter 2.5.4 showed that the introduction of six point mutations was sufficient to generate a monomeric LLhT variant (LLhT-6Mut). Since the selective assembly of a motor protein complex requires distinct and orthogonal coupling systems for foot A and foot B to the clapping machine (Chapter 2.5.3), an alternative ligation pair had to be introduced at the structural extension of foot A and B. For this purpose, SpyTag/Catcher had already been utilised to attach to the clapping machine in foot B. To maintain orthogonal heterodimerisation. DogTag/Catcher was chosen as the heterodimerisation module for foot B. Since none of the previously designed LLhT-Link variants was identified as suitable for use in motor protein assembly (Chapter 2.5.3), the construct LLhT-Lev7 was selected as the basis for further development, as it represented the soluble LLhT-derived extension variant. SpyCatcher was inserted within the Lev7 extension, while DogTag was attached at the C-terminus (Figure 74, subunit A). In addition, the six interface mutations were incorporated to achieve a monomeric oligomerisation state for the subunit. The resulting construct was termed LLhT-Lev7-SpyCatcher-6Mut-DogTag. The complementary subunit B was designed to contain DogCatcher at the C-terminus to establish the heterodimerisation interface (Figure 74, subunit B). In addition, its DNA-binding domain was modified to recognise the *galO* half-site, thereby enabling the heterodimer to bind to the asymmetric operator sequence. For this purpose, residues Y17, Q18, and R22 were replaced with T17, A18, and N22. Furthermore, the six mutations in the interface are present to form the monomeric basic construct. Resulting in a construct named LLhT-6-Mut-DogCatcher.

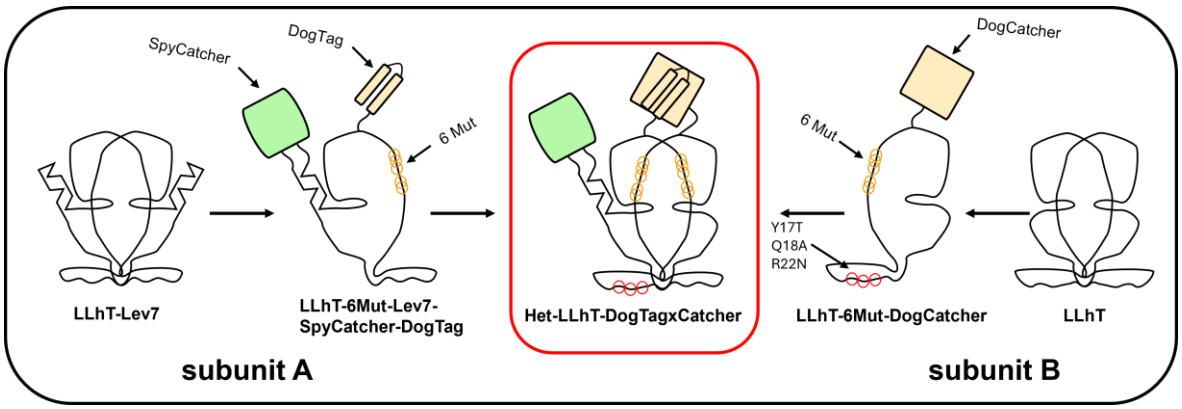


Figure 74. Schematic representation of the modular assembly of LLhT-based heterodimer. Starting from LLhT-Lev7 and LLhT as subunit A and B, six point mutations were introduced (orange highlights). Three additional point mutations were introduced into the DNA-binding domain of subunit B. DogTag and DogCatcher (yellow) were fused to the C-termini of LLhT-Lev7 and LLhT, respectively. The heterodimer formed between subunit A and B is highlighted in the red box.

The amount of recombinant protein in the soluble fraction after heterologous expression in *Escherichia coli* revealed a major asymmetry between the two constructs. Whereas LLhT-6Mut-DogCatcher is found in the soluble fraction (SI Figure 128), LLhT-Lev7-SpyCatcher-6Mut-DogTag was not observed in the soluble fraction. This finding indicates that the combination of multiple design features within the latter construct imposed a substantial burden on protein stability or folding. In principle, several factors may contribute to this behaviour. Insertion of SpyCatcher into the Lev7 extension introduces a large structural element into a scaffold that had already been destabilised relative to its parental protein LLhT. Besides the addition of the structural extension Lev7 into LLhT, which had previously been shown to reduce thermal stability by 10 °C (Chapter 2.5.2), and the introduction of the six mutations within the monomer-monomer interface causes a destabilisation of 13 °C (Chapter 2.5.4). It appears that combining these modifications into a single construct reduces the overall stability below the threshold compatible with soluble expression. At the same time, it cannot be excluded that insertion of SpyCatcher into the extension is intrinsically problematic, independent of the cumulative destabilisation caused by the additional mutations. Previous results obtained for the insertion of SpyCatcher into the loop of a structural extension within the lower PBP lobe of LLhT also did not yield a successful construct (Chapter 2.5.3). The observed loss of soluble expression reflects the incompatibility of the LLhT-derived scaffold and SpyCatcher, rather than SpyCatcher being a universal trigger for insolubility when incorporated into structural extensions.

Regarding the thermostability of subunit B (LLhT-6Mut-DogCatcher), no difference relative to LLhT-6Mut was observed, with the apparent melting temperature remaining at 44.5 °C (SI Table 40). This data suggests that the two additional modifications, namely the altered DNA-binding domain and the fusion of DogCatcher to the C-terminus, either compensate for one another in terms of thermostability or do not negatively affect the construct's thermal stability. Notably, unlike the thermal unfolding curves previously obtained for variants carrying SpyCatcher at the C-terminus, the DogCatcher-containing variant did not exhibit a second unfolding transition (SI Figure 150).

The soluble construct LLhT-6Mut-DogCatcher was analysed further with respect to its oligomeric state, to analyse whether the monomeric state achieved by the six point mutations is still present after the addition of DogCatcher. Based on the six interface mutations, the variant was expected to remain predominantly monomeric. However, SEC-MALS analysis did not confirm a clearly monomeric state (Figure 75). LLhT-6Mut-DogCatcher (Figure 75, blue) displayed a largely monodisperse SEC peak, indicating that the sample did not separate into multiple clearly distinguishable oligomeric species under the conditions analysed. However, the elution volume was markedly shifted compared to LLhT-6Mut (Figure 75, purple), from around 15.5 mL to around 13.0 mL. While an earlier elution is expected to some extent due to

the increased molecular size conferred by the attached DogCatcher domain (10 kDa), the magnitude of the shift appears too large to be explained by a purely monomeric LLhT-6Mut-DogCatcher species alone. This is further supported by comparison with dimeric LLhT (Figure 75, grey), which elutes later, at 14.0 mL, despite being expected to have a larger molecular mass than monomeric LLhT-6Mut-DogCatcher. Consistent with this discrepancy in elution behaviour, the experimentally determined molecular weight of LLhT-6Mut-DogCatcher was 72.8 kDa. This value falls between the sizes expected for a monomeric species (46 kDa) and a dimeric species (92 kDa). Table 43 in the SI summarises the expected and determined molecular weights of the MALS measurements. Under the tested conditions, the MALS-derived molecular weight indicates that LLhT-6Mut-DogCatcher does not behave as a monomeric species.

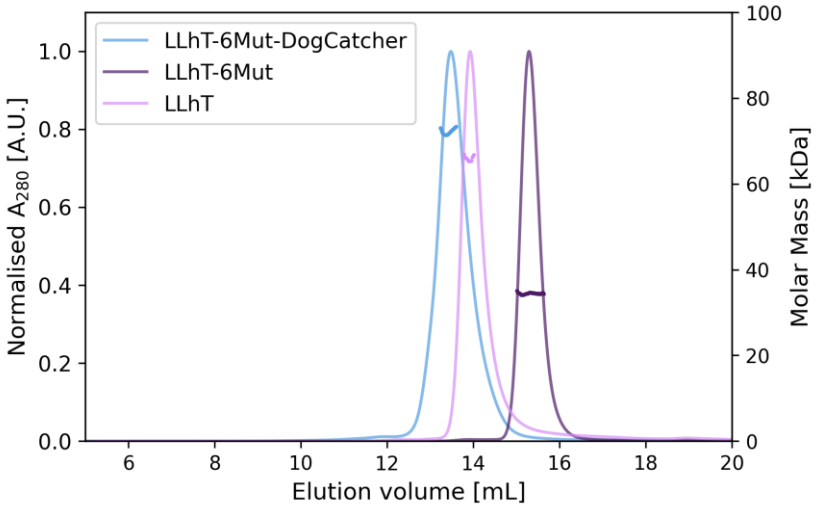


Figure 75. SEC-MALS characterisation of LLhT interface and DogCatcher fusion variants. SEC-MALS analysis was performed for LLhT (pink, expected MW: 78 kDa), LLhT-6Mut-DogCatcher (blue, expected MW: 46 kDa), and LLhT-6Mut (purple, expected MW: 36 kDa) using an analytical Superdex 200 Increase 10/300 GL column. Normalised absorbance at 280 nm is plotted against elution volume, with the corresponding MALS-derived molar mass shown on the secondary y-axis. Measurements were performed in Buffer C using a protein concentration of 0.5 mg·mL^{-1}.

SDS-PAGE analysis shows that no covalently linked higher-order species or contaminations were present (SI Figure 128). As the construct is similar to LLhT-6Mut, except for DogCatcher attached to the C-terminus, one possible explanation is that the DogCatcher domain itself contributes to self-association or otherwise perturbs the oligomeric equilibrium of the LLhT scaffold. A similar tendency may also be relevant to the clapping-machine constructs discussed in Chapter 2.6 and to the subunit of the heterodimer Het-DogTagxCatcher in Chapter 2.5.6. and will be further discussed there.

These results show that although the individual prerequisites for heterodimeric LLhT-based foot construction are established separately, their combination into a single construct proved substantially challenging. The C-terminal addition of DogCatcher might alter the oligomeric

behaviour of the monomerised LLhT scaffold, while simultaneous integration of the Lev7 extension, SpyCatcher, DogTag, and the six interface mutations abolished solubility altogether.

2.5.6 Assessment of SpyCatcher integration into the structural extension LacI-Lev1

The attachment of SpyCatcher to the structural extension of LLhT did not appear to be feasible (Chapter 2.5.3 and 2.5.5), as the corresponding constructs either exhibited higher oligomeric states or failed to express. To determine whether this limitation reflects a general incompatibility of SpyCatcher positioned at extensions within a PBP lobe, rather than a specific effect in the studied LLhT variants, SpyCatcher was incorporated into the LacI-A2-Link1 variant (Chapter 2.4.2) by replacing the Tag and Catcher modules. Starting from LacI-A2-Link1-SpyTag, DogTag was exchanged with SpyCatcher at the structural extension (Figure 76, subunit A). No further optimisation of linker sequences was performed. Instead, the genetic information encoding DogTag was directly exchanged with that encoding SpyCatcher while retaining the existing fusion architecture. To preserve the possibility of directed heterodimerisation in subsequent assembly steps, the C-terminal SpyTag was replaced by DogTag (Figure 76, subunit A). For the generation of a functional heterodimer, the subunit B was derived from LacI*-SpyCatcher by replacing SpyCatcher with DogCatcher (Figure 76, subunit B). The covalent connection to form the heterodimer Het-DogTagxCatcher is performed by the DogTagxCatcher module. SpyCatcher will be present for the assembly with the clapping machine (Chapter 2.7).

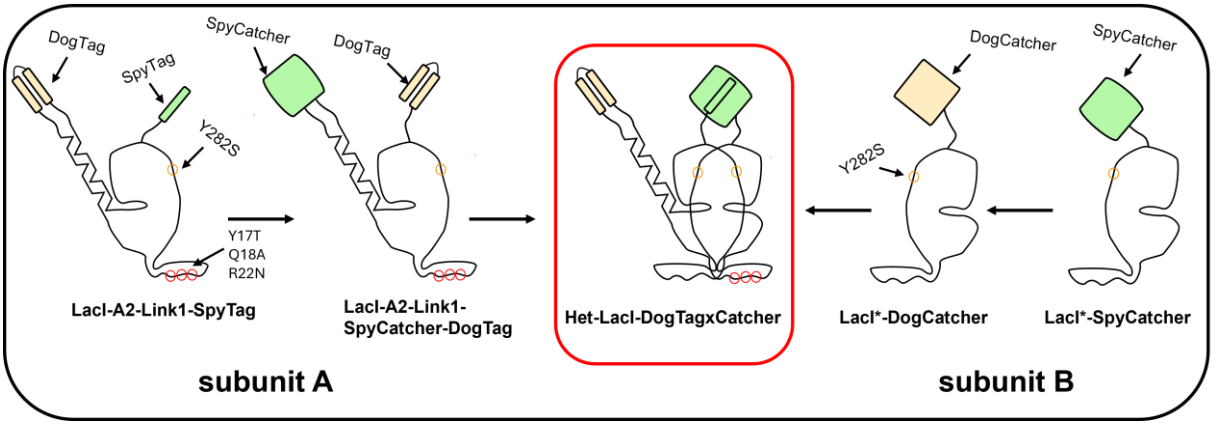


Figure 76. Schematic representation of the modular assembly of a LacI heterodimer mediated by DogTagxCatcher. Starting from LacI-A2-Link2-SpyTag and LacI*-SpyCatcher, the connector modules were exchanged to generate a DogTag/DogCatcher-based heterodimerisation strategy. DogTag and DogCatcher were used as C-terminal coupling modules to link the two LacI-derived subunits, whereas the DogTag located at the distal end of the structural extension of LacI-A2-Link1-SpyTag was replaced by SpyCatcher.

In contrast to the LLhT-derived constructs, subunit A and B were found in the soluble fraction after heterologous expression in *Escherichia coli* (SI Figure 129 and Figure 130), indicating that the incorporation of SpyCatcher into an α-helical bundle extension is possible. Although SpyCatcher was not originally intended for insertion into loop regions, the experimental results indicate that incorporation of SpyCatcher per se does not necessarily lead to insoluble expression.

The absence of LLhT-6Mut-Lev7-SpyCatcher-DogTag from the soluble fraction was attributed to a reduced stability caused by the introduced mutations (Chapter 2.5.5). However, it remained unclear whether the insertion of SpyCatcher into a structural extension generally decreases thermostability. Therefore, the change in thermal stability resulting from exchanging DogTag and SpyCatcher within the structural extension was analysed by monitoring the temperature-dependent change in molar ellipticity at 222 nm using CD spectroscopy (Figure 77). The addition of SpyCatcher did not appear to destabilise the construct. In comparison to LacI-Lev1, which lacks the DogTag connector element, the attachment of DogTag (LacI-A2-Link1-SpyTag) reduced the apparent melting temperature by only 1 °C (Table 15 and Chapter 2.4.1). A comparable behaviour was observed when DogTag was exchanged for SpyCatcher. In this case, no further destabilisation was detected (Table 15). The apparent melting temperature of LacI-A2-Link1-SpyCatcher-DogTag (45.7 °C) differed by only 0.5 °C from that of LacI-Lev1 (46.2 °C). These results indicate that the presence of SpyCatcher itself was most likely not the main reason for the failure of the LLhT-6Mut-Lev7-SpyCatcher-DogTag construct to remain soluble (Chapter 2.5.5). However, a deeper analysis, for example, by an exchange of SpyCatcher with other members of the Catcher/Tag family in the LLhT-6Mut-Lev7-SpyCatcher-DogTag construct would be required to confirm this conclusively.

In addition to showing that SpyCatcher does not further destabilise the LacI-based construct, the CD data revealed distinct unfolding behaviours for SpyCatcher- and DogCatcher-containing variants (Figure 77). If DogCatcher was attached to the C-terminus of LacI, the molar ellipticity remained at around −3000 deg·cm²·dmol⁻¹ and the transition midpoint is located at 46 °C (Figure 77, pink). In contrast, LacI-A2-Link1-SpyCatcher-DogTag showed an initial unfolding transition at 45.7 °C, but the molar ellipticity remained at −7000 deg·cm²·dmol⁻¹ even at temperatures up to 90 °C (Figure 77, purple). This suggests that a substantial fraction of secondary structure remains after the LacI-derived domain unfolds. The heterodimeric construct Het-DogTag×Catcher (Figure 77, orange) retained a molar ellipticity even more negative (−10000 deg·cm²·dmol⁻¹) after initial unfolding, followed by a second transition starting at 85 °C. Based on the thermal stability analysis of the covalently linked SpyTag×Catcher system (Chapter 2.2.4 and 2.4.2), this second transition is interpreted as the unfolding of the covalently linked DogTag×Catcher module, which appears to retain secondary structure until elevated temperatures.

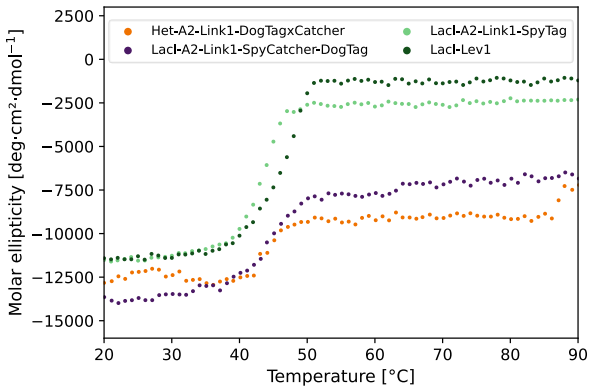


Figure 77. CD-based thermal unfolding analysis of LacI DogTagxCatcher heterodimer and its corresponding monomeric subunits. Thermal unfolding of LacI-A2-Link1-SpyCatcher-DogTag (purple), LacI*-DogCatcher (pink), Het-DogTag×Catcher (orange), LacI-Lev1 (green), and LacI-A2-Link1-SpyTag (light green) was monitored by the temperature-dependent change in molar ellipticity at 222 nm. Measurements were performed in Buffer C using a protein concentration of 0.2 mg·mL^{-1}.

Table 15. Apparent melting temperatures (T_{Melt}) of LacI DogTagxCatcher heterodimer and its corresponding monomeric subunits Values were determined from CD-based thermal unfolding experiments as the transition midpoint of the temperature-dependent change in molar ellipticity at 222 nm.

Construct	T_{Melt} [°C]
LacI-A2-Link1-SpyCatcher-DogTag	45.7
LacI*-DogCatcher	46.0
Het-DogTagxCatcher	46.6
LacI-Lev1	46.2
LacI-A2-Link1-SpyTag	45.5

A limitation that occurred when characterising LLhT-6Mut-DogCatcher was the oligomeric state of the variant (Chapter 2.5.5). To further assess the oligomerisation behaviour of DogCatcher in the context of the new construct, subunit A (LacI-A2-Lev1-SpyCatcher-DogTag) and subunit B (LacI*-DogCatcher) were purified separately, without prior combination of the lysates (SI Figure 129 and Figure 130). After Ni^{2+}-IMAC, neither construct eluted as a monodisperse species during SEC. For LacI-A2-Lev1-SpyCatcher-DogTag, a shoulder was observed at 130 mL (SI Figure 129, blue box). In contrast, LacI*-DogCatcher showed a clear separation into two pronounced peaks at 130 mL (blue box) and 150 mL (pink box) (SI Figure 130). This already indicates that DogCatcher may promote the formation of higher oligomeric states. Interestingly, and in contrast to LLhT-6Mut-DogCatcher, LacI*-DogCatcher showed a clear separation of two species, whereas the LLhT-based construct eluted as one peak (Chapter 2.5.5).

To further determine the oligomeric states of these species, SEC-MALS measurements were performed for the different SEC peaks (Figure 78). For both LacI-A2-Link1-SpyCatcher-DogTag (Figure 78A) and LacI*-DogCatcher (Figure 78B) samples collected from the shoulder at 130 mL or the first peak during the purification were referred to as DogTag-1stPeak or DogCatch-1stPeak (Figure 78, blue), whereas samples from the later eluting fractions are referred to as DogTag-2ndPeak or DogCatch-2ndPeak (Figure 78, pink). Two peaks around 12 mL and 14 mL were observed for DogTag-1stPeak, with the latter eluting peak being the dominant species. Two peaks at similar elution volumes were also observed for DogTag-2ndPeak. The peak at 14 mL retains the dominant peak, but the first peak is smaller than in DogTag-1stPeak. The early-eluting species is consistent with a population of dimeric species

with an apparent molecular size of around 100 kDa, whereas the later peak corresponds to the monomeric species (50 kDa). Thus, although the shoulder fraction was expected to exhibit a large fraction of dimeric species, the sample remained predominantly monomeric after reinjection. Showing that DogTag might lead to a concentration-dependent monomer-dimer equilibrium and is mostly found in the monomeric oligomerisation state at 0.5 mg·mL^{-1}.

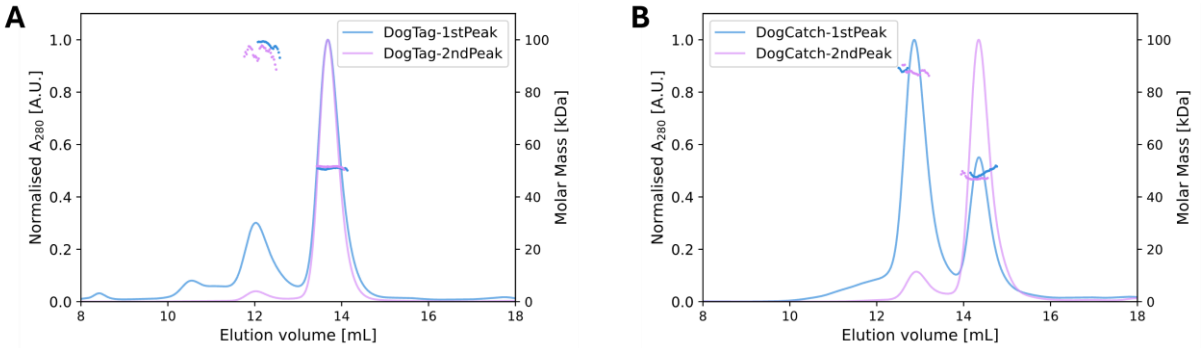


Figure 78. SEC-MALS analysis of LacI-A2-Link1-DogTag and LacI*-DogCatcher. Measurements were performed using an analytical Superdex 200 Increase 10/300 GL column. Normalised absorbance at 280 nm is plotted against elution volume, and the corresponding MALS-derived molar mass is shown on the secondary y-axis. Measurements were performed in Buffer C using a protein concentration of 0.5 mg·mL^{-1}. A) SEC-MALS analysis of two peaks separated during the final SEC purification step of LacI-A2-Link1-DogTag (SI Figure 130) (first peak, blue; second peak, pink). B) SEC-MALS analysis of two peaks separated during the final SEC purification step of LacI*-DogCatcher (SI Figure 129) (first peak, blue; second peak, pink).

A different behaviour was observed for LacI*-DogCatcher. The sample DogCatcher-1stPeak showed a dominant peak at around 13 mL, together with a smaller peak at around 15.0 mL. This indicates that part of the isolated dimeric fraction dissociated into monomers during SEC-MALS analysis. In contrast, SEC-MALS of the DogCatcher-2ndPeak revealed the opposite distribution. This sample showed a dominant later-eluting peak at 15 mL (50 kDa), consistent with a monomeric species (48 kDa), and only a minor early-eluting signal around 13 mL (93 kDa). This indicates that, even at a lower concentration of 0.5 mg·mL^{-1}, DogCatcher tends to remain in the dimeric oligomerisation state. Moreover, even though partial dissociation of the isolated dimeric fractions into monomeric species was observed, a substantial proportion of dimeric protein remained. Together with the results obtained for LLhT-6Mut-DogCatcher (Chapter 2.5.5) and CM-PotF (2.6), these data suggest that DogCatcher-containing constructs tend to form dimeric species. However, the extent and stability of this behaviour differ markedly between constructs. For LLhT-6Mut-DogCatcher, no clear separation between monomeric and dimeric species was achieved (SI Figure 128), whereas LacI*-DogCatcher (SI Figure 130) and CM-PotF (SI Figure 132) showed distinct monomeric and dimeric fractions.

The formation of higher oligomeric species is undesirable for the construction of artificial motor proteins, as defined oligomeric states are required for controlled assembly and function. Importantly, however, no additional higher-order oligomeric species were observed once DogTag and DogCatcher were covalently linked. This is demonstrated by the precisely defined

oligomeric state of the heterodimeric Het-Link1-DogTag×Catcher construct (Figure 79), as well as by the half-assembly between foot B and the clapping machine (Chapter 2.7). Thus, while isolated DogCatcher-containing modules may show DogCatcher-dependent dimerisation, covalent DogTag/DogCatcher ligation appears to stabilise the intended assembly state and prevents further oligomeric rearrangement. Therefore, the use of DogCatcher as a connecting module remains feasible without imposing constraints on the motor protein design.

Beyond providing insight into possible limitations associated with the design of the LLhT-based variants, the assembly of subunit A (LacI-A2-Link1-SpyCatcher-DogTag) and subunit B (LacI*-DogCatcher) are also valuable for addressing the broader conceptual question: whether different Tag/Catcher families can be used interchangeably for the assembly of the motor protein, meaning that formation of a heterodimeric LacI can be performed by SpyTagxCatcher as well as DogTagxCatcher. Furthermore, SpyTag and DogCatcher were attached to the clapping machine to enable analysis of specific interactions between the feet and their corresponding partner domains positioned at the structural extension. The availability of Het-SpyTagxCatcher (foot A) and Het-Link1-DogTagxCatcher (foot B) enables the evaluation of whether irreversible isopeptide bonds with the clapping machine can be formed in a specific and controlled manner.

To test whether SpyTagxCatcher and DogTagxCatcher can be functionally exchanged, the lysate of the monomeric variants LacI*-DogCatcher and LacI-A2-Link1-SpyCatcher-DogTag was combined overnight after heterologous expression in *Escherichia Coli* to examine whether heterodimer formation remains possible when the Spy system is replaced by the Dog system at the C-terminus. SDS-PAGE confirmed the successful formation of the isopeptide bond between DogTag and DogCatcher, as the construct was detected at a height corresponding to 100 kDa (Figure 79A). SEC-MALS analysis further supported heterodimer formation (Figure 79B). The majority of the sample eluted as a predominantly monodisperse peak with an experimentally determined molecular weight of 99.5 kDa. This value deviates by only 4 kDa from the theoretical molecular weight (103 kDa) and therefore agrees well with the expected heterodimeric species. Notably, this deviation is smaller than that observed for the corresponding Link1-derived heterodimer assembled via the Spy system, Het-Link1-SpyTag×Catcher, for which the difference between measured and expected molecular weight was 10 kDa (Figure 79B, dark grey), but comparable with the deviation from the expected molecular weight obtained for Het-Link2-SpyTag×Catcher (Figure 79B, grey). SpyTagxCatcher heterodimers were discussed in detail in Chapter 2.4.2. Additionally, a small shoulder at a higher elution volume (11 mL) was observed.

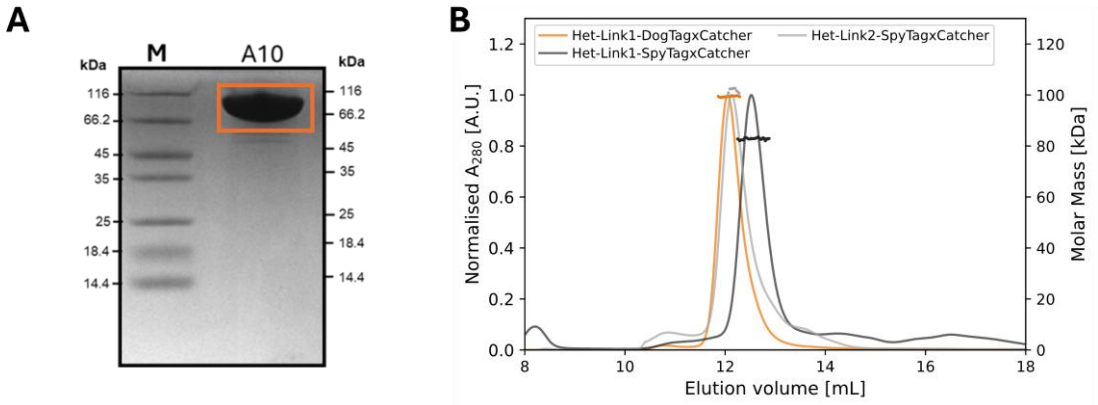


Figure 79. Validation of Het-Link1-DogTag×DogCatcher formation by SDS-PAGE and SEC-MALS. A) SDS-PAGE analysis of the main peak fraction obtained after analytical SEC of Het-Link1-DogTag×DogCatcher using an analytical Superdex 200 10/300 GL column (SI Figure 131). A band at 100 kDa is consistent with the formation of the covalently linked heterodimer. B) SEC-MALS analysis of Het-Link1-DogTag×DogCatcher (orange, expected MW: 103 kDa) compared with the SpyTag×Catcher-linked heterodimers Het-Link1-SpyTag×Catcher (dark grey, expected MW: 93 kDa) and Het-Link2-SpyTag×Catcher (light grey, expected MW: 92 kDa). Normalised absorbance at 280 nm is plotted against elution volume, and the corresponding MALS-derived molar mass is shown on the secondary y-axis. Measurements were performed in Buffer C using a protein concentration of 0.5 mg·mL^{-1}.

This data indicates that heterodimer formation via the DogTag/DogCatcher system is feasible. Functional characterisation of ligand binding revealed a behaviour comparable to that of the other heterodimeric LacI variants (Chater 2.4.2). Ligand affinity was moderately reduced by around 2 µM, but remained in a functionally relevant range, with an estimated K_D of 7 µM (SI Figure 140). The small decrease in affinity might be due to a larger attachment, DogCatcher instead of SpyTag, at the structural extension of LacI-Lev1.

Since the previously generated heterodimeric variants Het-Link1-SpyTag×Catcher and Het-Link2-SpyTag×Catcher no longer showed detectable DNA binding (Chapter 2.4.2), it was important to assess whether this loss of DNA affinity was also observed after replacing the heterodimerisation module. The DNA-binding properties of the DogTag/DogCatcher-based heterodimer were therefore analysed to determine whether the exchange of the covalent coupling system affected DNA interaction. No clear binding was detected even at elevated protein concentrations of 3 µM. However, compared with the SpyTagxCatcher heterodimers (Chapter 2.4.2), a small increase in the response unit is observed. However, the change in the response is too small to be attributed to binding to asymmetric DNA (Figure 80). This finding suggests that the loss of asymmetric DNA-binding capability is unlikely to arise specifically from the choice of the Tag/Catcher family. Instead, it more likely results from the combined architectural perturbation introduced by the heterodimeric interface together with the flexible linker connecting the monomer and the Tag/Catcher parts. However, it is also shown that small modifications can already alter DNA-binding affinity, as DogTagxCatcher exhibits a small interaction with DNA.

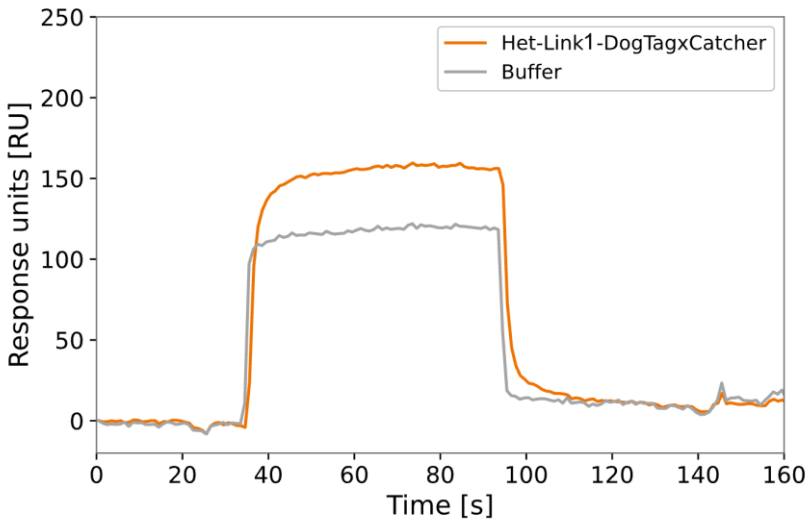


Figure 80. SPR analysis of Het-Link1-DogTagxCatcher binding to asymmetric DNA operator sequence. Sensorgrams were recorded at room temperature with a constant flow rate of 30 µL·min^{-1}. Het-Link1-DogTagxCatcher was injected at a concentration of 3 µM over immobilised double-stranded oligonucleotides containing different operator arrangements. Measurements were conducted by Vanessa Troßmann. Protein-containing sample curves are shown in orange, while buffer/reference curves are shown in grey.

Biochemical analysis of Het-Link1-DogTag×Catcher showed largely similar results to those obtained with its SpyTag×Catcher counterpart. Heterodimer formation was retained, and neither thermal stability nor ligand affinity was substantially altered. A major limitation of Het-Link1-SpyTag×Catcher is its limited interaction with DNA, which was not restored by replacing the heterodimerisation module with DogTagxCatcher. Thus, the central question can be answered as follows: SpyTag/Catcher and DogTag/Catcher systems are interchangeable in this context, but their exchange induces only limited changes in the relative orientation of the LacI subunits and is therefore insufficient to improve DNA binding.

2.6 Design and biochemical characterisation of the PotF-based clapping machine

A wide variety of proteins undergo conformational changes in response to external stimuli. However, the extent and nature of these structural rearrangements differ substantially across systems. For the design of the proposed non-autonomous walker (Chapter 2.1.2.1), proteins exhibiting conformational changes between different ligand-binding states appear particularly suitable as clapping-machine modules. One promising candidate in this context is PotF. Its open and closed conformations have been studied extensively, and PotF has furthermore been shown to provide a suitable starting point for protein engineering approaches.[121,122,444] Chapter 1.2.1 additionally highlights the advantages of PotF for the intended application.

To enable the covalent coupling of foot A and foot B to the clapping machine, a modular assembly approach was adopted. This strategy permits the specific and sequential attachment of the two feet of the motor protein, without the need to design and express a single-chain construct with a molecular weight greater than 200 kDa. By shifting the assembly of the full system to an *in vitro* ligation step, it also provides greater flexibility to combine different modules and study and introduce modifications to individual parts when required.

To modify PotF as a functional attachment point for both feet, several design questions, including the optimal placement of the connector sites, the linker length, and the structural properties, need to be addressed. Although internal positions within PotF may undergo larger relative displacements than the terminal regions, the N- and C-termini were selected as attachment points. In the closed conformation, these termini are located 12 Å closer together than in the open state, making them attractive candidates for transmitting ligand-induced motion to attached modules (Figure 81A). As the connector domains (DogCatcher and SpyTag) chosen for attaching the feet to PotF are not optimised for insertion into internal secondary-structure elements, the termini represent the most accessible and least disruptive positions for fusion. To simplify the design of the clapping machine, the linker length and structural properties were chosen conservatively. Although rigid or structurally defined linkers are assumed to be more suitable for transmitting conformational changes, as discussed for the foot extensions (Chapter 2.3) and for the general motor concept (Chapter 2.1.1), the first design of the clapping machine focused primarily on establishing PotF as an attachment platform. A $(GS)_4$-linker was introduced at both termini to connect PotF to the connector modules. This strategy was not intended to provide optimal mechanical coupling, but rather to minimise potential failure modes such as impaired folding or reduced expression. Since the complete motor function could not yet be assessed due to the loss of DNA-binding by the foot modules, the simpler, more robust flexible-linker design was selected for this initial proof-of-concept analysis. The resulting construct, in which DogCatcher is attached to the N-terminus and SpyTag to the C-terminus, was termed clapping machine-PotF (CM-PotF) (Figure 81B).

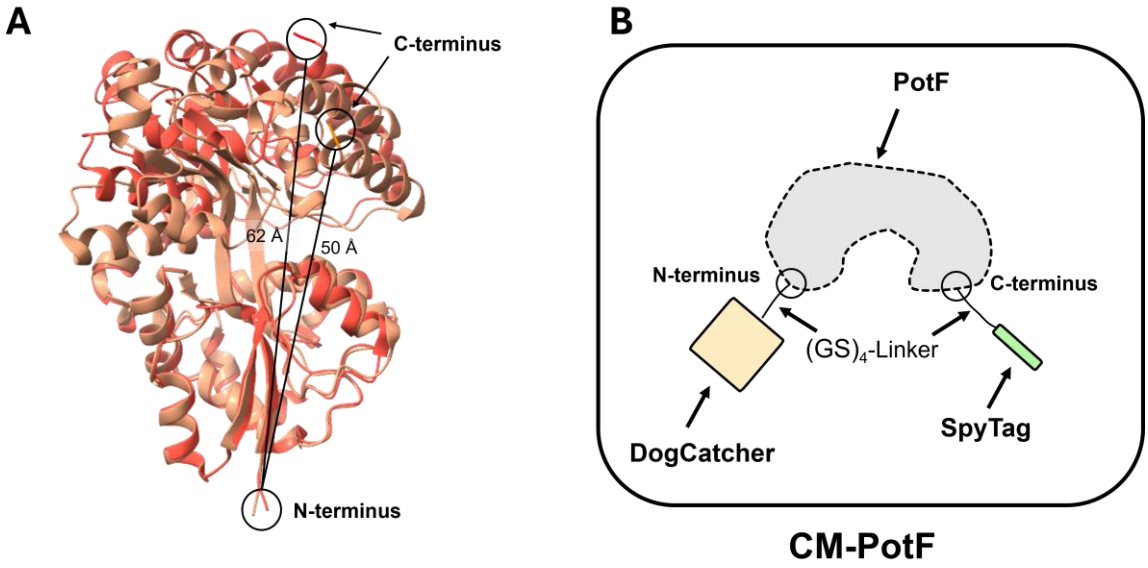


Figure 81. Conceptual and structural basis of the proposed clapping machine. A) Structural representation of the modified PotF for use as a clapping machine. B) PotF shown in two conformational states, closed (PDB: 6YEB[122], light orange) and open (PDB:6YED[122], dark orange) conformation. The superposition highlights a pronounced change in the distance between the N- and C-termini, which is also shown by measurements of the distance between the termini. B) Schematic illustration of the intended architecture of the clapping machine.

After heterologous expression of the resulting clapping-machine construct, CM-PotF, in *Escherichia coli* the protein of interest was present in the soluble fraction. Purification with Ni^{2+}-IMAC and SEC was successful (SI Figure 132). However, during the SEC purification, two distinct species were observed (Figure 82A, green and orange). Because both species displayed the same apparent molecular weight of around 55 kDa under denaturing SDS-PAGE conditions (Figure 82B), the most plausible interpretation is that they represent different oligomeric states rather than degradation products or truncated forms.

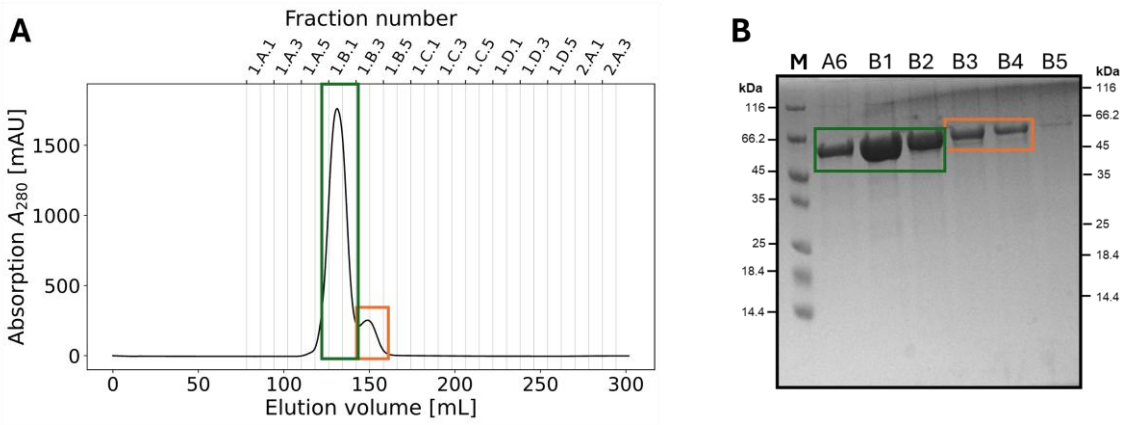


Figure 82. SEC and SDS-PAGE analysis of the PotF-based clapping-machine construct. A) SEC analysis of purified CM-PotF (MW: 55 kDa), showing two separated chromatographic species (green and orange). SEC was performed using a Superdex 75 Increase 10/300 GL column. B) SDS-PAGE analysis of the corresponding fractions. The similar apparent molecular weights of the bands indicate that the two SEC species differ primarily in oligomeric state.

Compared to unmodified PotF, which itself appears to give rise to both monomeric and dimeric species during purification (SI Figure 168), the proportion of the higher-molecular-weight

species was increased in the case of CM-PotF. This shift may reflect an increased tendency toward self-association due to the DogCatcher moiety, as discussed for foot B (Chapter 2.5.6). The second peak (Figure 82A, orange) was further analysed regarding its apparent molecular weight by SEC-MALS measurements (Figure 83, orange). SEC-MALS was used to assess whether the isolated species indeed remains monomeric under the analysed conditions or whether higher-order species are still populated.

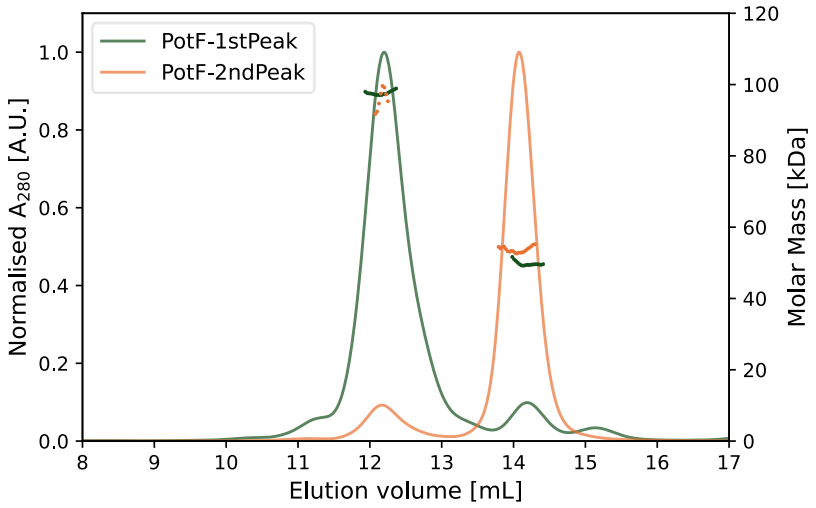


Figure 83. SEC-MALS analysis of the two CM-PotF species separated during purification The first and second SEC peaks of CM-PotF (expected MW: 55 kDa) were analysed separately using an analytical Superdex 200 Increase 10/300 GL column and are shown in green and orange, respectively. Normalised absorbance at 280 nm is plotted against elution volume, with the MALS-derived molar mass shown on the secondary y-axis. Colour coding corresponds to the highlighted SEC peaks in Figure 82. Measurements were performed in Buffer C using a protein concentration of 0.5 mg·mL^{-1}.

The analysis revealed two species, with the later-eluting peak at around 14 mL being more prominent than the earlier peak at around 12 mL (Figure 83, orange). The corresponding molecular weights were determined to be 95 kDa and 57 kDa, respectively (SI Table 44). This enables the assignment of the latter peak to monomeric CM-PotF (55 kDa). The peak at 12 mL corresponds to the dimeric form of CM-PotF. Although the experimentally determined molecular weight of 95 kDa deviates by around 15 kDa from the calculated dimeric mass of 110 kDa, this discrepancy is most likely caused by the low protein concentration across the peak at 12 mL, which reduces the accuracy of the molecular-weight determination by SEC-MALS. It should be noted that the preparative purification was performed on a Superdex S75 column, whereas the SEC-MALS was conducted using a Superdex S200 column. The improved separation range of the S200 column for species in this molecular-weight regime likely explains why two distinct peaks, corresponding to dimeric and monomeric CM-PotF, could be resolved in the SEC-MALS experiment. And the results do not support the possibility that monomeric CM-PotF at a concentration of 0.5 mg·mL^{-1} exhibits concentration-dependent dimer formation.

SEC-MALS measurements of the separately collected peak fraction at an elution volume of 130 mL (Figure 82, green) were conducted (Figure 83, green). In this sample, hardly any monomeric species (55 kDa) were detected, while the dimeric species remained dominant (110 kDa). This provides an additional indication that, once formed, the oligomeric states are stable at 0.5 mg·mL^{-1} as well. A definitive conclusion regarding a potential monomer-dimer equilibrium would, however, require a concentration-dependent analysis of CM-PotF. Within the concentration range analysed here, the monomeric fraction remained predominantly monomeric, supporting its use for subsequent assembly reactions at this concentration regime. Thus, later assembly experiments can be carried out with a monomeric clapping-machine population (Chapter 2.7).

For application in a functional clocked motor based on biased diffusion, the clapping machine must remain responsive to its ligand, putrescine. To assess the responsiveness DSC was performed, analysing the changes in thermal stability upon ligand interaction (Figure 84).

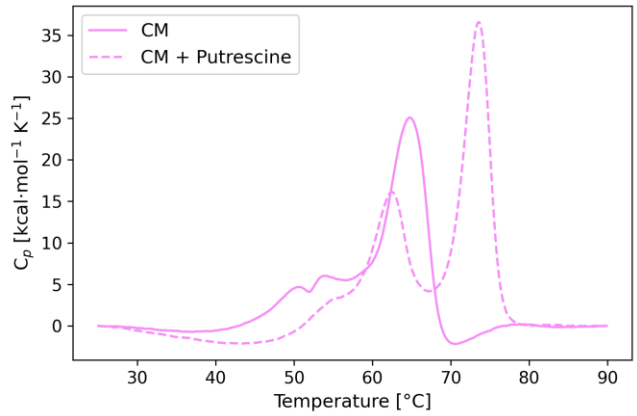


Figure 84. DSC analysis of the clapping-machine construct in the absence and presence of putrescine. Excess heat capacity (*C*p) is plotted as a function of temperature without ligand (solid line) and after addition of putrescine (dashed line). Measurements were performed in Buffer C using a protein concentration of 0.5 mg·mL^{-1}.

Table 16. Thermal transition temperatures of the PotF-based clapping-machine construct. The listed transition temperatures were derived from DSC profiles recorded both in the absence of ligand and in the presence of putrescine.

Construct	T_{Melt} [°C]
CM-PotF	50.7
	54.0
	64.8
CM-PotF + Putrecine	54.0
	62.4
	73.5

In the absence of the ligand putrescine, the DSC profile reveals several thermal transitions (Figure 84, solid line), consistent with a fusion construct comprising multiple domains with distinct intrinsic stabilities. The positions of the transition peaks are summarised in Table 16. Based on comparison with the melting temperature determined by circular dichroism spectroscopy, the peak at 64.8 °C is most likely attributable to PotF.[445] However, unequivocal assignment of the individual peaks to SpyTag, DogCatcher, and PotF would require additional thermal stability analyses of the isolated components. The observed ligand-induced shift in multiple DSC transitions (Figure 84, dashed line) suggests that putrescine binding affects the stability of the overall construct rather than stabilising only a local region. Since DSC transitions of multidomain proteins often correspond to cooperative or partially overlapping unfolding

events, individual peaks cannot necessarily be assigned to single domains. Therefore, the shift of several transitions is consistent with a ligand-induced stabilisation of the PotF-containing clapping-machine architecture and with thermodynamic coupling between the ligand-binding domain and adjacent structural elements. Putrescine binding likely stabilises the closed conformation of PotF and thereby alters interdomain contacts, conformational dynamics, or the relative stability of partially folded intermediates. As a result, the stabilising effect of ligand binding is propagated beyond the local binding site, affecting several unfolding transitions of the multidomain construct. The shift of all peaks within the DSC profile does not imply direct ligand interactions with all present domains but supports that the domains are conformationally and thermodynamically coupled. Furthermore, the retained responsiveness of CM-PotF to putrescine is validated. It was shown that PotF tolerates functionalisation of its termini with Tag/Catcher modules while retaining its ligand-binding properties, which is crucial for later applications of the clapping machine. Additionally, a precise oligomeric state is crucial for achieving a precise bipedal architecture of the motor protein; therefore, using the monomeric fraction of PotF is crucial. Overall, the construct represents a suitable starting point for assembling the motor protein; further optimisations, especially regarding the connections, might be necessary to increase its efficiency.

2.7 Assembly of the motor

Due to the reduced DNA-binding affinity of the heterodimeric feet, the proposed motor cycle could not yet be tested experimentally. Further optimisation of the foot constructs will therefore be required before a DNA-interacting walker can be realised. Nevertheless, a central objective of this thesis was to establish a modular assembly strategy that enables the straightforward exchange and recombination of individual components. To assess whether this modularity had been successfully implemented, assembly of the feet and the clapping machine was analysed by SDS-PAGE and SEC-MALS (Figure 86). This work established the design of the basic construct required to assemble the motor protein architecture proposed for a LacI-based motor protein (Figure 85C).

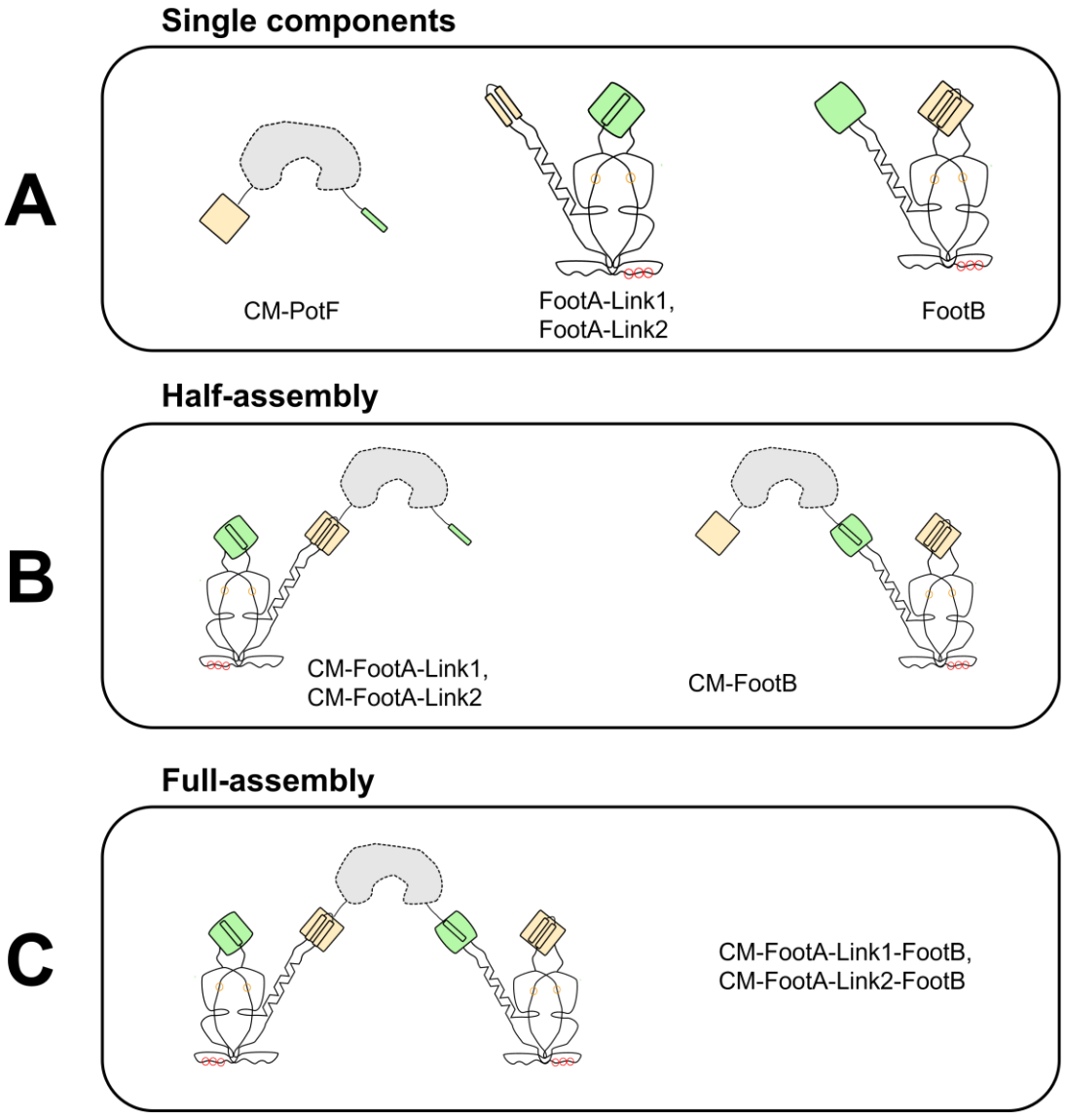


Figure 85. Schematic representation of the annealing steps. A) Basic building blocks (clapping machine (CM-PotF), Foot A variants (FootA-Link1 and FootA-Link2), and Foot B.) for an artificial motor protein. B) Partial assembly of one foot module with the clapping machine, yielding CM-FootA-Link1, CM-FootA-Link2, and CM-FootB. C) Complete motor assembly combining the clapping machine with both foot modules, yielding CM-FootA-Link1-FootB and CM-FootA-Link2-FootB.

The motor comprises two main building blocks: the LacI-derived feet (Chapter 2.4 and Chapter 2.5) and the ligand-responsive clapping machine (Chapter 2.6). In the design, the two feet differ in the connector systems used for foot heterodimerisation and for subsequent attachment to the clapping machine. Foot A is first generated as a SpyTag/SpyCatcher-mediated heterodimeric LacI foot module, leaving a DogTag within its structural extension available for coupling to the clapping machine (Figure 85A). In contrast, foot B is assembled through DogTag/DogCatcher-mediated heterodimerisation, while a SpyCatcher positioned in its structural extension remains available for attachment to the clapping machine (Figure 85A). Consequently, foot A is intended to connect to the clapping machine via DogTag, whereas foot B is intended to connect via SpyCatcher (Figure 85B).

The motor protein system was analysed for two foot A variants (Het-Link1-SpyTagxCatcher and Het-Link2-SpyTagxCatcher), one Foot B variant (Het-DogTagxCatcher), and the clapping machine construct (CM-PotF) (Figure 85), all of which have been characterised in detail (Chapter 2.4.2, Chapter 2.5.6 and Chapter 2.6). The monomeric fraction of the clapping machine was used for the assembly. The SEC-MALS results shown for the individual components again highlight that heterodimerisation of the feet is possible (Figure 86B), however, the feet themselves already exhibited minor peaks corresponding to higher-order oligomeric species eluting at around 11 mL, as discussed in Chapter 2.4.2. However, these species were not covalently linked, as no higher-oligomeric species were observed on SDS-PAGE for the single constructs (Figure 86A).

Initial analysis of the half-assemblies, comprising one foot module and one clapping machine, showed that all tested foot variants reacted with CM-PotF. This was reflected by a clear reduction of the peak corresponding to the unreacted foot module at 12 mL and an increase of the peak at around 11 mL, which is also supported by increase of the molecular weight to around 145 kDa. However, particularly for Het-Link2-SpyTag×Catcher (Figure 86C, grey) and Het-Link1-DogTag×Catcher (Figure 86C, orange), an additional elution peak at 10 mL was observed. Consistently, SDS-PAGE revealed additional bands above 250 kDa that were absent in the individual component samples. These findings suggest that a fraction of the components may undergo non-specific or over-stoichiometric reactions, potentially leading to the attachment of more than one foot module to a single clapping machine. In addition, a pronounced peak at 14 mL remained, which can be assigned to unreacted CM-PotF and is caused by the non-equimolar ratios (1:1.5 foot to CM-PotF) used during assembly. Taken together, the results indicate that half-assembly formation occurs efficiently across all variants, while some construct combinations also yield higher-molecular-weight side products.

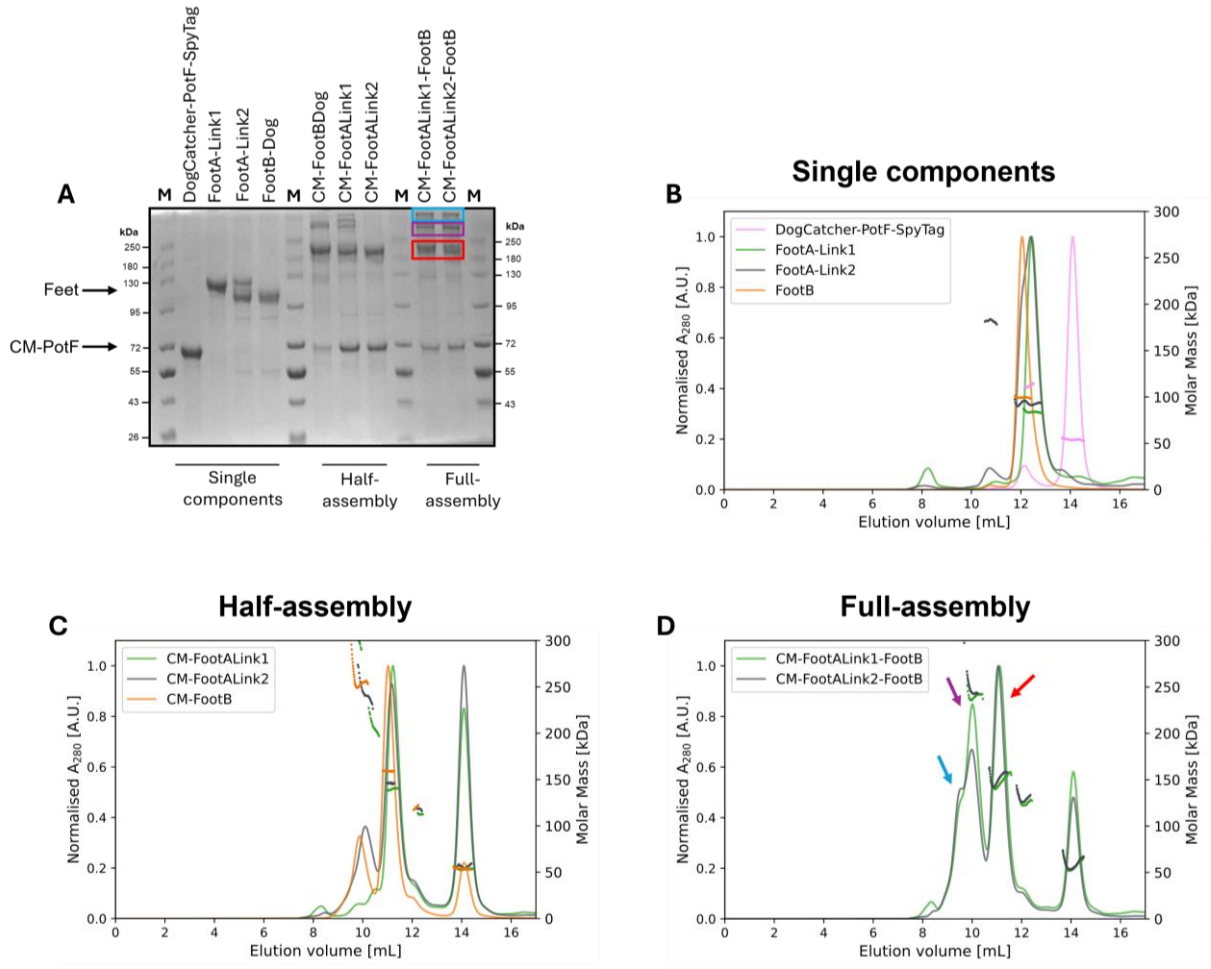


Figure 86. SDS-PAGE and SEC-MALS analysis of stepwise motor assembly. A) SDS-PAGE analysis of the individual constructs and assembly reactions. Highlighting the size of the individual components as well as the off-pathway interactions occurring for the full assembly state. B-D) SEC-MALS analysis of different assembly states performed using an analytical Superdex 200 Increase 10/300 GL column: B) individual constructs (expected MW: ~100 kDa and 55 kDa), C) partial assemblies (expected MW: ~150 kDa), and D) complete motor assemblies (expected MW: ~250 kDa). SEC-MALS measurements were performed in Buffer C using a protein concentration of 0.5 mg·mL^{-1}.

Assembly of the complete construct, consisting of one foot A variant, foot B, and the clapping machine, resulted in a further shift in both hydrodynamic radius and molecular weight, consistent with formation of the full motor assembly (Figure 86D). The experimentally determined molecular weight of 250 kDa for the increasing peak at 10 mL (magenta) is consistent with the expected size of a fully assembled motor protein (~260 kDa). At the same time, the data indicate that a proportion of the sample remained trapped at the half-assembly stage. This is supported by the SEC-MALS profile, in which the signal corresponding to the intermediate species (elution volume 11 mL) remained pronounced (Figure 86D, red arrow), and by the SDS-PAGE, where bands assigned to half assembly states were still clearly visible (Figure 86A, red box). In addition to the bands in the SDS-PAGE, which are assigned to the full assembly of the motor (Figure 86A, magenta box), higher-molecular-weight bands appeared during full assembly compared to half assembly (Figure 86A, blue box). Notably, the

gel pattern suggests the presence of several high-molecular-weight species, which is supported by the SEC-MALS data, which show an additional shoulder at 9 mL (Figure 85D, blue arrow) rather than a defined peak for the full assembled clapping machine.

Overall, these results demonstrate that assembly of the motor protein components is feasible, although some non-specific association appears to occur when two feet and the clapping machine are combined. This represents an important step towards constructing the basic framework of a bipedal walker and highlights the potential of the modular design strategy. At the same time, the data indicate that further optimisation will be required during the assembly process itself. In particular, improvements may be achieved by adjusting assembly stoichiometries and introducing additional purification steps to enrich the desired final product.

3 Discussion

3.1 Context dependence and modular control of engineered LacI heterodimerisation

3.1.1 Divergence between *in vivo* assignment and *in vitro* behaviour

Although re-engineering the LacI monomer-monomer interface has led to results that are not in accordance with the formation of a heterodimeric species *in vivo*,[68] the *in vitro* data did not support stable heterodimer formation through interface modification alone (Chapter 2.2.1). A likely explanation is that subunit B, which was intended to establish a novel interface with monomeric subunit A, still retained a substantial tendency to homodimerise. Residues C281 and Y282 have been identified as key determinants of LacI dimerisation, both in the literature[106,110,442] and in the present study. However, in the variants LacI-B1, LacI-B2 and LacI-B3, the additional interface mutations neither induced complete monomerisation nor promoted the expected heterodimerisation. One possible explanation is that some of the substituted residues occur naturally in other members of the LacI/GalR family, such as a lysine at position 255 in LacI-B3, which is naturally occurring in TreR or phenylalanine at position 282 of variant LacI-B1, which is also present in TreR or RbsR,[106] indicating that these positions might form an alternative monomer-monomer interface rather than necessarily destabilising it. A second explanation is the pronounced plasticity of the LacI dimer interface. A previous work has shown that the monomeric LacI variant Y282D recovers dimerisation through compensatory second-site mutations, including substitutions in the immediate structural environment of residue 282.[106] Since the four additional substitutions introduced in variant B are located close to this position, it appears plausible that they partially restored productive intermolecular contacts, even though the exact substitution at position 282 differs from that analysed by Swint-Kruse and coworkers.[106] Thus, rather than selectively enforcing heterodimerisation, the redesigned interface may have generated an alternative interaction surface still compatible with homodimer formation.

It should further be noted that not all variants reported by Daber *et al.*[68] were analysed in the present work, and the data presented here, therefore, does not invalidate the overall concept of the underlying *in vivo* study. From the number of variants tested, the current results suggest that the heterodimeric state assigned *in vivo* may depend strongly on the physicochemical context of the cell and may not be preserved after purification. In the cellular environment, proteins are embedded in a highly crowded and heterogeneous medium, with total macromolecular concentrations reaching approximately 300 $g \cdot L^{-1}$ in *Escherichia coli*.[446] Under such conditions, proteins are subject to two broad classes of intermolecular effects: hard-core repulsions and soft interactions.[447] These transient, context-dependent interactions contribute to what has been termed quinary structure, namely the network of weak macromolecular contacts that exists in the crowded cellular interior, but is largely lost upon lysis and purification

into dilute buffer.[448] Hard-core repulsions arise from excluded-volume effects: two macromolecules cannot occupy the same space simultaneously, and therefore entropically favour compact conformations and assemblies.[448–450] In contrast, soft interactions comprise weak attractive and repulsive chemical contacts, including electrostatic, hydrophobic, water-mediated, and other surface-based interactions.[447,450–453] Although individually weak, these interactions become highly influential in the cellular context due to the extreme macromolecular concentration. Importantly, it is the balance between steric repulsion and weak chemical attraction or repulsion that shapes association equilibria between competing oligomeric states *in vivo*. An assembly that is favoured, stabilised, or even only transiently enriched *in vivo,* may no longer remain the dominant or even a detectable species *in vitro* due to changes in the intramolecular effects. In this sense, the discrepancy between the heterodimeric assignment reported *in vivo* and the present *in vitro* observations does not necessarily imply that the original observation was incorrect. Rather, it may reflect the loss of the cellular interaction landscape that supported, biased, or kinetically enabled formation of the heterodimeric state.

3.1.2 Distinct effects of splitGFP and SpyTag/SpyCatcher on engineered LacI heterodimerisation

In the present study, two different engineered split proteins were fused to the C-terminus to generate a heterodimerisation domain: splitGFP[154] and SpyTag/SpyCatcher[204]. Although both systems rely on self-assembly, they differ fundamentally in their mode of assembly and therefore in their impact on the underlying LacI scaffold. The association of the splitGFP halves is followed by irreversible fluorophore maturation.[454] In the case of splitGFP, heterodimerisation of the bipart system cannot be considered independent of attached scaffolds. Even though self-assembly occurs, the association rate (0.0042 μM^{-1} min^{-1}) is far below the diffusion limit.[455] To enhance the splitGFP assembly, two fusion partners should be in close spatial proximity and an appropriate relative orientation towards each other.

For LacI, only when the splitGFP system was attached to the C-terminus of LacI-A2 and LacI-B2, GFP maturation occurred. Those variants had previously been identified as heterodimers *in vivo*[68] but lost their heterodimeric oligomerisation *in vitro*. (Chapter 3.1.1) At first glance, it appears that LacI-A2 forms a heterodimer with LacI-B2, as evidenced by the *in vivo*-based BiFC assay. In this setting, the central question was whether splitGFP would preserve and reinforce the heterodimeric state formed during heterologous expression in *Escherichia coli* and in the absence of the cellular context after purification. Higher-order oligomerisation states have been identified for this variant LacI-splitGFP (Chapter 2.2.3). These results imply that although heterodimer formation is enhanced by splitGFP, it cannot prevent off-pathway formation of individual subunits.

Analysis of the variant LacI-splitGFP after removal of the splitGFP *in vitro* suggests that irreversible splitGFP maturation stabilises the LacI-A2/LacI-B2 heterodimer; however, heterodimer formation is not retained if the splitGFP is removed, and the two subunits LacI-A2 and LacI-B2 dissociate into their natural oligomerisation states. LacI-splitGFP interacts with both symmetric and asymmetric DNA, consistent with the presence of higher-order oligomeric assemblies detected by SEC-MALS. This result highlights that subunit LacI-B2 still exhibits residual homodimer formation in the absence and the presence of splitGFP. The results obtained for LacI-splitGFP highlight an important limitation of splitGFP in this context. While it serves as a useful heterodimer stabiliser and detection module, it is only suitable if the underlying protein scaffold already exhibits protein-protein interactions, due to its self-association rate of 0.0042 μM^{-1} min^{-1}.[455]

In comparison, SpyTag/SpyCatcher differ fundamentally in this regard. SpyTag/SpyCatcher functions as a proximity-generating heterodimerisation module, as it forms a spontaneous covalent isopeptide bond between the two fusion partners.[456] Heterodimerisation is initiated by the attached split protein to the C-terminal domains themselves and does not rely on prior selective association of the LacI scaffold. Given the near diffusion-limited association kinetics reported for optimised SpyTag/SpyCatcher variants,[192] this strategy provides a considerably more robust route to defined heterodimer formation than slow and geometry-dependent splitGFP complementation. This is a major conceptual advantage for the present engineering problem, as it removes the need to generate a fully selective heterodimeric LacI interface by redesign of the native monomer-monomer interface alone. Thus, monomeric LacI variants (e.g. LacI-A2) have been used as the starting construct for the attachment of SpyTag/SpyCatcher. The advantages of not requiring PPI for the interaction for specific and efficient interactions of the split proteins are reflected by the experimental results. SEC-MALS confirmed the presence of a discrete heterodimer, while DNA binding analysis with SPR showed selective interaction with asymmetric, but not symmetric, operator DNA. Together, these findings strongly support the conclusion that SpyTag/SpyCatcher-mediated assembly yielded a functionally defined heterodimer without the requirement of protein-protein interactions present in the underlying protein scaffold LacI and has major advantages over the use of a heterodimerisation domain in comparison to splitGFP.

3.2 RFdiffusion-guided structural extension of an allosterically active LacI domain

While RFdiffusion has already proven highly effective for *de novo* binder generation, motif scaffolding, and the design of new protein architectures, examples of its use to introduce structured extensions into existing proteins remain comparatively limited.[257,457–460] The available studies nevertheless illustrate two important points. On the one hand, RFdiffusion[257] and ProteinMPNN[252] have been used to generate rigid fusion modules for G protein-coupled receptors (GPCRs), demonstrating that *de novo* designed domains can be structurally integrated into a pre-existing protein framework.[461] On the other hand, recent work on minimal TIM barrels showed that RFdiffusion[257] can also be applied to generate comparatively small extensions, such as short α-helical additions, rather than only larger domains.[247] However, both cases differ fundamentally from the present study. In the GPCR context, the inserted element primarily serves as a structurally stabilising fusion partner, whereas in the TIM-barrel example, a highly stable scaffold that does not undergo major internal conformational rearrangements is involved.[461] For designing a LacI-based motor protein, a rigid connection between the feet and the clapping machine appears to be a key prerequisite for mechanical and functional communication between them. This work, therefore, focused on introducing a structured linker into the LacI backbone. A structural extension was designed using RFdiffusion[257] and ProteinMPNN[252]. The novelty of the present approach is not the addition of the structural element itself, but the introduction of the designed rigid extension into a functionally sensitive region. Therefore, structural compatibility and conformational coupling need to be satisfied simultaneously. The central challenge was not simply to obtain a folded extension, but to do so without abolishing the native functional properties of LacI, namely ligand binding, DNA binding, and allosteric responsiveness. In this respect, the results are highly informative. Ligand binding was retained for LacI-Lev1 and LacI-Lev2, despite the proximity of the extension of LacI-Lev1 to the ligand-binding pocket, indicating that the local architecture of the regulatory domain remained sufficiently intact to preserve this function, and that the binding site was not blocked. The designs did not abolish LacI function, but they were not functionally neutral either. Rather, the results indicate that insertion of a structured extension into the N-terminal effector subdomain (LacI-Lev1) weakens the interaction network required for tight DNA binding, while still preserving the overall functional integrity of the protein. The consequences for DNA binding are discussed in more detail in Chapter 3.4. But in principle, integrating a rigid structural element retained functionality.

The structural data for PBP1 provide additional support for the design's success. Most notably, the intended continuation of an internal LacI α-helix was structurally realised, resulting in a rigid protrusion emerging from the N-terminal PBP subdomain. This is important because it demonstrates that the RFdiffusion-guided design was able to extend a native secondary-structure element in a geometrically coherent manner within an allosteric protein context. In

addition, the crystal structure provided insight into the elongation of the conformational changes occurring upon ligand binding toward the structural element. The solved structure indicates that the structural element is rigid. Moreover, the different ligand states of chain A and B highlight that, depending on the absence or presence of a ligand, the structural element moves accordingly to the PBP domain. This observation supports a central design principle for the artificial motor protein: the transmission of conformational changes from the LacI scaffold to the coupling element and vice versa. This is particularly important because directional stepping requires that ligand-induced conformational changes in LacI or in the clapping-machine module be mechanically propagated to the feet. The coupling of movements from the clapping machine to the extension reduces the spatial search volume of a detached foot and thereby increases the likelihood of productive rebinding to the next binding site on the DNA track. In this way. The structures support the idea that the extension could not only serve as a passive attachment point but also contribute to converting local conformational changes into the mechanical bias required for directional motion.

The present study expands the demonstrated application space of RFdiffusion beyond binder design, scaffold diversification, or catalytic functionalisation towards the introduction of motion-relevant structural elements into dynamic allosteric proteins. The current successful design further makes LacI-Lev1 a suitable candidate for the underlying motor concept, given the conformational change of the structural extension with PBP domain motion.

3.3 Selective effector recognition as a basis for ligand orthogonality

Orthogonality of effector recognition is a central requirement for the design of a non-autonomous motor, as different feet and clapping modules must be addressable in a selective manner (Chapter 2.1.2.1). Natural regulatory systems demonstrate that such selectivity is achieved even in a highly crowded intracellular environment. In *Escherichia coli*, for example, the repertoire of DNA-binding transcriptional regulators has been estimated to be 314 repressors, illustrating the extent to which specific signal discrimination is required for robust cellular control.[462] To explore ligand combinations suitable for an orthogonally controlled motor design, LacI-based variants previously characterised *in vivo* were analysed *in vitro* for their ligand selectivity. These included, on the one hand, a LacI variant carrying three point mutations in the ligand-binding domain (LacI-Gentio), which is reported to recognise gentiobiose and on the other hand, a chimeric construct in which the PBP-like domain of LacI had been replaced with the one of TreR (LLhT).[25,94] The present results show that IPTG and T6P act orthogonally with respect to the effector domains of LacI and LLhT. In contrast, gentiobiose still exhibits measurable interaction with LacI, but does not with LacI-Gentio. As LacI-Gentio did not exhibit a measurable gentiobiose interaction, the allosteric response of gentiobiose was not examined in the present work. It is therefore shown that IPTG and

gentiobiose don't have a structurally constrained ligand space. Detailed information on whether gentiobiose is a neutral ligand, an anti-inducer, or an inducer is not possible. In this context, the limited discrimination of gentiobiose to LacI may be explained by the absence of selective pressure in *Escherichia coli* to discriminate strictly against this sugar,[463] whereas T6P is associated with a native regulatory system and therefore likely underwent more specific evolutionary partitioning.

In addition to the evolutionary conservation of ligand orthogonality, comparison of the ligand structures provides a useful framework for understanding the observed selectivity. Notably, the remaining compatibility of gentiobiose with LacI is consistent with its greater structural similarity to lactose-derived effectors (Figure 87).

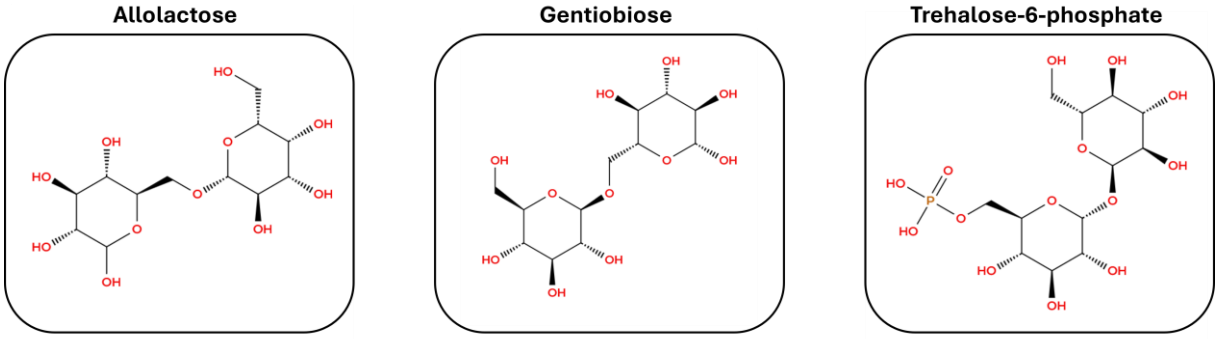


Figure 87. Structural comparison of disaccharide ligands of LacI and variants analysed in this thesis. Chemical structures of allolactose, gentiobiose, and T6P. Allolactose represents the natural inducer of LacI, whereas gentiobiose and T6P were used as alternative carbohydrate ligands to investigate ligand-dependent effects on LacI-derived constructs and TreR-derived effector domains. The comparison highlights the different glycosidic connectivities and functional groups that may influence ligand recognition and allosteric regulation.

The physiological LacI inducer allolactose is a β-(1-6)-linked galactose-glucose disaccharide, whereas gentiobiose is a β-(1-6)-linked glucose-glucose disaccharide. By contrast, T6P is an α,α-1,1-linked glucose-glucose disaccharide that additionally carries a phosphate group at C6. Thus, gentiobiose is topologically much closer to allolactose than T6P. Allolactose and gentiobiose share a β-(1-6) topology but differ in epimerisation at C4. It seems that the epimerisation at C4 is not sufficient to abolish the interaction with LacI. Raffinose provides a useful counterexample: it contains a 1-6-linked galactose-glucose substructure. It acts in other bacterial sugar-regulatory systems, yet it has been reported not to act as an inducing ligand in *Escherichia coli*.[464,465] Thus, the presence of a 1-6-linked galactosidic motif alone cannot account for productive LacI recognition. In contrast to raffinose, gentiobiose preserves a free O6 hydroxyl group which is also present in allolactose. Structural studies showed that the O6 hydroxyl plays a key role in LacI induction by supporting a water-mediated hydrogen-bonding network between the N- and C-terminal core subdomains.[30,67] At the same time, it is clearly not the sole determinant of ligand binding, because the anti-inducer ONPF lacks the O6 hydroxyl yet still binds to LacI.[67] This suggests that the O6 position is important for productive allosteric

regulation and might contribute to binding, but not solely. Since the allosteric consequences of gentiobiose binding were not analysed in detail in the present work, it cannot be concluded whether the O6-containing scaffold of gentiobiose would likewise support an allosteric response upon binding.

A comparison of the binding pockets of LacI and TreR interacting with the chemical group at C6 further highlights this position as important for ligand specificity. Both proteins employ different recognition principles for the chemical group at C6. In LacI, the O6 hydroxyl is embedded in a predominantly neutral, water-mediated hydrogen-bonding network that supports allosteric coupling between the N- and C-subdomains.[30] In TreR, the corresponding ligand region is accommodated within a phosphate-recognition environment that combines hydrogen-bonding residues with a positively charged anchor for electrostatic stabilisation of the phosphoryl group.[115]

Accordingly, the reduced compatibility of T6P with LacI may plausibly reflect the absence of the interaction pattern required to accommodate a negatively charged phosphate group. Conversely, in TreR and LLhT, the positive charge in the pocket and the phosphate-coordinating environment are likely to disfavour binding of galactosidic ligands such as IPTG or allolactose. The comparison of the ligands highlights that specific ligand recognition in LacI results from the coordinated interplay of multiple precise interactions between the protein and its ligand. The present findings show that three point mutations are insufficient to achieve complete discrimination between two ligands with closely related chemical architectures. In contrast, the use of a markedly distinct ligand-binding domain, especially one that has evolved and thus been further refined for selective recognition, appears to be the more robust strategy for preserving orthogonality.

3.4 Kinetic origins of impaired DNA binding in heterodimeric motor feet

The design of LacI-based motor feet requires the coordinated combination of different LacI modifications. For a bipedal motor architecture, the feet must not only retain DNA-binding functionality but also be connected to the central clapping machine with defined geometry. In this context, the integration of a structural element into the N-terminal subdomain of the effector-binding domain represents a useful building block. Since the connection to the clapping machine should be introduced into only one subunit of each foot, a heterodimeric LacI architecture is essential. This enables subunit-specific modification and restricts the attachment point for the clapping machine to one defined position within the foot. A heterodimer is generated by the attachment of SpyTag/Catcher to the C-terminus of a monomeric LacI-variant (LacI-A2-SpyTag, LacI*-SpyCatcher).

A major obstacle to successful motor generation was the reduced DNA-binding affinity of the heterodimeric foot variants (Het-Link1-SpyTagxCatcher, Het-Link2-SpyTagxCatcher and Het-

Link1-DogTagxCatcher). When the individual design features required to generate these heterodimeric foot variants are considered separately, it becomes evident that both heterodimer formation and the introduction of structural extensions already reduce DNA-binding affinity on their own. The reduced affinity observed for the foot variants most likely reflects the combination of unfavourable features of both main building blocks. Modifying LacI towards a heterodimer reduces the DNA-binding affinity by 17-fold, while the insertion of the structural extension decreases the affinity by 77-fold. Focusing on the kinetic parameters provides a deeper insight. While heterodimer formation mainly influences DNA association, structural extension influences both association and dissociation. The essentially non-measurable affinity of the heterodimeric feet with SPR, therefore, appears to arise from slower association induced by the heterodimer formation and faster dissociation from DNA due to the structural extension within the PBP domain. This raises the question of why distinct structural interventions in LacI affect different kinetic parameters so differently. A central requirement for the successful generation of a defined heterodimeric assembly, while avoiding the formation of higher-order oligomeric species, was the use of monomeric LacI variants in which the dimerisation domain was fused to the C-terminus. Dimerisation domains are connected to the C-terminus via a flexible $(GS_2)_4$-linker. In principle, high-affinity operator recognition by LacI requires the coordinated engagement of two DNA-binding domains with the palindromic operator.[435] When starting from two monomeric LacI subunits, the DNA-binding domains are not initially positioned in close spatial proximity. Although the dimerisation domains bring the two monomeric subunits into close spatial proximity in solution, the $(GS_2)_4$-linker confers substantial conformational flexibility on the assembly. Due to the increased diffusion space, the possibility that both feet are present in the geometry necessary for specific operator binding is reduced. Only if both DNA-binding domains are present association happens, as a single LacI subunit is insufficient for stable sequence-specific operator recognition. The dimer is the minimal functional DNA-binding unit.[435] The observed decrease in the association rate is therefore fully consistent with the design's structural characteristics.

Although not derived from LacI itself, previous studies on other DNA-binding proteins support the general principle that linker architecture strongly influences DNA recognition. Van Leeuwen *et al.* showed that linker length and composition in the bipartite Oct-1 DNA-binding domain, formed by helix-turn-helix motifs, alter binding geometry and specificity.[466] Likewise, Varming *et al.* demonstrated for the TP901-1 CI phage repressor that its flexible interdomain linker modulates DNA-binding affinity, and their abstract explicitly notes decreased DNA-binding affinity for the variant with the longer flexible linker studied *in vitro*.[467] These studies support the idea that linker-mediated changes in interdomain positioning and flexibility have strong effects on productive DNA association, even when the DNA-recognition motifs themselves remain intact.

In contrast, extension of the α-helices in the N-terminal PBP subdomain led to markedly faster dissociation from the DNA target. An interpretation is that this modification destabilises the protein-DNA complex. In LacI, tight operator binding depends not only on direct contacts between the DNA-binding domain and DNA, but also on an extensive network of inter-subunit interactions that links the DNA-binding headpiece and hinge region to the N-terminal core subdomain. Important protein-protein interactions have been identified between the hinge-helix of one monomer and the N-terminal subdomain of the effector domain of the opposing monomer. Specifically, residues 46-51, found in the hinge helix, are central elements of the crucial interactions, occurring after the disordered to ordered transition.[28,468,469] Accordingly, high-affinity operator binding relies on an integrated interaction network rather than on any single local contact. Consistent with this view, Zhan *et al.* showed that substitutions at V52, although this position does not directly contact the DNA or the core domain, can markedly reduce operator affinity.[469] However, this interaction pattern is not conserved in a sequence-specific manner throughout the LacI/GalR protein family. Subsequent chimera studies showed that functional coupling is not contingent upon a single, precisely defined, and strictly conserved interface.[25] Exchange of the regulatory PBP-like domains between different LacI/GalR family members demonstrated that preservation of the exact residue identities at these contact sites is not required. Instead, what appears to be essential is the formation of a structurally and functionally compatible interaction network that sustains effective communication between the DNA-binding and regulatory regions.[25]

The importance of the interaction network between the hinge helix and the effector binding domain was further highlighted by the insertion of a glycine spacer immediately after the hinge helix, which caused a pronounced loss of operator affinity while leaving IPTG responsiveness intact, indicating that perturbation of the interaction interface is consistent with the selective destabilisation of the DNA-bound state.[470] Summarising this shows that for stable DNA-protein interactions, an interaction network between the hinge-helix and the PBP domain is necessary. Although the crystal structure of the truncated construct, lacking the hinge helix and DNA-binding domain, does not reveal major structural disruption of the N-terminal PBP subdomain upon insertion of the extension, this observation cannot rule out functionally relevant perturbations in a protein that harbours the DNA-binding domain. Interaction patterns and amino acid orientations may differ in the presence of the DNA-binding domain in the crystal structure, as additional headpiece-hinge-core interactions could be formed. Therefore, only a crystal structure including the DNA-binding domain would allow a detailed analysis of how interactions within the protein-DNA complex differ in the presence and absence of the structural extension. However, the reduced DNA-binding affinity observed for variants carrying a structural extension in the N-terminal lobe indicates that interactions required to form a stable protein-DNA complex are affected, thereby shortening the lifetime of the DNA-bound state. The

reduced DNA-binding affinity observed for the final heterodimeric foot variants likely results from a combination of destabilising effects arising from heterodimerisation and structural extension.

3.5 Thermal stability costs of implementing motor-protein design features

One striking and recurrent challenge throughout this project was the loss of thermostability across several design routes. This trend was observed across multiple subprojects and likely also contributed to the poor solubility of the LLhT-6Mut-Link1-SpyCatcher-DogTag variant, one of the two monomers required for heterodimeric LLhT-based foot formation, and required for non-autonomous motor assembly (Chapter 2.5.5). Computational tools such as PROSS[471] and ThermoMPNN[472] support the increase of the stability of analysed variants by combining evolutionary information, atomistic energy functions, or learned structural representations. However, there is no single universal determinant that enhances thermostability across all proteins. Rather, thermostability emerges from a context-dependent interplay of packing, electrostatics, solvation, conformational dynamics, and functional constraints.[473,474] In the present case, the central question is why the thermostability of LacI, and likewise of the related chimera LLhT, was affected strongly by the point mutations introduced within the C-terminal subdomain of the effector binding domain and by the addition of structural extensions, especially in the N-terminal lobe of the effector domain. Mutational studies across many proteins have shown that stability effects are highly position-dependent and are often most severe in buried or tightly packed environments.[475] Although the residues targeted in the present work are not classical core residues, interface residues occupy a similarly packing-sensitive environment and are therefore likely to be destabilised by loss of complementarity, disruption of favourable contacts, or suboptimal quaternary packing.[476] Structural analyses of LacI indicate that the C-terminal subdomains remain comparatively fixed, whereas larger ligand-associated rearrangements are primarily accommodated by the N-subdomains and the DNA-binding/linker regions.[28,30] Consistent with this architecture, previous LacI studies showed that mutations at the dimerisation interface affect assembly, interface stability, DNA binding, and, in selected cases, thermostability.[100,108,477] Notably, stabilisation of the dimerisation interface was previously associated with a substantial increase in thermal stability, suggesting a close link between interface packing and overall fold stability.[100] The decrease in thermostability observed for several variants with mutations present in the monomer-monomer interface in the present study (Chapter 2.2.1 and Chapter 2.5.4) is consistent with the interpretation that disruption or weakening of this monomer-monomer interface contributes to global destabilisation. However, the present data do not support a simple one-to-one relationship between monomerisation and reduced thermostability. The variants LacI-B2, LacI-B3, and LLhT-6Mut each showed a decrease in thermal stability of 10 °C, although the

underlying mutations differed substantially in their effects on the monomer-dimer equilibrium. While LacI-B2 and LacI-B3 primarily perturbed the equilibrium between monomeric and dimeric states, LLhT-6Mut abolished detectable monomer-monomer interactions entirely. The positions of the mutations in those variants are similar, with four modifications incorporated at identical positions. On the other hand, the number of mutations influences thermal stability, as shown by the decreased stability of LacI-B2 and LacI-B3 compared to LacI-B1, LacI-A1, and LacI-A2, which have only three or one mutation. Directly comparing the LacI-A1 and LacI-A2, which vary only in the amino acid exchanged, also highlights that the amino acid itself can have an influence on the stability. Instead, individual positions contribute to interface stability to different extents. Important features include their positions within the interface, amino acid context of surrounding residues, and the extent to which these changes influence packing and intersubunit interactions.

A second major determinant of reduced stability was the introduction of structural extensions. In both LacI and LLhT, insertion into the N-terminal PBP subdomain caused a substantially stronger loss of thermostability than extension at the C-terminus. Although the number of constructs analysed here is limited, this pattern strongly suggests that the insertion site, rather than the mere presence of an added segment, is a major determinant of stability in this system. In principle, these observations are consistent with the analysis of insertion and deletion mutations reported by Savino *et al*. The tolerance for insertions is not uniformly distributed across the protein fold but varies strongly with structural context. Flexible loops and terminal regions are generally more permissive to indels, whereas α-helices and β-sheets show reduced tolerance.[478]

This general picture is further supported by large-scale evolutionary analyses showing that indels are depleted in regions predicted to form secondary structure and enriched in regions without defined secondary structure, with the strongest depletion observed for β-strands.[479] Together, these findings support the view that perturbations within pre-organised structural elements are often substantially more disruptive than modifications in flexible or peripheral regions. Within this framework, the minor effect of the C-terminal extension in LacI-Lev2 is readily explained by the more peripheral and comparatively permissive character of this region.[480] While a terminal extension is tethered to the scaffold by only one linkage and therefore retains greater orientational freedom, an inserted element is constrained by two flanking connections.

In contrast, insertion into a pre-existing α-helical segment, as in LacI-Lev1 and LLhT-Lev7, is expected to impose a markedly greater structural penalty, because the introduced sequence must be accommodated within an already organised local environment. Several features required for the proposed motor protein appear to impose a stability penalty, including interface engineering to generate monomeric subunits and the incorporation of a structural extension as

a coupling element. However, the thermostability of individual variants cannot be explained by a single dominant feature. Instead, it reflects the cumulative contribution of multiple design elements combined within a given LacI or LLhT variant.

3.6 Effects of connector modules and structural extension on oligomeric state

For an effective bipedal protein motor, it is important to control the oligomeric state of the assembled motor and prevent off-target oligomerisation of individual motor protein components, which could lead to unspecific motor protein assemblies. It is therefore important that the oligomeric state of the individual components is controlled. Different challenges regarding the oligomeric state occurred during different subprojects. One major issue was observed for the clapping-machine construct, PotF, with the connector modules DogCatcher and SpyTag attached to the N- and C-terminus. Native PotF is a periplasmic binding protein that is generally considered monomeric, although *in vitro* analyses indicated a small fraction of dimeric species already for the unmodified scaffold. Upon introduction of the connector modules, the dimeric fraction increased markedly.

Comparison of different variants suggests that this behaviour is primarily associated with DogCatcher rather than with PotF or SpyTag. Constructs carrying SpyTag alone did not show a comparable increase in dimer formation, in comparison to comparable variants without (e.g. LacI-A2, LacI-A2-SpyTag). In contrast, the only other available DogCatcher-containing construct, LacI*-DogCatcher, displayed a pronounced dimeric peak during purification, while the construct without (LacI-A2) is monomeric. Taken together, these observations indicate that DogCatcher-containing fusion proteins in the present design context exhibit a greater tendency toward self-association. No direct precedent of self-association appears to exist in the current literature. No comparable behaviour was observed for the connector module SpyCatcher-containing variants, consistent with previous biophysical characterisation of SpyCatcher003 as predominantly monomeric in solution.[481]

Having higher oligomeric species impairs the bipedal motor architecture, but the behaviour of the covalently linked variants of DogTagxCatcher indicates that the self-association of DogCatcher is diminished upon complex formation. Fully coupled DogTag×DogCatcher (e.g. Het-Link1-DogTagxCatcher) assemblies no longer displayed a pronounced excess of higher oligomeric species relative to the corresponding SpyTagxCatcher control (e.g. Het-Link1-SpyTagxCatcher). As the tendency of unfused DogCatcher to dimerise is not retained after covalent assembly with DogTag, this suggests that the DogTag/DogCatcher system remains suitable for constructing artificial motor proteins. No compromises are required in the motor architecture towards a defined oligomeric assembly state.

A second type of undesired oligomeric behaviour was observed for the extension variants LacI-Lev1 and LLhT-Lev7. In both cases, SEC-MALS yielded apparent molecular weights between

the monomer and dimer values, whereas SEC did not resolve separate monomer and dimer peaks. This behaviour is consistent with a shifted monomer-dimer equilibrium rather than with the presence of two fully separated oligomeric species. This effect is not primarily problematic for the final motor concept, because the natural LacI dimer is deliberately destabilised in later design steps to generate asymmetric feet. Nevertheless, the behaviour is mechanistically informative because it shows that structural extension in the N-terminal PBP subdomain alters the energetic balance of assembly. While the C-terminal subdomain is the main contributor towards a stable monomer-monomer interface, the N-subdomain interface, including residues such as K84 and positions 94-98, appears to act less as a rigid packing nucleus and more as a structurally responsive interface that participates in allosteric communication.[67,102,106,477] Mutations in the N-subdomain modulate interface strength and thermostability, but their effects are closely intertwined with altered conformational coupling rather than reflecting disruption of the primary assembly itself.[100] Within this framework, the effect of the extension on the oligomeric state is most plausibly interpreted as an indirect destabilisation of the dimer. A direct steric disruption of key interactions required for formation of a stable protein–DNA complex appears unlikely, as the structural extension is located far from the canonical C-terminal dimerisation hotspot. Rather, the extension in the N-terminal lobe may alter the conformational and energetic properties of this domain, thereby affecting how the N-terminal subdomain contributes to stabilisation of the assembled dimeric state.In other words, the extension likely does not redefine the dimer interface itself but instead shifts the balance between stabilising and destabilising contributions within the full-length protein, making dissociation of the dimeric state more favourable.

A minor limitation to the efficient generation of the bipedal motor is that disruption of the LacI monomer-monomer interface does not appear to be fully complete. The point mutation Y282S was present in all heterodimeric variants. Even though it has been shown to form mostly monomeric species, a small amount of dimeric species also remains for LacI-Y282S (LacI-A2) (Chapter 2.2.1). The small proportion of homodimerisation is retained for more advanced variants with the Y282S point mutations, for example, LacI-A2-SpyTag, LacI*-SpyCatcher, but also for the heterodimeric assemblies LacI-A2-SpyTag×Catcher and LacI-A2-DogTag×Catcher. Across the different LacI-A2-containing variants, the relative extent of this population remained similar and did not increase substantially for any particular connector module, with the expected exception of DogCatcher. As connector modules are also present in the advanced variant, higher oligomeric states of theses variants may result from nonspecific covalent linkage of the connector module. Since no corresponding covalently linked side products were detected by SDS-PAGE, it was shown that connector modules interact with a high selectivity and do not lead to higher oligomeric states if present in the single components. Also some limitations are observed in regards of idealised oligomeric behaviour, the limitations

are either not present in the assembled feet or only have a small influence on the efficiency of the assembly of the motor protein.

3.7 Plasticity of the LacI interface and its implications for monomerisation of LLhT

For LacI, the monomer-monomer interface has been characterised in detail and reveals a remarkably plastic mode of dimer stabilisation. While single point mutations at position 278, 281 or 282 lead to the generation of monomeric LacI-variants,[100,106,443] additional mutations further compensate for the dimerisation.[106]

The monomer-monomer interface is structurally conserved across the LacI/GalR protein family and exhibits similar interaction patterns. But the structural conservation does not translate into sequence conservation.[106] Therefore, the question arises whether mutations introduced into the monomer-monomer interface of LacI to disrupt dimerisation can be transferred to other members of the LacI/GalR protein family. In the present study, several residues of LLhT were therefore selected by alanine substitution. Specifically, positions 268, 271, and 272 were analysed, as mutations at the corresponding positions in LacI are known to affect the monomer-dimer equilibrium. In addition, position 244 was included in the mutational analysis, despite the lack of an established precedent in LacI. This residue was identified as a key interaction partner within the monomer-monomer interface network. It forms contacts with positions 271 and 272 of the opposite monomer. However, none of the individual alanine or serine substitutions was sufficient to shift LLhT into a predominantly monomeric state. A further five mutations at positions within the monomer-monomer interface, forming conserved interactions throughout the family, needed to be mutated to alanine to successfully generate a monomeric LLhT variant.

These findings support the notion that interface stability in LacI/GalR proteins is encoded by cooperative, context-dependent interaction networks that are specific to each family member. Accordingly, even if LLhT residues are structurally analogous to LacI positions and participate in comparable intersubunit contacts, the energetic contribution of each position to dimer stability is not necessarily the same throughout the LacI/GalR protein family. The effects of interface mutations identified in LacI cannot, therefore, be assumed to transfer directly to LLhT. Even within LacI itself, dimer stability is not limited to a single invariant interaction network; instead, different combinations of interface interactions can support dimer formation.[106] Highlighting that for LacI itself, dimer stability is not restricted to a single invariant interaction network. Instead, different combinations of interface interactions can support the formation of a stable dimer. This is consistent not only with previous reports describing multiple alternative ways to form a LacI dimer,[106] but also with the LacI-B variants analysed in this thesis, which demonstrate that alternative interface architectures can support dimer formation. These results highlight that mutations at structurally analogous positions within the monomer-monomer

interface do not necessarily affect the monomer-dimer equilibrium in the same way across different members of the LacI/GalR protein family. Accordingly, monomerising mutations identified in LacI cannot be directly transferred to LLhT to achieve an equivalent outcome. This suggests that a universal monomerisation strategy for the LacI/GalR family is unlikely, since the interface architecture and the energetic contribution of individual residues are protein-specific.

3.8 Orthogonal Tag/Catcher systems as modular connectors for motor assembly: opportunities and architectural limitations

Tag/Catcher domains constitute versatile protein-based coupling systems that have been successfully used to covalently link distinct protein fragments.[183] Their suitability for the proposed motor architecture was systematically assessed. Two orthogonal Tag/Catcher systems were used as the heterodimerisation domain and the connector module between the feet and the clapping machine. The connector modules were interchanged between the feet, providing a promising strategy for stepwise assembly of the motor architecture.

At the level of the isolated feet, this concept proved highly effective: SpyTag/SpyCatcher and DogTag/DogCatcher remained sufficiently orthogonal to permit selective heterodimer formation without detectable accumulation of higher-order oligomeric species, even if the heterodimerisation domain (e.g. SpyTagxCatcher) and the linker attachment module (e.g. DogTag) were present. This supports the results obtained by Keeble *et al.*, which showed that their newly generated DogTag/Catcher modules do not interact with SpyTag/Catcher.[481]

Likewise, attachment of a single foot to the central clapping machine remained highly efficient, indicating that both connector pairs retained their functionality in the fusion context. The situation changed substantially when both feet and the clapping machine were combined into a single reaction. Although formation of the half-assembled intermediate still proceeded efficiently, addition of the second foot was markedly reduced, such that only about 30% of the complexes reached the fully assembled motor state. These findings show that orthogonal Tag/Catcher domains are, in principle, highly effective connector modules for modular motor assembly. Still, they reveal an important limitation of the present design: the successful formation of the complete complex is determined not only by biochemical orthogonality but also by geometric accessibility within the multi-component architecture. Results indicate that the interaction between the first foot and the clapping machine hinders access to the second connector element. Due to the flexible linkers present in the current clapping machine construct (CM-PotF), a possible explanation for the nonetheless limited number of fully assembled constructs is that the steric hindrance might only be present within a specific assembly order. Meaning if foot A interacts first, foot B cannot interact, while the other way, the connectors remain accessible. In this view, incomplete assembly would not primarily reflect insufficient

intrinsic reactivity of the Tag/Catcher domains themselves, but rather context-dependent blocking imposed by the architecture of the partially assembled intermediate. It is further conceivable that this effect is asymmetric. The binding of foot A may hinder the subsequent attachment of foot B, whereas the reverse order of assembly may remain more permissive and still allow completion. If so, the fully assembled motors observed here may arise predominantly from one favourable assembly pathway, while the alternative route becomes kinetically trapped in a half-assembled state. At present, however, this interpretation remains speculative because no kinetic dissection of the individual assembly steps was performed.

A more detailed mechanistic understanding of this effect would require consideration of the effective association kinetics of the individual connector pairs in their specific fusion context. Even though association kinetics for SpyTag/SpyCatcher and DogTag/DogCatcher were determined in simplified systems, these kinetics may not directly reflect the behaviour of the present fusion constructs.[197,481] For DogTag/DogCatcher in particular, the measured second-order rate constants vary substantially with structural context, indicating that local presentation and steric accessibility can strongly modulate effective reactivity.[197] Thus, in the present system, context-dependent architectural effects are likely to be at least as important as the intrinsic chemistry of the connector pairs themselves.

The present study also extends previous applications of DogTag and SpyTag-based connector systems. DogTag has been used primarily as a modular ligation handle, including in structurally constrained contexts such as loops.[197,482–484] Here, it was incorporated productively both at a terminus and within an engineered structural region. SpyTag/SpyCatcher, by contrast, has been used predominantly in terminal fusion formats,[456,485] but the present data show that function is retained when the SpyCatcher module is embedded within a loop. Importantly, the longer-term aim of the present design is not merely covalent coupling, but the establishment of a connection that may ultimately allow transmission of conformational changes. This would move orthogonal Tag/Catcher systems from modular ligation systems towards their prospective use in mechanically coupled protein assemblies of an artificial motor protein. The Tag/Catcher-based assembly strategy provides an experimentally validated route towards the modular construction of the proposed artificial motor protein.

The results show that the basic motor architecture can be generated from individually produced components. This has important implications for future motor design. The modular architecture allows individual components to be optimised separately, facilitating iterative improvement of stability, solubility, DNA-binding behaviour, and mechanical coupling. At the same time, the observed assembly limitation highlights the need to design the spatial arrangement of the connector modules more carefully, for example, by adjusting linker length, rigidity, orientation, or attachment sites. But the present work highlights both the feasibility and the key design constraints of a modular Tag/Catcher-based strategy for the construction of the artificial motor.

4 Conclusion and Outlook

The overall aim of this thesis was to contribute to the development of an artificial protein motor within the ArtMotor project. Establishing LacI as a functional and engineerable component of a motor architecture was the main purpose of this thesis. The present study provides several key advances towards this goal and creates the basis for future molecular motor construction based on LacI.

A first major achievement was the establishment of a robust, modular strategy for generating heterodimeric LacI. While redesign of the native LacI dimer interface alone did not yield a sufficiently stable and selective heterodimer *in vitro*, the use of Tag/Catcher provided a reliable solution to this problem. This strategy enabled the formation of a defined heterodimeric LacI species and, importantly, established selective binding to asymmetric operator DNA. In this way, the work created a route toward asymmetric motor feet, which is a central prerequisite for directional behaviour in a bipedal motor system.

Beyond establishing a heterodimerisation strategy, it was shown that the LacI architecture enables significant structural expansion through RFdiffusion-guided design. The successful implementation of a rigid extension in the N-terminal PBP subdomain showed that LacI tolerates the introduction of motion-relevant structural elements even within an allosterically active domain. Although this intervention compromised reduced DNA binding and thermostability, it nevertheless represents a substantial conceptual advance that ligand binding and allosteric control are preserved. The successful design of a rigid connector element establishes an important basis for coupling internal conformational changes from the clapping machine to the feet. This rigid connection, which is intended to transmit conformational changes, should enhance the processivity of both non-autonomous and autonomous motors by shifting the mode of interaction from purely Brownian motion towards more biased movement.

The functional coordination of a bipedal motor in a non-autonomous way requires a high degree of ligand control. The present work further identified selective effector recognition as a realistic basis for ligand orthogonality in future motor designs. In particular, the clear orthogonality observed between LacI and TreR recognition of IPTG and T6P supports the feasibility of independently addressable motor components. Since orthogonal input signals are essential for a non-autonomous stepping mechanism, this finding removes a central conceptual obstacle for future implementation of a ligand-controlled motor cycle. The use of Tag/Catcher pairs as connector elements between protein-based motor components proved to be an effective modular assembly strategy. SpyTag/Catcher and DogTag/Catcher could be implemented within the same building block while maintaining orthogonality under the conditions tested. Although formation of the irreversible isopeptide bond was generally successful, coupling efficiency was strongly influenced by geometric accessibility and by the

architectural context of the respective fusion construct.

Beyond these achievements, the thesis also yielded broader mechanistic insight into LacI and related LacI/GalR-family proteins. The results further support the view that the LacI dimer interface is highly context-dependent and cannot be reprogrammed by simply transferring individual “monomerising” residues to homologous systems. Likewise, the work showed that distinct design interventions impair DNA binding by different kinetic routes: heterodimerisation primarily reduced productive association with DNA, whereas structural extension mainly accelerated dissociation of the DNA-bound state. Furthermore, thermostability was impaired across several design routes. The introduction of motor-relevant features, therefore, appears to compromise protein stability, indicating that their implementation must be viewed as a biophysical trade-off. Accordingly, required functional elements need to be balanced against their destabilising effects on thermostability.

Although a functional molecular walker was not achieved, this work established several key requirements for the development of a LacI-based artificial molecular motor. It demonstrates the feasibility of generating heterodimeric LacI variants for asymmetric DNA binding, shows that a rigid structural extension can be incorporated into the allosterically sensitive N-terminal lobe of LacI, and identifies orthogonal ligand-protein interactions that may be exploited in a non-autonomous motor concept. Importantly, the work also defines precise limitations of the current design, providing a basis for the targeted optimisation of biophysical properties in future iterations toward a functional artificial motor protein.

Building on these findings, the most immediate priority for future work is to improve the DNA binding of the engineered motor feet, both by enhancing the reduced association of the heterodimeric constructs and by limiting the increased dissociation associated with incorporation of the structural extension. Furthermore, a close look needs to be taken at the thermostability and the influence on the thermal stability of the mutation, to enable successful formation of an LLhT-based foot. Although motor assembly has already been achieved, further optimisation of connector rigidity will likely be critical for efficient, potentially processive motor function. In addition, refinement of the clapping-machine module, or its replacement with an enzymatic element, represents an important next step towards a more advanced and ultimately autonomous motor design.

5 Materials and Methods

5.1 Materials

5.1.1 Bacterial Strain

E. coli BL21(DE3)	*E. coli* B dcm ompT hsdS(rB-mB-) gal (Merck KGaA, Darmstadt)
E. coli DH5α	*E. coli* fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17 (NEW ENGLAND BIOLABS, Frankfurt a. M.)
E. coli Top10	*E.coli* F^-, mcrA, Δ(mrr-hsdRMS-mcrBC), φ80lacZΔM15, ΔlacX74, recA1, araD139, Δ(ara-leu)7697, galU, galK, rpsL(StrR), endA1, nupG (Thermo Fischer Scientific GmbH, Schwerte)

5.1.2 Plasmids

pET21b(+)[486]	T7 promoter, ampicillin resistance, C-terminal His_6-tag
pDUET^{M} [487]	Dual-expression vector, two cloning sites, T7 promoter, ampicillin resistance. Modified in this work with splitGFP incorporated into the reading frames, strand 1-9 within reading frame 1 and strand 10-11 in reading frame 2.
pBAD/HisA[488]	Arabinose-inducible pBAD promotor; ampicillin resistance; C-terminal His_6-tag
pCDF-pBAD[489]	CDF-derived compatible origin, arabinose-inducible pBAD promoter, streptomycin resistance

5.1.3 Sequencing Primers

Table 17. Overview of the used sequencing primers.

Primer name	Sequence 5'-3'
DuetDOWN1	GAT TAT GCG GCC GTG TAC AA
DuetUP2	TTG TAC ACG GCC GCA TAA TC
pBAD-FP	ATG CCA TAG CAT TTT TAT CC
pBAD-RP	GAT TTA ATC TGT ATC AGG
petup	ACG ATG CGT CCG GCG TAG A
T7	TAA TAC GAC TCA CTA TAG GG
T7term	GCT AGT TAT TGC TCA GCG G

5.1.4 Cloning Primers

Primer for SDM

Table 18. Overview of the primers used for SDM.

Primer name	Sequence 5'-3'
fw-LacI-Y282A	GGTCAGCGGCGGAATAGCACAACTACTATCTTCGGTATCATC
rv-LacI-Y282A	GATGATACCGAAGATAGTAGTTGTGCTATTCCGCCGCTGACC
fw-LacI-R197G	GTGCACGCCTGGGCCTGGCAGGC
rv-LacI-R197G	GCCTGCCAGGCCCAGGCGTGCAC
fw-LacI-Y282S	GTCAGCGGCGGAATAAAACAACTACTATCTTCGG
rv-LacI-Y282S	CCGAAGATAGTAGTTGTTTTATTCCGCCGCTGAC
fw-LacI-D278N	GGCGGAATAAAACAACTACTATTTTCGGTATCATCATAACCAACC
rv-LacI-D278N	GGTTGGTTATGATGATACCGAAAATAGTAGTTGTTTTATTCCGCC
fw-LacI-L251M	GCGCATTGCGCCCATTGCCACTGATCATTTGCAACC
rv-LacI-L251M	GGTTGCAAATGATCAGATGGCAATGGGCGCAATGCGC
fw-LacI-Y17T	GCGGCTAACGGTCTGAGTGCTAACGCCGGCATAT
rv-LacI-Y17T	ATATGCCGGCGTTAGCACTCAGACCGTTAGCCGC
fw-LacI-Q18A	AACGCGGCTAACGGTCGCAGTGCTAACGCCGGCATATTCG
rv-LacI-Q18A	CGAATATGCCGGCGTTAGCACTGCGACCGTTAGCCGCGTT
fw-LacI-R22N	GCTTGCCTGATTAACAACGTTGCTAACGGTCGCATAGCTA
rv-LacI-R22N	TAGCTATGCGACCGTTAGCAACGTTGTTAATCAGGCAAGC
fw-LacI-Q291H	CCTTTAACGACTATAAAGCACGATTTCCGTCTTTTGGGTCA
rv-LacI-Q291H	TGACCCAAAAGACGGAAATCGTGCTTTATAGTCGTTAAAGG
fw-LacI-T276L-S279G	GGACATCTCCGTCATAGGATACGACGATCTGGAAGACGGTGCCTGTTATATACC
rv-LacI-T276L-S279G	GGTATATAACAGGCACCGTCTTCCAGATCGTCGTATCCTATGACGGAGATGTCC
fw-LLhT-271S	CAATTTCCGGATGCAGGCTTTTCATCAGAGGGGTATTGCCAACACTG
rv-LLhT-271S	CAGTGTTGGCAATACCCCTGCTGATGAAAAGCCTGCATCCGGAAATTG
fw-LLhT-244A	TACTTGCACCCGCTGCCAGGGTATCTGTTGCGCA
rv-LLhT-244A	TGCGCAACAGATACCCTGGCAGCGGGTGCAAGTA
fw-LLhT-268A	CGGATGCAGAAATTTCATCGCAGGGGTATTGCCAACACTG
rv-LLhT-268A	CAGTGTTGGCAATACCCCTGCGATGAAATTTCTGCATCCG
fw-LLhT-271A	AATTTCCGGATGCAGAGCTTTCATCAGAGGGGTATTGCCAACA
rv-LLhT-271A	TGTTGGCAATACCCCTCTGATGAAAGCTCTGCATCCGGAAATT
fw-LLhT-272A	CAATTTCCGGATGCGCAAATTTCATCAGAGGGGTATTGCCAAC
rv-LLhT-272A	GTTGGCAATACCCCTCTGATGAAATTTGCGCATCCGGAAATTG
Fw-LLhT-271-272 (after 268A, 244A)	GTAACAATTTCCGGATGCGCGGCTTTCATCGCAGGGGTATTGCCAACACTGGC
Rv-LLhT-271-272 (after L268A, L244A)	GCCAGTGTTGGCAATACCCCTGCGATGAAAGCCGCGCATCCGGAAATTGTTAC
fw-LLhT-217-218	TTTGGCCACATTTTCATAACCCTGCGCGGCTGCAAGACCCGGCAGTGCTGCCAC

Primer name	Sequence 5'-3'
rv-LLhT-217-218	GTGGCAGCACTGCCGGGTCTTGCAGCCGCGCAGGGTTATGAAAATGT GGCCAAA
Fw-LLhT-221 (after 217A, 218A)	CACTTTGGCCACATTTTCCGCACCCTGCGCGGCTGCAAG
Rv-LLhT-221 (after 217A, 218A)	CTTGCAGCCGCGCAGGTGCGGAAAATGTGGCCAAAGTG

Primer for RE-based cloning

Amplification of pDUET^{M}

Table 19. Overview of primer used for the amplification of linearised pDUET with RE-site overhang within the first and second reading frame.

Primer name	Sequence 5'-3'
fw-pDUET-NcoI-1st	CGAGGTATCCATGGATCATATGAAGCCGCTGACCCTGTATGATGTG
rv-pDEUT-SacI-1st	CGTATTAGAGCTCGTCAGTCCTGGTGGTGGTGGTGCTC
rv-pDUET-AvrII-2nd	CCATATCCTAGGTCAGTGGTGGTGGTGGTGGTGCTCG
fw-pDUET-NdeI-2nd	CGTGCAGTGATACATATGAAGCCCGGTGACCCTGTATGATGTGGC

Amplification of inserts required for splitGFP

Table 20. Overview of primers used for the amplification of the inserts for the incorporation into both reading frames of pDUET to generate variants required for splitGFP-attached proteins. Primers named LacI can be used for the LacI-A-splitGFP, WT-splitGFP and LacI-splitGFP.

Primer name	Sequence 5'-3'
fw-TruncII-EcoRI-1st	CCAGTAGAATTCATGCTGGAAGATGCAAAAGAAATTCCGTA
rv-TruncII-SacI-1st	CGTTATGAGCTCTTTACGACCGTATGCGGTATATTTATC
fw-LD-NdeI-2nd	CTAATTCATATGCTGGAAAAAGGCAAAATGGCAATTG
rv-LD-EcoRV-2nd	GTACATGATATCCTTTTCTTTCAGGAATTTAACGAAGTATTC
fw-LacI-EcoRI-1st	GTTAGCGATGGACAGGAATTCCACATGAAACCC
rv-LacI-SacI-1st	GCTACATGAGCTCTTCCAGTGCGAGTGTTG
fw-LacI-NdeI-2nd	CAGATGGCTCATATGAAACCCGTCAC
rv-LacI-EcoRV-2nd	CTGCAATGATATCTTCTAACGCTAACGTAG

Amplification of pBAD-HisA

Table 21. Overview of primers used for the amplification of linearised pBAD-HisA with *SalI* and *XhoI* RE-sites

Primer name	Sequence 5'-3'
fw-BAD-SalI	GTCGATAATGCTCAATTGATCTCGAGGGCCAGGGTGGTTTTACG
rv-BAD-XhoI	CTGCAAGTCGACGGATCAATTAGAAGTTTTGTTTCAAGGTCCATGGAGC CAC

Amplification SpyCatcher

Table 22. Overview of primers used for the amplification of SpyCatcher, adding RE-sites *SalI* and *XhoI*.

Primer name	Sequence 5'-3'
fw-SpyCatcher-SalI	CTCTTAGTCGACGGCAGCGGTAGTGGCAGTGGCAG
rv-SpyCatcher-XhoI	GAAGCTCCATCGAGAATATGTGCGCCTTTGGTTGCTTTGCCATTAAC

Amplification of pCDF-pBAD

Table 23. Overview of primers used for the amplification of linearised pCDF-BAD with *SalI* and *EcoRI* RE-sites

Primer name	Sequence 5'-3'
fw-CDF-SalI	CATTAGGTCGACCACTCGAGAGCTTGGCTGTTTTGGCGGATG
rv-CDF-EcoRI	GTCTCCTGAATTCCTCGAGGGCCAGGGTGGTTTTACGTTTAACC

Amplification of SpyTag

Table 24. Overview of primers used for the amplification of SpyTag, adding RE-sites *SalI* and *EcoRI*.

Primer name	Sequence 5'-3'
fw-SpyTag-EcoRI	GTAGCTGAATTCGCCCATATTGTGATGGTGGATGCCTAC
rv-SpyTag-SalI	GTAATTGTCGACCAGCTTCCTTTCGGGCTTTGTTAGCAG

Amplification of pDUET^{M} structural extension without the DNA-binding domain

Table 25. Overview of primers used for the removal of the DNA-binding domain for LacI-Lev1 and LacI-Lev2 to generate PBP1 and PBP2 variants.

Primer name	Sequence 5'-3'
fw-removeDNABD-Lev1	GTTAACCATATGTTATTGATTGGGGTTGCAACATC
fw-removeDNABD-Lev2	GTTAACCATATGCTGTTAATTGGCGTCGCAAC
rv-removeDNABD-Lev1+2	GATCCAGATATCCATATGTATATCTCCTTCTTATACTTAAC

5.1.5 Molecular Weight Marker

DNA Molecular Weight Marker

1 kb DNA Ladder

Polynucleotide size [bp]:
10000, 8000, 6000, 5000, 4000, 3000, 2000, 1500, 1000

New England Biolabs GmbH, Frankfurt a. M.

Protein Molecular Weight Marker

Pierce™ Unstained Protein Marker

Protein size [kDa]:	Thermo Fischer Scientific GmbH, Schwerte
116.0, 66.2, 45.0, 35.0, 25.0, 18.4, 14,4	

Blue Prestained Protein Standard Broad Range

Protein size [kDa]:	
250.0, 180.0, 130.0, 95.0, 72.0, 55.0, 43.0, 34.0, 26.0, 17.0, 11.0	New England Biolabs GmbH, Frankfurt a. M

5.1.6 Enzymes

DNA-Polymerases

KAPA HiFi DNA Polymerase	ROCHE Deutschland Holding GmbH, Mannheim
Phusion™ High-Fidelity DNA Polymerase	New England Biolabs GmbH, Frankfurt a. M.
Taq DNA Polymerase	New England Biolabs GmbH, Frankfurt a. M.

Type II Restriction Endonucleases

EcoRI	New England Biolabs GmbH, Frankfurt a. M.
SacI	New England Biolabs GmbH, Frankfurt a. M.
NdeI	New England Biolabs GmbH, Frankfurt a. M.
XhoI	New England Biolabs GmbH, Frankfurt a. M.
DpnI	New England Biolabs GmbH, Frankfurt a. M.

Ligase

T4 DNA Ligase	New England Biolabs GmbH, Frankfurt a. M.

5.1.7 Chemicals

BLDpharmatech Ltd., Shanghai, China	biotin, β-cyclodextrin
Carl ROTH GmbH + Co. KG, Karlsruhe	acetate, agar-agar, agarose, ampicillin, L-arabinose, bis-Tris, bromophenol blue, dithiothreitol (DTT), ethylenediaminetetraacetic acid (EDTA), isopropyl β-D-1-thiogalactopyranoside (IPTG), phosphoric acid, sodium dodecyl sulfate (SDS), Tris

Fischer Scientific GmbH, Schwerte	glycerol, sodium chloride (NaCl)
Grüssing GmbH, Filsum	disodium hydrogen phosphate (Na_2HPO_4), sodium dihydrogen phosphate (NaH_2PO_4)
Merck KGaA, Darmstadt	ammonium persulfate (APS)
AppliChem GmbH, Darmstadt	29:1 acrylamid-bisacrylamid
Sigma-Aldrich Biochemie GmbH, Hamburg	ethanol, β-D-galactose, imidazole, streptavidin, T6P
SERVA Electrophoresis GmbH, Heidelberg	Coomassie Brilliant Blue G-250
Thermo Fisher Scientific GmbH, Schwerte	2-(N-morpholino)-ethanesulfonic acid-SDS-running buffer, (MES-SDS-running buffer)
VWR Chemicals, Darmstadt	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)

5.1.8 Kits

DNA Clean & Concentrator-100 (DCC-100)	ZYMO Research Europe GmbH, Freiburg
NucleoSpin® Plasmid Kit	MACHERY-NAGEL GmbH & Co. KG, Düren
NucleoSpin® Gel and PCR Clean-up	MACHERY-NAGEL GmbH & Co. KG, Düren

5.1.9 Consumables

AlphaLaboratories, Hampshire, UK	Transfer pipettes Pastette®
B. Braun Melsungen AG, Melsungen	Disposable syringe (1 mL, 5 mL, 10 mL, 50 mL)
Hampton Research, Aliso Viejo, USA	Intelli-Plate 96-3 LVR, VDX Plate with sealant
MACHEREY-NAGEL GmbH & Co. KG, Düren	NucleoSpin® Gel and PCR Clean-up Kit, NucleoSpin® Plasmid Kit
Merck KGaA, Darmstadt	Amicon® Ultra centrifuge filter (MWCO: 10 kDa) Millipore Express® PLUS 0.22 µm PES Membrane

NIPPON Genetics EUROPE GmbH, Düren	Centrifuge Tube 15 mL, 50 mL
Repligen Corporation, California, USA	Spectra/Por® Dialysis Membrane Standard RC Tubing (MWCO 6-8 kDa)
Sarstedt AG & Co. KG, Nümbrecht	Petri dish (60 x 15 mm) SafeSeal tube (2 mL, 1.5 mL, 0.5 mL, 0.2 mL) Serological pipette (1 mL, 5 mL, 10 mL, 25 mL)
Sigma-Aldrich Biochemie GmbH, Hamburg	Parafilm “M” Laboratory Film
Starlab GmbH, Hamburg	Bevelled Tip (10 µL, 200 µL, 1000 µL)
Thermo Fischer Scientific GmbH, Schwerte	Sterile PVDF Syringe Filter 0.45 µm
Zymo Research Europe GmbH, Freiburg	DNA Clean & Concentrator-100 (DCC-100) Kit

5.1.10 Equipment

Beckmann Coulter GmbH, Krefeld	Avanti JXN-26; rotor IDs: JLA-8.1000 and JA-25.50
Binder GmbH, Tuttlingen	BD 53 heating cabinet
Bio-Rad Laboratories GmbH, Feldkirchen	MiniPeprean® Tetra Cell PowerPac™ Basic
Branson Ultrasonics GmbH & Co. OHG, Dietzenbach	Sonifier 102C (CE)
Cleaver Scientific, Warwickshire, UK	POWER Pro
Cytiva Europe GmbH, Freiburg im Breisgau	Äkta™ pure; Column: HiLoad 26/60 Superdex™ 75 prep grade, HiLoad 26/60 Superdex™ 200 prep grade, Superdex™ 200 10/300 GL, Superdex™ 75 10/300 GL Gel Filtration Calibration Kits (LMW)
Eppendorf AG, Hamburg	BioPhotometer® Centrifuge 5427R: Rotor FA 45-48-11 Centrifuge 5920R; Rotor: S-4x750

	Research Plus® 1000/100/10/2.5 & 8-Channel 300/10 µCuvette G1.0 ThermoMixer® F2.0
Formulatrix Inc., Bedford, USA	Formulatrix® Rock Imager®
GE HealthCare GmbH, Düsseldorf	HisTrap™ FF Columns 5 mL ÄKTA™ Pure HisTrap™ FF Column 5 mL
Hanna Instruments Deutschland GmbH, Vöhringen	pH 211 Microprocessor pH Meter
Heidolph Scientific Products GmbH, Schwabach	MR3000 magnetic stirrer plate
Hellma GmbH & Co. KG, Müllheim	Cuvette QX 110 – macro cuvette 1 mm, Cuvette QS 105.250- ultra micro cuvette 10 mm
Hirschmann Laborgeräte GmbH & Co. KG, Eberstadt	Pipetus®
IBA Lifescience GmbH, Göttingen	Strep-Tactin®XT 4Flow® FPLC column
JASCO Deutschland GmbH, Pfungstadt	J-1500 CD Spectrometer FP-6500 Spectrofluorometer
KERN & SOHN GmbH, Balingen-Frommern	CKE 16K0.1 laboratory Balance
MALVERN, Kassel	Microcal peaq-DSC
New Brunswick Scientific GmbH, Nürtingen	C24 Incubator Shaker Innova® 44 Incubator Shaker
NIPPON Genetics EUROPE GmbH, Düren	FastGene FAS-DIGI PRO
Systec GmbH, Linden	Autoclave Vx-150
Tecan Trading AG, Männedorf, Switzerland	Spark® Multimode Microplate Reader
Thermo Fischer Scientific GmbH, Schwerte	TKA GenePure ultra-pure water system
Vacuubrand GmbH + Co. KG, Wertheim	Vacuum Pumping Unit PC 3004 Vario
Vilber GmbH, Eberhardzell	E-Box VX 20M Gel Imaging System

Willis Towers Watson GmbH, Frankfurt a.M.	inoLab® pH7110 pH-Meter
WYATT Technology, Santa Barbara, USA	MALS system (Agilent 1260 Infinity II HPLC system, miniDAWN MALS detector, Optilab differential refractive index detector)

5.1.11 Culture Medium

Lysogeny broth (LB)-medium 5 g·L^{-1} yeast extract, 10 g·L^{-1} tryptone, 10 g·L^{-1} NaCl	Carl ROTH, Karlsruhe

5.1.12 Buffer and solutions

Unless stated otherwise, all buffers were sterilised by filtration and stored at room temperature. Bidistilled water (ddH2O) was used throughout the preparation.

Antibiotics

After sterile filtration, the prepared antibiotics were stored at -20 °C.

Ampicillin stock solution (1000x)	100 mg·mL^{-1} dissolved in water
Streptomycin stock solution (1000x)	50 mg·mL^{-1} dissolved in water

Inducer

IPTG-stock solution	1 M IPTG dissolved in water, sterile filtered and stored at -20 °C
L-Arabinose-stock solution	20% (w/v) dissolved in water, sterile filtered and stored at room temperature

Agarose gel electrophoresis

Agarose (1%)	1 g agarose dissolved in 100 mL 1x TAE, boiled and stored at 60 °C
TAE (1x)	40 mM Tris-base, 20 mM acetic acid, 1 mM EDTA
6x purple gel loading dye	New England Biolabs GmbH, Frankfurt a. M

DNA modification

dNTP solution (10 mM)	New England Biolabs GmbH, Frankfurt a. M

SDS-Polyacrylamide electrophoresis

Ammonium persulfate (APS) stock solution (10%)	10% (w/v) APS dissolved in water and stored at -20 °C
Coomassie Staining Dye	110 mg·L^{-1} Coomassie BlueG-250, 50 g·L^{-1} ethanol, 80 g·L^{-1} phosphoric acid, 10 g·L^{-1} β-Cyclodextrin
SDS-PAGE separating gel (13.5%)	333 mM Bis-Tris pH 6.5, 13.5% 29:1 acrylamid-bisacrylamide
SDS-PAGE separating gel (8%)	333 mM Bis-Tris pH 6.5, 8% 29:1 acrylamid-bisacrylamide
SDS-PAGE stacking gel (5%)	333 mM Bis-Tris pH 6.5, 5% 29:1 acrylamide-bisacrylamide
1x MES running buffer	20x MES-SDS running buffer (Thermo Scientific)
SDS-PAGE sample buffer (2x)	100 mM Tris pH 6.8, 20% glycerol, 4% SDS, 200 mM DTT, 1 µg bromophenol blue

Protein Biochemistry

Buffer for Ni^{2+}-Metal affinity chromatography

Running Buffer (Buffer A_{IMAC})	50 mM sodium phosphate buffer, pH 7.4, 300 mM NaCl, 50 mM Imidazole
Elution Buffer (Buffer B_{IMAC})	50 mM sodium phosphate buffer, pH 7.4, 300 mM NaCl, 500 mM Imidazole

Buffer for reverse Ni^{2+}-Metal affinity chromatography

Running Buffer (Buffer C)	50 mM sodium phosphate buffer, pH 7.4, 300 mM NaCl
Elution Buffer (Buffer B_{IMAC})	50 mM sodium phosphate buffer, pH 7.4, 300 mM NaCl, 500 mM Imidazole

Buffer for Strep-Trap affinity chromatography

Running Buffer (Buffer A_{Strep})	50 mM sodium phosphate buffer, pH 8.0, 300 mM NaCl, 1 mM EDTA
Elution Buffer (Buffer B_{Strep})	50 mM sodium phosphate buffer, pH 8.0, 300 mM NaCl, 1 mM EDTA, 50 mM biotin
Regeneration Buffer	3 M $MgCl_2$

Buffer for SEC, SEC-MALs, DSC, CD, Fluorescence Titration

Buffer C	50 mM sodium phosphate buffer, pH 7.4, 150 mM NaCl

Buffer for SPR

Protein Preparation	25 mM HEPES, pH 7.4, 150 mM NaCl
SPR-Running Buffer	25 mM HEPES, pH 7.4, 200 mM KCl, 5 mM $MgCl_2$, 0.2 mM EDTA, 0.05% Tween
Normalisation solution	70% (v/v) glycerol
Chip cleaning solution	50 mM NaOH
Amine coupling buffer	50 mM sodium acetate, pH 4.5
Strep-Buffer	10 mM HEPES, pH 7.4, 150 mM NaCl, 5 mM $MgCl_2$
Protein regeneration buffer	25 mM HEPES, pH 7.4, 500 mM KCl, 5 mM $MgCl_2$, 0.2 mM EDTA; 0.05% (v/v) Tween
DNA-regeneration buffer	10 mM glycine, pH 2.5

Buffer for X-Ray crystallography

Protein Preparation	50 mM HEPES, pH 7.4, 150 mM NaCl
Crystallisation condition 1	0.1 M HEPES, pH 6.5, 0.8 M $(NH_4)_2SO_4$
Crystallisation condition 2	0.1 M Bis-Tris, pH 5.5, 1 M $(NH_4)_2SO_4$, 1% (w/v) PEG 3350

5.1.13 Software

ASTRA 8.0.2.5	WYATT Technology Europe GmbH, Dernbach
Biacore X Control Software (version 2.2)	Cytiva Europe GmbH, Freiburg im Breisgau
BIAevaluation software (version 4.1)	Cytiva Europe GmbH, Freiburg im Breisgau
MicroCal	Malvern Panalytical GmbH, Kassel
Spectra Manager	JASCO Deutschland GmbH, Pfungstadt
UNICORN	Cytiva Europe GmbH, Freiburg im Breisgau

5.2 Methods

5.2.1 Experimental Methods

5.2.1.1 Molecular Biology Methods

5.2.1.1.1 Site-Directed Mutagenesis (SDM) for the generation of point mutants

Site-directed mutagenesis (SDM) was employed to introduce point mutations into target genes. To incorporate mutations into the wild-type *lacI* gene, the pBAD vector was used as the template to avoid unintended alterations within the lac regulatory element. For designed variants or constructs not derived from *lacI,* pET21b(+) or pDuet^{M} vectors carrying the respective gene of interest were used as the template vectors.

Polymerase chain reaction (PCR) was performed in a total volume of 50 µL using the components and final concentrations listed in Table 26. Primers were designed using the QuikChange primer design tool (Agilent Technologies)[490] and are listed in Table 18. PCR cycling conditions used for SDM are summarised in Table 27.

Table 26. Composition of PCR for site-directed mutagenesis

Components	Final Concentration
Template	~ 100 ng
Forward (fw)-Primer	0.3 µM
Reverse (rv)-Primer	0.3 µM
dNTPs	200 µM
KAPA HiFi Fidelity Buffer	1x
KAPA HiFi DNA Polymerase	1 U
PCR-grade water	ad 50 µL

Table 27. PCR cycling parameters for site-directed mutagenesis

Step	Temperature	Duration	Cycles
Initial denaturation	95 °C	3 min	1
Denaturation	98 °C	20 s	
Annealing	*	15 s	30
Extension	72 °C	45 s/kb	
Final Extension	72 °C	10 min	1

* The annealing temperature was adjusted for each primer pair and set to 3 °C below the calculated melting temperature (Oligo Analysis Tool, Eurofins Genomics)[491]. Primers are summarised in Table 18.

Following PCR amplification, the parental methylated template DNA was selectively digested by adding 1 µL *DpnI* (20 U·µL^{-1}) and incubating at 37 °C for 1 h. Site-directed mutagenesis reaction mixtures generated using pET21b(+), pBAD/HisA, or pDUET^{M} as templates were

used to transform Top10 cells, while those obtained with pCDF-pBAD as the template were used to transform DH5α cells (5.2.1.1.12). Successful introduction of desired mutations was verified by plasmid isolation (5.2.1.1.10) and subsequent Sanger sequencing (5.2.1.1.11).

5.2.1.1.2 Polymerase Chain Reaction (PCR) for the addition of restriction enzyme recognition sites

The PCR reactions were carried out to amplify short DNA fragments as well as linearised plasmids while simultaneously introducing recognition sequences for specific restriction endonucleases (RE) at the 5' and 3' ends. Reactions were assembled in a total volume of 50 µL, with components and final concentrations detailed in Table 28, while thermal cycling parameters are summarised in Table 29.

Table 28. Composition of PCR for the incorporation of restriction enzyme recognition sites

Components	Final Concentration
Template	~ 100 ng
Forward (fw)-Primer	0.5 µM
Reverse (rv)-Primer	0.5 µM
dNTPs	200 µM
Phusion™ GC-Buffer	1x
Phusion™ High-Fidelity DNA Polymerase	0.5 U
PCR-grade water	ad 50 µL

Table 29. PCR cycling parameters for the incorporation of restriction enzyme recognition sites

Step	Temperature	Duration	Cycles
Initial denaturation	98 °C	30 s	1
Denaturation	98 °C	10 s	
Annealing	*	15 s	30
Extension	72 °C	30 s/kb	
Final Extension	72 °C	10 min	1

* The annealing temperature was adjusted for each primer pair and set to 3 °C below the calculated melting temperature of the primer part, which overlaps with the template DNA (Oligo Analysis Tool, Eurofins Genomics)[491]. Primers used for the amplification are summarised in Table 19-Table 25.

Following amplification, PCR products were digested with appropriate restriction enzymes (5.2.1.1.3) and cleaned (5.2.1.1.5) before ligation (5.2.1.1.4). Successful cloning steps were subsequently verified by colony PCR (5.2.1.1.9) and Sanger sequencing (5.2.1.1.11).

5.2.1.1.3 Restriction Endonuclease (RE) digestion

Purified DNA fragments, obtained as described in Chapter 5.2.1.1.6, as well as plasmid DNA extracted as described in Chapter 5.2.1.1.7, were subjected to digestion with Type II restriction endonucleases (Type II-RE) to generate 5'-cohesive ends. The use of two distinct REs enabled the directional insertion of the DNA fragment into the target plasmid.
The components and final concentrations for the double-digest reaction are summarised in Table 30. Reactions were incubated at 37 °C for 1 h, followed by purification of the digested DNA via agarose gel extraction (5.2.1.1.5) before ligation (5.2.1.1.4). Further, specific combinations of RE were selected to enable directional insertion of DNA fragments into the reading frames of the respective plasmids (Table 31).

Table 30. Composition of cleavage reaction mix for the reaction with restriction endonucleases (RE I and RE II).

Components	Final Concentration
Plasmid or PCR product	1 µg
rCutSmart Buffer	1x
RE I	1 U
RE II	1 U
PCR-grade water	ad 50 µL

Table 31. Summary of restriction endonucleases used for vector-specific cloning.

Vector	RE I	RE II
pET21b(+)	*NdeI*	*XhoI*
pDuet^{M} 1st Reading Frame	*EcoRI*	*SacI*
pDuet^{M} 2nd Reading Frame	*NdeI*	*XhoI*
pCDF-BAD	*NdeI*	*SacI*
pBAD/HisA	*EcoRI*	*XhoI*

5.2.1.1.4 Ligation of DNA fragments with 5' cohesive ends

Before ligation, both vector and insert DNA were purified from agarose gels (5.2.1.1.5). For the ligation of DNA fragments possessing 5' cohesive ends, the reaction was assembled on ice using a molar ratio of 1:3 (vector to insert). The composition of the complete ligation mix is summarised in Table 32. The reaction was incubated at room temperature for 30 min. The entire reaction mixture was used for transformation (5.2.1.1.12).

Table 32. Composition of the ligation reaction for a molar ratio of 1:3

Components	Final Concentration
T4 DNA Ligase Buffer	1x
Vector DNA	0.020 pmol
Insert DNA	0.060 pmol
T4 DNA Ligase	400 U
PCR-grade water	ad 10 µL

5.2.1.1.5 Agarose gel electrophoresis for DNA separation and analysis

A 1% (w/v) agarose gel was prepared by dissolving agarose powder in 1x TAE (tris-acetate-EDTA) buffer through heating until complete dissolution was achieved. After cooling the agarose solution to around 50 °C, Midori Green Xtra DNA stain was added at a final amount of 1 µL per 50 mL gel solution. The agarose was then cast and allowed to solidify at room temperature. DNA samples were combined with 6x purple loading dye before loading onto the gel. Electrophoresis was performed in 1x TAE buffer at a constant voltage of 80 V for 45 min (POWER pro, Cleaver Scientific). DNA bands were visualised using the FastGene FAS-DIGI PRO gel documentation system (Nippon Genetics). Agarose gel electrophoresis was employed both for qualitative assessment of DNA fragments and integrity and for preparative separation followed by gel extraction.

5.2.1.1.6 Purification of DNA in preparative amounts (Midiprep)

For the purification of DNA following PCR amplification, the DNA Clean & Concentrator-100 (DCC-100) Kit from Zymo Research was employed according to the manufacturer's instructions. [492]

5.2.1.1.7 Extraction and purification of DNA from agarose gels

DNA was extracted from agarose gels using the NucleoSpin® Gel and PCR Clean-up from Macherey-Nagel according to the manufacturer's instructions. [493]

5.2.1.1.8 Quantitative determination of DNA concentration

DNA concentrations were determined photometrically by measuring the absorbance at 260 nm using a NanoDrop Plate in combination with a Tecan Spark multimode reader. Based on the Lambert-Beer law, an absorbance value of 1.0 at 260 nm corresponds to a concentration of 50 µg/mL for double-stranded DNA.[494] DNA concentrations were calculated according to equation (1).

$$c_{dsDNA} = \frac{A_{260} \cdot 50 \cdot DF}{1000} \quad (1)$$

c_{dsDNA}: concentration dsDNA, A_{260}: Absorption at 260 nm, **DF**: dilution factor

DNA purity was assessed by calculating the absorbance ratios A_{260}/A_{280} and A_{260}/A_{230}. High-quality DNA typically exhibits A_{260}/A_{280} ratios of 1.8 and A_{260}/A_{230} ratios between 2.0 and 2.2, indicating minimal contamination by proteins, salts or organic solvents.[495]

5.2.1.1.9 Colony PCR for screening of recombinant clones

To verify the successful insertion of DNA fragments into the vectors, colony PCR was performed. Primers were selected according to the vector and reading frames (Table 33). Reactions were assembled on ice with the components and concentrations listed in Table 34. Bacterial colonies were dissolved in water before the addition of the remaining elements. Thermal cycling was performed according to the parameters summarised in Table 35. The length of the PCR products was analysed by running a 1% agarose gel. (5.2.1.1.5)

Table 33. Summary of primers used for colony PCR

Vector	fw-Primer	rv-Primer
pET21b(+)	T7	T7term
pDuet^{M} 1st Reading Frame	petup	DuetDOWN1
pDuet^{M} 2nd Reading Frame	DuetUP2	T7term
pCDF-BAD	pBAD-FP	pBAD-RP
pBAD/HisA	pBAD-FP	pBAD-RP

Table 34. Colony PCR reaction composition

Components	Final Concentration
Bacterial Colony	
fw-Primer	0.5 µM
rv-Primer	0.5 µM
dNTPs	200 µM
Standard *Taq* Reaction Buffer	1x
Taq DNA Polymerase	2 U
PCR-grade water	ad 20 µL

Table 35. Colony PCR cycling parameters

Step	Temperature	Duration	Cycles
Initial denaturation	95 °C	5 min	1
Denaturation	95 °C	20 s	
Annealing	55 °C	15 s	30
Extension	72 °C	1 min/kb	
Final Extension	72 °C	10 min	1

5.2.1.1.10 Isolation of plasmid DNA from bacterial cell cultures

Single bacterial clones were used to inoculate 5 mL LB medium supplemented with appropriate antibiotic (ampicillin, 100 µg·mL^{-1}; streptomycin, 50 µg·mL^{-1}). Cultures were incubated at 37 °C with shaking at 180 rpm for 6 h or overnight. Cells were harvested by centrifugation at 4000 x g for 5 min at room temperature. Plasmid DNA was subsequently extracted using the NucleoSpin® Plasmid Kit from Macherey-Nagel according to the manufacturer's instructions.[496]

5.2.1.1.11 Sanger Sequencing

Plasmid DNA (~100 µg·mL^{-1}) was submitted to Eurofins Genomics for Sanger sequencing. Sequencing was performed using primers listed in Table 8. All primers were supplied by Eurofins during the sequencing process, except for pBAD-RP, which was premixed with the template at a final concentration of 13.3 µM.

5.2.1.1.12 Transformation of bacterial cells

Synthesised plasmids, SDM reaction products, and ligation mixtures were used to transform *E.coli* cells. Strain selection was based on the intended downstream application: *E.coli* DH5α or Top10 strains were used for plasmid amplification, whereas *E.coli* BL21 cells were employed for protein expression. For each transformation, 100 ng of plasmid DNA or the complete SDM or ligation reaction mixture was incubated with chemically competent cells, previously stored at -80 °C, for 5 min on ice. Cells were subsequently subjected to heat shock at 42 °C for 45 s, followed by incubation on ice for 10 min. After the addition of 800 µL Lysogeny Broth (LB) medium, cells were allowed to recover by shaking at 37 °C and 800 rpm for 1 h. Finally, cells were plated onto LB agar plates supplemented with appropriate antibiotics depending on the used plasmid (ampicillin: 100 µg·mL^{-1}, streptomycin: 50 µg·mL^{-1}).

5.2.1.2 Protein biochemistry methods

5.2.1.2.1 Heterologous gene expression in bacterial systems

Overnight starting cultures were grown at 30 °C in 50 mL LB medium in Erlenmeyer flasks with appropriate antibiotics (ampicillin: 100 µg·mL^{-1}, streptomycin: 50 µg·mL^{-1}). For protein expression, 1 L LB medium in a 2 L baffled Erlenmeyer flask was inoculated with 10 mL overnight culture and incubated at 37 °C until OD_{600} reached a value of 0.7. Protein expression was induced by the addition of either β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.2 mM or L-arabinose to a final concentration of 0.2% (w/v), depending on the respective promotor. Following induction, cultures were incubated at 20 °C with shaking at 160 rpm for 20 h. Cells were harvested by centrifugation at 4000 x g at 15 °C for 15 min, washed once with buffer C (50 mM sodium phosphate buffer, pH 7.4, 150 mM NaCl), and stored at -20 °C until further use.

5.2.1.2.2 Cell lysis with sonication

Stored cell pellets were resuspended in buffer A. Either lysis buffer for immobilised Ni^{2+} affinity chromatography (50 mM sodium phosphate buffer pH 7.4, 300 mM NaCl, 50 mM imidazole, Buffer A_{IMAC}) or Strep-Tactin affinity chromatography (50 mM sodium phosphate buffer pH 8.0, 300 mM NaCl, 1 mM EDTA, Buffer A_{Strep}) was used, depending on the first purification step. A summary of the purification steps performed for each construct is found in Table 37. Resuspended cells were lysed by sonication (Sonifier 102C (CE), Branson Ultrasonics) using a tapered microtip (tip diameter 6.5 mm) (Branson Ultrasonics) at an output setting of 4 and a duty of 40%. Sonication was performed for three cycles of 3 min each on ice, with 3 min cooling intervals between cycles. Lysates were clarified by centrifugation at 18000 x g and 4 °C for 1 h. The sample to be purified was filtered through a 0.45 µm filter before application to the column.

5.2.1.2.3 Immobilised Ni^{2+}-Affinity Chromatography (IMAC)

The clarified lysate or heterodimeric LacI variant after the first purification step (5.2.1.2.5) was applied to a 5 mL HisTrapHP column (GE HealthCare) equilibrated with buffer A_{IMAC} (50 mM sodium phosphate buffer pH 7.4, 300 mM NaCl, 50 mM imidazole) operated on the ÄKTApure system (Cytiva) at a flow rate of 5 mL·min^{-1}. After washing the column with 15 column volumes (CV) buffer A_{IMAC} and 5 CV with 8% buffer B_{IMAC} (50 mM sodium phosphate buffer pH 7.4, 300 mM NaCl, 500 mM imidazole), the protein was eluted with a linear gradient over 20 CV to 80% buffer B_{IMAC}. Fractions containing the protein of interest were pooled, concentrated with centrifugal concentrators (10 kDa molecular weight cutoff (MWCO), Merck), and subsequently either subjected to TEV protease cleavage (5.2.1.2.7) or to further purification by size exclusion chromatography (5.2.1.2.6), depending on features present in the protein of interest. A

summary of all purification steps performed in order is found in Table 37.

5.2.1.2.4 Reverse Immobilised Ni^{2+}-Affinity Chromatography (rvIMAC)

TEV protease-cleaved proteins (5.2.1.2.7) were applied to a 5 mL HisTrapHP column (GE HealthCare) equilibrated with buffer C (50 mM sodium phosphate buffer, 150 mM NaCl) and operated on the ÄKTApure system (Cytiva) at a flow rate of 5 $mL \cdot min^{-1}$. Protein lacking the His_6-tag, including the protein of interest, was collected in the flow-through, and during the wash with 5 CV of buffer C. Remaining His_6-tagged protein fragments, as well as the TEV protease, were retained on the column and eluted with 100% buffer B_{IMAC} over 5 CV.

5.2.1.2.5 Immobilised Strep-Tactin Affinity Chromatography (Strep-Tactin AC)

The supernatant after cell lysis (5.2.1.2.2) was applied to a 5 mL Strep-Tactin®XT 4Flow® FPLC column (IBA Lifescience) equilibrated with buffer A_{Strep} (50 mM sodium phosphate buffer pH 8.0, 300 mM NaCl, 1 mM EDTA), and operated on the ÄKTApure system (Cytiva) at a flow rate of 4 mL min^{-1}. The column was washed with 10 CV of buffer A_{Strep}, and the protein was eluted with 100% buffer B_{Strep} (50 mM sodium phosphate buffer pH 8.0, 300 mM NaCl, 50 mM Biotin, 1 mM EDTA) over 5 CV. Columns were regenerated by sequential washing with 10 CV ddH_2O at 2 $mL \cdot min^{-1}$, followed by 20 CV regeneration solution (3 M $MgCl_2$) at 1 mL min^{-1}, and a final 10 CV of ddH_2O at 2 $mL \cdot min^{-1}$. Fractions containing the protein of interest were pooled, concentrated with centrifugal concentrators with a 10 kDa MWCO (Merck), and either subjected to size exclusion chromatography (5.2.1.2.6) or immobilised Ni^{2+} affinity chromatography (5.2.1.2.3) as listed in Table 37.

5.2.1.2.6 Size-Exclusion Chromatography (SEC)

Concentrated protein solutions were loaded onto a HiLoad 26/600 Superdex 75 or HiLoad 26/600 Superdex 200 (Cytiva), depending on the molecular size of the protein of interest, preequilibrated in buffer C (50 mM sodium phosphate buffer pH 7.4, 150 mM NaCl). Elution was performed with 1 CV of buffer C at 2.5 $ml \cdot min^{-1}$. Fractions containing the protein of interest were pooled and concentrated using centrifugal concentrators with a 10 kDa MWCO (Merck). Proteins were either flash-frozen in liquid nitrogen and stored at -80 °C or used directly for analytical and biophysical methods (5.2.1.3).

5.2.1.2.7 TEV Protease-Mediated Cleavage of Affinity Tags

To remove the His_6-purification tag and parts of the splitGFP system attached to A1-splitGFP variants, TEV protease-mediated cleavage was performed overnight at 4 °C. A protein-to-TEV ratio of 1:20 (w/w) was used. The protein of interest and cleavage products, as well as His_6-tagged TEV protease, were separated by reverse immobilised Ni^{2+}-affinity chromatography (5.2.1.2.4).

5.2.1.2.8 Dialysis to exchange the buffer

For experiments performed under buffer conditions other than buffer C, e.g. crystallography or SPR measurements, the buffer was exchanged to the corresponding buffer by performing two dialysis steps. Samples were loaded into dialysis tubing with 6-8 kDa MWCO (Repligen Corporation). Dialysis was performed at 4 °C with continuous gentle stirring of the dialysis buffer overnight for the first step and for 3 h for the second step, with a protein-to-dialysis buffer ratio of 1:200 for each step.

5.2.1.2.9 Ligation of the Tag and Catcher protein family

For the ligation of members from the Tag and Catcher protein family, two different approaches were employed. Heterodimeric constructs were ligated *in vitro* before purification. After cell lysis, supernatants containing the monomeric constructs with Tag and Catcher were mixed and incubated overnight at 4 °C. In a second approach, the irreversible ligation of heterodimeric ‘feet’ with the ‘clapping machine’ was performed *in vitro* after initial individual purification of both components. The ‘feet’ and the ‘clapping machine’ were combined and incubated overnight at 4 °C. Components were combined in a ratio of 1:1.5 ‘foot’ to ‘clapping machine’.

5.2.1.3 Analytical and biophysical methods

5.2.1.3.1 Quantitative determination of protein concentration

The protein concentration was determined from the absorbance at 280 nm measured using a BioPhotometer® (Eppendorf) in a 1 mm pathlength cuvette (Eppendorf). Concentrations were calculated according to Lambert-Beer law (2)) .[497]

$$A_{280} = \varepsilon \cdot c \cdot l \quad (2)$$

A_{280}: Absorbance at 280 nm in [A.U], **c**: protein concentration in [mol·L^{-1}], **l**: pathlength in [cm], **ε**: molar extinction coefficient in [L mol^{-1}·cm^{-1}] calculated with the online tool ProtParam tool from expasy[498] and summarised in Table 38

5.2.1.3.2 Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE was used to analyse fractions from different purification steps as well as to assess the successful ligation of Tag and Catcher tagged proteins. Depending on the protein size, either 13.5% (333 mM Bis-Tris pH 6.5, 13.5% 29:1 acrylamid-bisacrylamide) or 8% separating gels (333 mM Bis-Tris pH 6.5, 8% 29:1 acrylamid-bisacrylamide) were employed. Electrophoresis was performed at a constant current of 70 mA per gel for 35 min. Gels were stained with Coomassie staining dye (110 mg·L^{-1} Coomassie BlueG-250, 50 g·L^{-1} ethanol, 80 g·L^{-1} phosphoric acid, 10 g·L^{-1} β-Cyclodextrin) at room temperature and imaged with E-Box VX 20M Gel Imaging System (Vilber GmbH).

5.2.1.3.3 Secondary structure analysis with far-ultraviolet Circular Dichroism (CD) spectroscopy

CD spectra were collected with a JASCO J-1500. Experiments were performed in buffer C (50 mM sodium phosphate buffer pH 7.4, 150 mM NaCl) using a protein concentration of 0.2 mg·mL^{-1}. Far-Ultraviolet-CD spectra were recorded in the range of 200-260 nm at 20 °C in a 1 mm cuvette (Hellma) with a 1 nm bandwidth, 1 s response time and scanning speed of 100 nm min^{-1}. For each protein, 5 spectra were accumulated. Data were normalised by subtraction of a buffer spectrum and conversion to mean residue molar ellipticity (equation (3)).

$$[\theta_{MRE}] = \frac{\theta_{obs} \cdot M \cdot 100}{d \cdot c \cdot n} \quad (3)$$

θ_{MRE}: molar ellipticity per residue in [deg·cm^2·dmol^{-1}], **θ_{obs}**: measured ellipticity in [mdeg], **M**: molecular weight in [Da] (summarised in Table 38), **d**: cuvette pathlength in [cm], **c**: concentration of the protein in [mg·mL^{-1}], **n**: number of amino acids of the protein (summarised in Table 38)

Thermal unfolding was assessed by monitoring the temperature-dependent change in molar ellipticity at 222 nm, a characteristic marker of α-helical unfolding. Samples were heated up to 90 °C with a rate of 1 °C ·min^{-1}. Measured unfolding curves were normalised as shown above. The melting temperature (T_m) of the protein was determined by fitting the unfolding curve to a

sigmoidal Boltzmann function (equation (4)). T_m was defined as the temperature at which 50% of the protein was unfolded, corresponding to the midpoint of the fitted curve.

$$\theta_{MRE}[T] = A + \frac{B - A}{1 + e^{\left(\frac{T_m - T}{k}\right)}} \quad (4)$$

θ_{MRE}: molar ellipticity per residue [deg cm^2 dmol^{-1}], **T**: temperature in [°C], **T_m**: melting temperature in [°C], **A**: lower plateau of sigmoid curve in [deg cm^2 dmol^{-1}], **B**: higher plateau of sigmoid curve [deg cm^2 dmol^{-1}], **k**: slope [°C],

5.2.1.3.4 Differential Scanning Calorimetry (DSC) for thermodynamic stability assessment

For DSC measurements, proteins were prepared at a concentration of 0.5 mg·mL^{-1} in buffer C. Measurements were performed using a MicroCal Peaq-DSC system (Malvern) at a scanning speed of 90 K·h^{-1} in non-feedback mode. Before each protein-buffer scan, two buffer-buffer scans were conducted for instrument equilibration. The buffer-buffer baseline was subtracted from each endotherm, and the resulting thermograms were normalised to the protein concentration. Data analysis was performed using MicroCal software (Malvern). For measurements in the presence of the ligand, ligands were used at a final concentration of 1 mM and were premixed with the protein before the measurement.

5.2.1.3.5 Analytical Size-Exclusion Chromatography (SEC) for oligomeric state determination

Protein samples (200 µL, 0.5 mg·mL^{-1}) were loaded onto a Superdex 200 Increase 10/300 GL column (Cytiva). Columns were equilibrated in buffer C. Elution was performed with 1 CV buffer C at a flow rate of 0.8 mg·mL^{-1}. For column calibration, the Gel Filtration Calibration Kit low molecular weight (Cytiva) was used. A mixture of 0.5 mg·mL^{-1} Blue Dextran 2000, Aprotinin, Ribonuclease A, Carbonic anhydrase, Ovalbumin, and Conalbumin, each at 0.5 mg·mL^{-1}, was prepared (total volume 200 µL) and used for column calibration. The distribution coefficient (K_{av}) of each protein from the calibration kit was calculated according to equation (5).

$$K_{av} = \frac{V_e - V_0}{V_c - V_0} \quad (5)$$

K_{av}: distribution coefficient, **V_e**: elution volume of protein in [mL], **V_0**: void volume in [mL], **V_c**: total column volume in [mL]

A calibration curve of K_{av} versus the logarithm of the molecular weight of the proteins was prepared using a semi-logarithmic plot. The molecular weight of the protein of interest was determined from the regression curve.

5.2.1.3.6 Size-Exclusion Chromatography with Multi-Angle Light Scattering (SEC-MALS) for absolute molecular mass determination

SEC-MALS measurements were performed using either a Superdex 75 Increase 10/300 GL column or a Superdex 200 Increase 10/300 GL column (Cytiva), depending on the molecular weight of the proteins of interest. Columns connected to an Agilent 1260 Infinity II HPLC system (Wyatt) coupled to a miniDAWN multiangle light scattering detector (MALS) and an Optilab differential refractive index detector (dRI) were equilibrated with buffer C. Samples were injected at a protein concentration of 0.5 mg·mL^{-1}, with an injection volume of 100 µL. The flow rate was set to 0.8 mL·min^{-1}. Bovine serum albumin (BSA, 2 mg·mL^{-1}) was used as a standard to normalise the MALS detector, to correct for peak alignment and broadening, and to verify reproducibility. Data collection and analysis were performed using the ASTRA 8.0.2.5 software (Wyatt). Measurements were conducted at room temperature, and the protein-specific refractive index increment (dn/dc) of 0.185 mL·g^{-1} was used for molar mass calculation.

5.2.1.3.7 Fluorescence-based titration assay for ligand binding analysis

Fluorescence measurements were performed using a JASCO FP-6500 Spectrofluorometer. Excitation was set at 280 nm with a 3 nm bandwidth, and emission spectra were recorded from 300 to 450 nm with a scanning speed of 500 nm·min^{-1} and a bandwidth of 5 nm. Protein samples (170 µL, 0.2 mg·mL^{-1}) were subjected to cumulative ligand titration by stepwise addition of ligand solution in 1 min intervals. Fluorescence signals arising from buffer addition alone were used for background correction and subtracted from all measurements. The change in fluorescence intensity at 360 nm was used to determine the binding constant by fitting the data to a hyperbolic binding model (equation (6)).

$$\Delta F_{360} = \Delta F_{max,360} \cdot \frac{[L]_{free}}{K_d + [L]_{free}} \quad (6)$$

ΔF_{360}: Change in Fluorescence at ligand concentration in [A.U], **$\Delta F_{max,360}$**: maximal change of Fluorescence at saturation in [A.U], **$[L]_{free}$**: free ligand concentration in [µM] with the assumption that [L]free ≈ $[L]_{total}$, **K_d**: dissociation constant in [µM]

5.2.1.3.8 Measurement of GFP fluorescence in splitGFP complementation assays

GFP fluorescence was measured directly from bacterial cultures using a Tecan Spark plate reader. For each measurement, 150 µL of cell culture was transferred into clear-bottom 96-well plates. The absorbance at 600 nm was recorded to determine the corresponding OD600 values. GFP fluorescence was then measured in the same culture samples without prior cell lysis or additional sample processing. Fluorescence was recorded after excitation at 485 nm, and emission was detected at 520 nm. The excitation and emission bandwidths were set to 10 nm. A dichroic 510 mirror was used for wavelength separation. For each well, 30 flashes

were recorded with an integration time of 40 µs. Gain and z-position were adjusted manually for the respective measurement.

To account for differences in cell density, GFP fluorescence values were normalised to the corresponding OD600 values (equation 7).

$$F_{520,norm} = \frac{E_{520}}{OD_{600}} \quad (7)$$

$F_{520,norm}$: normalised fluorescence at 520 nm, **E_{520}:** Emission at 520 nm, **OD_{600}**: Absorbance at 600 nm (optical density)

5.2.1.3.9 Surface Plasmon Resonance (SPR) for DNA binding interactions

All SPR experiments were conducted by Dr. Vanessa Troßmann. Buffers were filtered through 0.22 µm membrane filters and degassed before use. SPR measurements were performed on a BiacoreX instrument at 25 °C using the Biacore X Control Software (version 2.2).

Before functionalisation, the chip was normalised using 70% (w/w) glycerol and subsequently equilibrated in running buffer (25 mM HEPES, 200 mM KCl, 5 mM MgCl2, 0.2 mM EDTA, 0.05% (v/v) Tween, pH 7.4) until a stable baseline was obtained. The surface was cleaned by three injections of 50 mM NaOH and activated by using a 1:1 EDC/NHS mixture for 10 min at a flow rate of 10 µL·min^{-1}, resulting in a signal increase of at least 100 response units (RU). For streptavidin immobilisation, a fresh coupling solution was prepared from 70 µL of amine coupling buffer (50 mM sodium acetate, pH 4.5), 60 µL of water, and 10 µL of streptavidin stock solution (5 mg·mL^{-1} in 10 mM HEPES, 150 mM NaCl, 5 mM MgCl2, pH 7.4). Streptavidin (final concentration: 0.5 mg·mL^{-1}) was immobilised by amine coupling for 10 min resulting in a change of the response by ~4000 RU, followed by blocking of remaining active esters with 1 M ethanolamine for 10 min.

Biotinylated single-stranded DNA (anchor strand, 10 µM in running buffer) was injected until saturation (increase of ~4000 RU), and unbound DNA was removed by washing with 10 mM glycine (pH 2.5). Subsequently, double-stranded DNA constructs (250 nM), containing a single-stranded overhang complementary to the immobilised anchor strand, were hybridised to the surface of the non-reference cell at flow rates between 10 and 30 µL·min^{-1}, yielding an additional increase of the response of 50-120 RU.

The DNA constructs contained either a symmetric *lac* or *gal* operator or an asymmetric *lac/gal* hybrid sequence (Table 36).

Table 36. DNA sequences used for SPR measurements , with the ssDNA overhang complementary to the anchor strand highlighted in pink and the DNA-binding sequences for symmetric and asymmetric LacI shown in green.

DNA construct description	**Sequence 5'-3'**
Anchor	ACCTTAGGGCGTGG-Bio
Symmetric lacO with anchor-complementary overhang	CCACGCCCTAAGGTCTATAGGGGAATTGTGAGCGGATAACAATTCCCCATCTT
Complementary lacO DNA	AAGATGGGGAATTGTTATCCGCTCACAATTCCCCTATAG
Symmetric galO with anchor-complementary overhang	CCACGCCCTAAGGTCTATAGGGGGTAAGCGCTTACCCCCATCTT
Complementary galO DNA	AAGATGGGGGGTAAGCGCTTACCCCCTATAG
Asymmetric lacO/galR DNA with anchor-complementary overhang	CCACGCCCTAAGGTCTATAGGGGAATTTTAAGCGCTCACAATTCCCCATCTT
Complementary lacO/galR DNA	AAGATGGGGAATTGTGAGCGCTTAAAATTCCCCTATAG

Protein-DNA interaction measurements were performed at a flow of 30 µL·min^{-1} using single-cycle kinetics. Proteins were prepared in 25 mM HEPES (pH 7.4), 150 mM NaCl. As 50 µL samples were applied to the system for 60 s (association phase), followed by running buffer for at least 60 s (dissociation phase). Association and dissociation phases were recorded in real time, and reference subtraction was applied to correct for bulk refractive index changes and nonspecific binding.

DNA regeneration was achieved by injection of 1 mM IPTG or 1 mM T6P (15-60 s, 30 µL min^{-1}) to remove the bound protein. If necessary, additional regeneration was performed using the running buffer supplemented by 0.5 M KCl until the baseline was restored. To enable reuse of the modified CM5 chip with different DNA constructs, hybridised double-stranded DNA templates were removed by three injections of 10 mM glycine buffer (pH 2.5) at flow rates of 30 to 100 µL·min^{-1} for 15 to 60 s.

Kinetic parameters were determined from concentration-dependent single-cycle measurements using at least five protein concentrations (starting from 1 to 3 µM, 1:1 serial dilutions). Data were analysed using the BIAevaluation software (version 4.1). Buffer- and reference-corrected curves were normalised manually to the same injection time and baseline. Association and dissociation phases were defined manually. To fit the SPR data, a 1:1 binding model with mass transfer and a constant RI of 0 was applied to determine R_{max}, k_{on}, k_{off}, and K_D (equation 8, equation 9).

$$\frac{dR}{dt} = k_{on} C_s (R_{max} - R) - k_{off} R \quad (8)$$

dR/dt: rate of change of the SPR response over time, $\mathbf{k_{on}}$: association rate constant, $\mathbf{C_s}$: Analyte concentration at the sensor surface, $\mathbf{R_{max}}$: maximal SPR response, **R**: measure SPR response, $\mathbf{k_{off}}$: dissociation rate constant

$$K_D = \frac{k_{off}}{k_{on}} \quad (9)$$

$\mathbf{K_D}$: equilibrium dissociation constant, $\mathbf{k_{on}}$: association rate constant, $\mathbf{k_{off}}$: dissociation rate constant

5.2.1.3.10 X-Ray crystallography for high-resolution structure determination

Initial crystallisation screens were performed using commercially available sparse-matrix screens (JSGC Core I-IV and JCSG+ Suite) in 96-well sitting drop plates (3-drop Intelli-Plates, Hampton Research). Drops were prepared by pipetting protein and reservoir solution in 1:1, 1:2 and 2:1 volume ratios, using a Phoenix pipetting robot (Art Robbins Instruments). Before using proteins for crystallisation, they were dialysed in 25 mM HEPES pH 7.4, 150 mL NaCl (5.2.1.2.8). Protein concentrations of 15 mg·mL^{-1} for the variants PBP1 and 10 mg·mL^{-1} for PBP2 were used. LacI-Lev1 and LacI-Lev2 were screened at a concentration of 10 mg·mL^{-1} in the presence of 50 nM double-stranded DNA containing the operator sequence O1 with a total length of 50 bp. Crystallisation plates were incubated at 293 K. Initial hits for PBP1 were observed after 7 days at the following conditions: 1) 0.1 M HEPES pH 6.5, 0.8 M $(NH_4)_2SO_4$, 2) 0.1 M Bis-Tris pH 5.5, 1 M $(NH_4)_2SO_4$, 1% (w/v) 3350 PEG.

Conditions were further optimised using the initial hit and setting up the hanging drop vapour diffusion method in 24-well VDX Plate with sealant (Hampton Research). 5 µL drops were set up at a protein-to-well ratio of 1:2. The best diffracting crystals were obtained at 8.0 mg·mL^{-1} PBP1 in 0.1 M HEPES pH 6.5, 0.8 M $(NH_4)_2SO_4$, supplemented with 0.5 µL H_2O. Crystals reached dimensions of 100-200 µm and were manually mounted using nylon cryo-loops and cryoprotected by brief immersion in mother liquor supplemented with 20% ethylene glycol before flash-freezing with liquid nitrogen. X-ray diffraction data were collected at 100 K at beamlines BL 14.1 and 14.2 at the synchrotron BESSY II, Helmholtz-Zentrum Berlin.[499]

Phases were solved by molecular replacement using *Phaser*[500] with the corresponding AlphaFold2[501] structure as the search model. The resulting models were manually inspected and rebuilt using *Coot*[502]*,* followed by iterative refinement using *phenix.refine*[503]*.* Refinement cycles included positional, individual B-factor, and TLS refinement as appropriate. Model quality was assessed throughout refinement using standard validation metrics.

5.2.2 Computational Methods

5.2.2.1 Calculation of dihedral angles

Dihedral angles were calculated using BioPhyton[504]. Structural models representing the open (1LBG)[28] and the closed (2P9H)[30] conformational states of LacI were first aligned based on Cα-atoms. Backbone dihedral angles Φ (phi) and ψ (psi) were computed for each amino acid residue (i) using standard atom definitions:

$$\phi_i = \prec (C'_{i-1}, N_i, C_{\propto i}, C'_i) \tag{10}$$

Φ_i: phi angle of amino acid i, **C'**: carbonyl carbon, **Cα**: alpha carbon atom, **N**: backbone amide nitrogen, **i-1**: of the receding residue i-1, **i**: number of amino acid investigated

$$\psi_i = \prec (N_i, C_{\propto i}, C'_i, N_{i+1}) \tag{11}$$

Ψ_i: psi angle of amino acid i, **C'**: carbonyl carbon, Cα: alpha carbon atom, N: nitrogen, i-1: of the receding residue i-1, i: amino acid investigated

Dihedral angles were calculated using a vector-based arctangent formulation and reported in degrees within the range (−180°, +180°]. Residues lacking the required atoms, including terminal residues, were excluded from the analysis. Residue-wise conformational changes between the open and closed states were quantified by calculating the differences in backbone dihedral angles:

$$\Delta\phi_i = \phi_i^{\text{closed}} - \phi_i^{\text{open}}, \Delta\psi_i = \psi_i^{\text{closed}} - \psi_i^{\text{open}} \tag{12}$$

$\Delta\Phi_i$: change in phi angle of residue i, **$\Delta\Psi_i$**: change of psi angle of residue i, **closed**: closed conformation, **open**: open conformation, **i**: number of amino acid investigated

For each residue, the overall backbone dihedral change was defined as the sum of the absolute changes in φ and ψ:

$$\Delta\theta_i = \mid \Delta\phi_i \mid + \mid \Delta\psi_i \mid \tag{13}$$

$\Delta\Theta_i$: combined angular change of the backbone conformation at residue i, **$\Delta\Phi_i$**: change in phi angle of residue i, **$\Delta\Psi_i$**: change of psi angle of residue i

Δθ values were used to identify regions undergoing pronounced backbone rearrangements between the open and closed conformational states.

5.2.2.2 Design of backbone extension in the PBP lobe from the LacI-GalR protein family

The RFdiffusion generative model[257] was used to generate structured insertions (10-25 AS) within LacI. Insertions were designed either at the C-terminus or at the insertion site identified by residue-wise dihedral angle analysis (Δθ) (5.2.2.1). Structure generation was performed using RFdiffusion inference with chain A of the experimentally derived PBP-domain of LacI as an input (PBD: 2P9H[30]). 250 potential structures were sampled. Generated designs were screened for the presence of a well-defined secondary structure element within the newly

designed segment. Candidates were retained if the insertion formed a continuous α-helix or β-strand/β-hairpin architecture and did not exhibit obvious backbone discontinuities or severe steric clashes upon visual inspection.

5.2.2.3 Sequence design of backbone extensions and screening

Filtered backbone models generated with RFdiffusion[257] (5.2.2.2) contained diverse insertions composed of poly-glycine residues. To assign amino acid sequences to these designed backbone extensions, ProteinMPNN[252] was applied.

The sequence of the PBP-domain was fixed and kept identical to the wild-type. For each backbone, 20 sequences were sampled using a sampling temperature of 0.3 and backbone noise of 0.2 to allow for sequences with a more conservative sequence design and limited exploration around the provided backbone geometry.

Designed sequences were subsequently filtered based on their amino acid composition to reduce the likelihood of aggregation-prone and unstructured variants. For each sequence, the relative content of aromatic, hydrophobic, charged residues and alanine was calculated. Only sequences whose composition in each category fell within ± standard deviation of the corresponding distribution of the AS features across the backbone extension were retained. Sequences passing the filter were selected for downstream structural evaluation and experimental testing.

5.2.2.4 Structure prediction and evaluation

Two different structure prediction methods were employed:

1) Protein structure predictions for designed sequences were generated using AlphaFold2[501] implemented through ColabFold[505]. ColabFold enabled efficient structure prediction by replacing the standard AlphaFold2 homology search with a substantially faster MMseqs2-based workflow, while retaining AlphaFold2's neural network inference for structure generation.[258] Structure prediction was performed using ColabFold with default settings, which include MMseqs2-based MSA generation, five AlphaFold2 models with three recycles, standard confidence-based model ranking, and no Amber relaxation. Predicted structures were evaluated based on the model score pLDDT (predicted Local Distance Difference Test) provided. Only designs with a mean pLDDT of above 80 for the backbone extension were retained for further analysis.

2) The three-dimensional structures of ProteinMPNN-designed sequences were predicted using Chai[259]. For each designed sequence, structure prediction was performed by using three recycles and 200 diffusion timesteps. Using the Chai MSA server, multiple sequence alignments were used for structure prediction. Additionally, ESM models were incorporated into the prediction to enhance prediction accuracy. Predicted structures were evaluated based on

the model confidence score ipTM provided by Chai. Only designs exceeding an ipTM of the backbone extension of 0.8 were retained for further analysis.

To assess structural compatibility with the designed backbone, the predicted models were aligned to the corresponding RFdiffusion-generated backbone, and the root-mean-square deviation[506] (RMSD) over backbone atoms was calculated.

Designs with a backbone RMSD of ≤ 4.0 Å relative to the backbone designed with RFdiffusion were considered structurally consistent and retained. Final candidate selection was based on a combination of predicted structure quality scores, backbone RMSD, and visual inspection to assess secondary structure integrity and overall fold plausibility. Up to five designs meeting all criteria were selected for experimental characterisation.

5.2.2.5 Addition of Tag/Catcher building blocks into the backbone extension

For later attachment of DogTag or SpyCatcher to backbone extensions, the pipeline consisting of RFdiffusion (5.2.2.2), ProteinMPNN (5.2.2.3) and structure prediction (5.2.2.4) was reused. However, settings were adjusted to the specific case. The length diffusing region was set to 5-10 amino acids. As an insertion site, the ends of the alpha-helical extensions were used. Otherwise, the same settings were kept as before.

6 References

(1) Nilsson, P. F.; Zink, A.; Laprévote, O. M. C.; Korosec, C. S.; Gustafsson, N.; Robertson, N. O.; Forde, N. R.; Eichhorn, R.; Höcker, B.; Curmi, P. M. G.; Linke, H. Walking by Design: How to Build Artificial Molecular Motors Made of Proteins. *Nano Futur.* **2024**, *8* (4), 042501. https://doi.org/10.1088/2399-1984/ad7d7e.

(2) Jacob, F.; Monod, J. Genetic Regulatory Mechanisms in the Synthesis of Proteins. *J. Mol. Biol.* **1961**, *3*, 318–356. https://doi.org/10.1016/s0022-2836(61)80072-7.

(3) Swint-Kruse, L.; Matthews, K. S. Allostery in the LacI/GalR Family: Variations on a Theme. *Curr. Opin. Microbiol.* **2009**, *12* (2), 129–137. https://doi.org/10.1016/j.mib.2009.01.009.

(4) Meng, L. M.; Nygaard, P. Identification of Hypoxanthine and Guanine as the Co-Repressors for the Purine Regulon Genes of Escherichia Coli. *Mol. Microbiol.* **1990**, *4* (12), 2187–2192. https://doi.org/10.1111/j.1365-2958.1990.tb00580.x.

(5) Weickert, M. J.; Adhya, S. A Family of Bacterial Regulators Homologous to Gal and Lac Repressors. *J. Biol. Chem.* **1992**, *267* (22), 15869–15874. https://doi.org/10.1016/S0021-9258(19)49615-4.

(6) Colmer, J. A.; Hamood, A. N. Characterization of ptxS, a Pseudomonas Aeruginosa Gene Which Interferes with the Effect of the Exotoxin A Positive Regulatory Gene, ptxR. *Mol. Gen. Genet. MGG* **1998**, *258* (3), 250–259. https://doi.org/10.1007/s004380050729.

(7) Henkin, T. M. The Role of the CcpA Transcriptional Regulator in Carbon Metabolism in Bacillus Subtilis. *FEMS Microbiol. Lett.* **1996**, *135* (1), 9–15. https://doi.org/10.1111/j.1574-6968.1996.tb07959.x.

(8) Kornberg, H. L. Routes for Fructose Utilization by Escherichia Coli. *J. Mol. Microbiol. Biotechnol.* **2001**, *3* (3), 355–359.

(9) Elsen, S.; Swem, L. R.; Swem, D. L.; Bauer, C. E. RegB/RegA, a Highly Conserved Redox-Responding Global Two-Component Regulatory System. *Microbiol. Mol. Biol. Rev.* **2004**, *68* (2), 263–279. https://doi.org/10.1128/mmbr.68.2.263-279.2004.

(10) Gilbert, W.; Müller-Hill, B. ISOLATION OF THE LAC REPRESSOR. *Proc. Natl. Acad. Sci. U. S. A.* **1966**, *56* (6), 1891–1898. https://doi.org/10.1073/pnas.56.6.1891.

(11) Majumdar, A.; Adhya, S. Demonstration of Two Operator Elements in Gal: In Vitro Repressor Binding Studies. *Proc. Natl. Acad. Sci.* **1984**, *81* (19), 6100–6104.

(12) Elliott, K.; FitzSimons, D. W. *Purine and Pyrimidine Metabolism*; John Wiley & Sons, 2009.

(13) Horlacher, R.; Boos, W. Characterization of TreR, the Major Regulator of the Escherichia coli Trehalose System *. *J. Biol. Chem.* **1997**, *272* (20), 13026–13032. https://doi.org/10.1074/jbc.272.20.13026.

(14) Monod, J.; Wyman, J.; Changeux, J. P. ON THE NATURE OF ALLOSTERIC TRANSITIONS: A PLAUSIBLE MODEL. *J. Mol. Biol.* **1965**, *12*, 88–118. https://doi.org/10.1016/s0022-2836(65)80285-6.

(15) Nguyen, C. C.; Saier, M. H. Phylogenetic, Structural and Functional Analyses of the LacI-GalR Family of Bacterial Transcription Factors. *FEBS Lett.* **1995**, *377* (2), 98–102. https://doi.org/10.1016/0014-5793(95)01344-X.

(16) Harrison, S. C. A Structural Taxonomy of DNA-Binding Domains. *Nature* **1991**, *353* (6346), 715–719. https://doi.org/10.1038/353715a0.

(17) Jones, S.; Barker, J. A.; Nobeli, I.; Thornton, J. M. Using Structural Motif Templates to Identify Proteins with DNA Binding Function. *Nucleic Acids Res.* **2003**, *31* (11), 2811–2823. https://doi.org/10.1093/nar/gkg386.

(18) Luscombe, N. M.; Austin, S. E.; Berman, H. M.; Thornton, J. M. An Overview of the Structures of Protein-DNA Complexes. *Genome Biol.* **2000**, *1* (1), REVIEWS001. https://doi.org/10.1186/gb-2000-1-1-reviews001.

(19) Pabo, C. O.; Sauer, R. T. Transcription Factors: Structural Families and Principles of DNA Recognition. *Annu. Rev. Biochem.* **1992**, *61*, 1053–1095. https://doi.org/10.1146/annurev.bi.61.070192.005201.

(20) Harrison, S. C.; Aggarwal, A. K. DNA RECOGNITION BY PROTEINS WITH THE HELIX-TURN-HELIX MOTIF. *Annu. Rev. Biochem.* **1990**, *59* (Volume 59, 1990), 933–969. https://doi.org/10.1146/annurev.bi.59.070190.004441.
(21) Schumacher, M. A.; Allen, G. S.; Diel, M.; Seidel, G.; Hillen, W.; Brennan, R. G. Structural Basis for Allosteric Control of the Transcription Regulator CcpA by the Phosphoprotein HPr-Ser46-P. *Cell* **2004**, *118* (6), 731–741. https://doi.org/10.1016/j.cell.2004.08.027.
(22) Kalodimos, C. G.; Biris, N.; Bonvin, A. M. J. J.; Levandoski, M. M.; Guennuegues, M.; Boelens, R.; Kaptein, R. Structure and Flexibility Adaptation in Nonspecific and Specific Protein-DNA Complexes. *Science* **2004**, *305* (5682), 386–389. https://doi.org/10.1126/science.1097064.
(23) Schumacher, M. A.; Choi, K. Y.; Lu, F.; Zalkin, H.; Brennan, R. G. Mechanism of Corepressor-Mediated Specific DNA Binding by the Purine Repressor. *Cell* **1995**, *83* (1), 147–155. https://doi.org/10.1016/0092-8674(95)90243-0.
(24) Bondos, S. E.; Swint-Kruse, L.; Matthews, K. S. Flexibility and Disorder in Gene Regulation: LacI/GalR and Hox Proteins *. *J. Biol. Chem.* **2015**, *290* (41), 24669–24677. https://doi.org/10.1074/jbc.R115.685032.
(25) Meinhardt, S.; Manley, M. W.; Becker, N. A.; Hessman, J. A.; Maher, L. J.; Swint-Kruse, L. Novel Insights from Hybrid LacI/GalR Proteins: Family-Wide Functional Attributes and Biologically Significant Variation in Transcription Repression. *Nucleic Acids Res.* **2012**, *40* (21), 11139–11154. https://doi.org/10.1093/nar/gks806.
(26) Campitelli, P.; Swint-Kruse, L.; Ozkan, S. B. Substitutions at Nonconserved Rheostat Positions Modulate Function by Rewiring Long-Range, Dynamic Interactions. *Mol. Biol. Evol.* **2021**, *38* (1), 201–214. https://doi.org/10.1093/molbev/msaa202.
(27) Fukami-Kobayashi, K.; Tateno, Y.; Nishikawa, K. Domain Dislocation: A Change of Core Structure in Periplasmic Binding Proteins in Their Evolutionary History. *J. Mol. Biol.* **1999**, *286* (1), 279–290. https://doi.org/10.1006/jmbi.1998.2454.
(28) Lewis, M.; Chang, G.; Horton, N. C.; Kercher, M. A.; Pace, H. C.; Schumacher, M. A.; Brennan, R. G.; Lu, P. Crystal Structure of the Lactose Operon Repressor and Its Complexes with DNA and Inducer. *Science* **1996**, *271* (5253), 1247–1254. https://doi.org/10.1126/science.271.5253.1247.
(29) Wilson, C. J.; Zhan, H.; Swint-Kruse, L.; Matthews, K. S. The Lactose Repressor System: Paradigms for Regulation, Allosteric Behavior and Protein Folding. *Cell. Mol. Life Sci.* **2007**, *64* (1), 3–16. https://doi.org/10.1007/s00018-006-6296-z.
(30) Daber, R.; Stayrook, S.; Rosenberg, A.; Lewis, M. Structural Analysis of Lac Repressor Bound to Allosteric Effectors. *J. Mol. Biol.* **2007**, *370* (4), 609–619. https://doi.org/10.1016/j.jmb.2007.04.028.
(31) Bairoch, A.; Apweiler, R. The SWISS-PROT Protein Sequence Database and Its Supplement TrEMBL in 2000. *Nucleic Acids Res.* **2000**, *28* (1), 45–48. https://doi.org/10.1093/nar/28.1.45.
(32) Tretyachenko-Ladokhina, V.; Cocco, M. J.; Senear, D. F. Flexibility and Adaptability in Binding of E. Coli Cytidine Repressor to Different Operators Suggests a Role in Differential Gene Regulation. *J. Mol. Biol.* **2006**, *362* (2), 271–286. https://doi.org/10.1016/j.jmb.2006.06.085.
(33) Fukami-Kobayashi, K.; Tateno, Y.; Nishikawa, K. Parallel Evolution of Ligand Specificity Between LacI/GalR Family Repressors and Periplasmic Sugar-Binding Proteins. *Mol. Biol. Evol.* **2003**, *20* (2), 267–277. https://doi.org/10.1093/molbev/msg038.
(34) Miller, J. H.; Schmeissner, U. Genetic Studies of the *Lac* Repressor. *J. Mol. Biol.* **1979**, *131* (2), 223–248. https://doi.org/10.1016/0022-2836(79)90074-3.
(35) Markiewicz, P.; Kleina, L. G.; Cruz, C.; Ehret, S.; Miller, J. H. Genetic Studies of the Lac Repressor. XIV. Analysis of 4000 Altered Escherichia Coli Lac Repressors Reveals Essential and Non-Essential Residues, as Well as “Spacers” Which Do Not Require a Specific Sequence. *J. Mol. Biol.* **1994**, *240* (5), 421–433. https://doi.org/10.1006/jmbi.1994.1458.
(36) Coulondre, C.; Miller, J. H. Genetic Studies of the *Lac* Repressor. *J. Mol. Biol.* **1977**, *117* (3), 525–567. https://doi.org/10.1016/0022-2836(77)90056-0.

(37) Coulondre, C.; Miller, J. H. Genetic Studies of the *Lac* Repressor. *J. Mol. Biol.* **1977**, *117* (3), 577–606. https://doi.org/10.1016/0022-2836(77)90059-6.
(38) Nelson, B. D.; Manoil, C.; Traxler, B. Insertion Mutagenesis of the Lac Repressor and Its Implications for Structure-Function Analysis. *J. Bacteriol.* **1997**, *179* (11), 3721–3728. https://doi.org/10.1128/jb.179.11.3721-3728.1997.
(39) Swint-Kruse, L.; Zhan, H.; Fairbanks, B. M.; Maheshwari, A.; Matthews, K. S. Perturbation from a Distance: Mutations That Alter LacI Function through Long-Range Effects. *Biochemistry* **2003**, *42* (47), 14004–14016. https://doi.org/10.1021/bi035116x.
(40) Glasgow, A.; Hobbs, H. T.; Perry, Z. R.; Wells, M. L.; Marqusee, S.; Kortemme, T. Ligand-Specific Changes in Conformational Flexibility Mediate Long-Range Allostery in the Lac Repressor. *Nat. Commun.* **2023**, *14* (1), 1179. https://doi.org/10.1038/s41467-023-36798-1.
(41) Riggs, A. D.; Suzuki, H.; Bourgeois, S. *Lac* Repressor-Operator Interaction. *J. Mol. Biol.* **1970**, *48* (1), 67–83. https://doi.org/10.1016/0022-2836(70)90219-6.
(42) Riggs, A. D.; Bourgeois, S.; Cohn, M. The Lac Repressor-Operator Interaction. 3. Kinetic Studies. *J. Mol. Biol.* **1970**, *53* (3), 401–417. https://doi.org/10.1016/0022-2836(70)90074-4.
(43) Falcon, C. M.; Matthews, K. S. Engineered Disulfide Linking the Hinge Regions within Lactose Repressor Dimer Increases Operator Affinity, Decreases Sequence Selectivity, and Alters Allostery. *Biochemistry* **2001**, *40* (51), 15650–15659. https://doi.org/10.1021/bi0114067.
(44) Meinhardt, S.; Jr, M. W. M.; Parente, D. J.; Swint-Kruse, L. Rheostats and Toggle Switches for Modulating Protein Function. *PLOS ONE* **2013**, *8* (12), e83502. https://doi.org/10.1371/journal.pone.0083502.
(45) Elf, J.; Li, G.-W.; Xie, X. S. Probing Transcription Factor Dynamics at the Single-Molecule Level in a Living Cell. *Science* **2007**, *316* (5828), 1191–1194. https://doi.org/10.1126/science.1141967.
(46) Hammar, P.; Leroy, P.; Mahmutovic, A.; Marklund, E. G.; Berg, O. G.; Elf, J. The Lac Repressor Displays Facilitated Diffusion in Living Cells. *Science* **2012**, *336* (6088), 1595–1598. https://doi.org/10.1126/science.1221648.
(47) Wilson, C. J.; Zhan, H.; Swint-Kruse, L.; Matthews, K. S. Ligand Interactions with Lactose Repressor Protein and the Repressor-Operator Complex: The Effects of Ionization and Oligomerization on Binding. *Biophys. Chem.* **2007**, *126* (1), 94–105. https://doi.org/10.1016/j.bpc.2006.06.005.
(48) Lamerichs, R. M.; Boelens, R.; Van der Marel, G. A.; Van Boom, J. H.; Kaptein, R. Assignment of the 1H-NMR Spectrum of a Lac Repressor Headpiece-Operator Complex in H2O and Identification of NOEs. Consequences for Protein-DNA Interaction. *Eur. J. Biochem.* **1990**, *194* (2), 629–637. https://doi.org/10.1111/j.1432-1033.1990.tb15662.x.
(49) Flynn, T. C.; Swint-Kruse, L.; Kong, Y.; Booth, C.; Matthews, K. S.; Ma, J. Allosteric Transition Pathways in the Lactose Repressor Protein Core Domains: Asymmetric Motions in a Homodimer. *Protein Sci.* **2003**, *12* (11), 2523–2541. https://doi.org/10.1110/ps.03188303.
(50) Daily, M. D.; Gray, J. J. Allosteric Communication Occurs via Networks of Tertiary and Quaternary Motions in Proteins. *PLOS Comput. Biol.* **2009**, *5* (2), e1000293. https://doi.org/10.1371/journal.pcbi.1000293.
(51) Parente, D. J.; Swint-Kruse, L. Multiple Co-Evolutionary Networks Are Supported by the Common Tertiary Scaffold of the LacI/GalR Proteins. *PLOS ONE* **2013**, *8* (12), e84398. https://doi.org/10.1371/journal.pone.0084398.
(52) Xu, J.; Liu, S.; Chen, M.; Ma, J.; Matthews, K. S. Altering Residues N125 and D149 Impacts Sugar Effector Binding and Allosteric Parameters in Escherichia Coli Lactose Repressor. *Biochemistry* **2011**, *50* (42), 9002–9013. https://doi.org/10.1021/bi200896t.
(53) Elowitz, M. B.; Leibler, S. A Synthetic Oscillatory Network of Transcriptional Regulators. *Nature* **2000**, *403* (6767), 335–338. https://doi.org/10.1038/35002125.

(54) Gardner, T. S.; Cantor, C. R.; Collins, J. J. Construction of a Genetic Toggle Switch in Escherichia Coli. *Nature* **2000**, *403* (6767), 339–342. https://doi.org/10.1038/35002131.
(55) Robinett, C. C.; Straight, A.; Li, G.; Willhelm, C.; Sudlow, G.; Murray, A.; Belmont, A. S. In Vivo Localization of DNA Sequences and Visualization of Large-Scale Chromatin Organization Using Lac Operator/Repressor Recognition. *J. Cell Biol.* **1996**, *135* (6 Pt 2), 1685–1700. https://doi.org/10.1083/jcb.135.6.1685.
(56) Spronk, C. A. E. M.; Slijper, M.; van Boom, J. H.; Kaptein, R.; Boelens, R. Formation of the Hinge Helix in the Lac Represser Is Induced upon Binding to the Lac Operator. *Nat. Struct. Biol.* **1996**, *3* (11), 916–919. https://doi.org/10.1038/nsb1196-916.
(57) Lactose Repressor Protein: Functional Properties and Structure. In *Progress in Nucleic Acid Research and Molecular Biology*; Academic Press, 1997; Vol. 58, pp 127–164. https://doi.org/10.1016/S0079-6603(08)60035-5.
(58) Riggs, A. D.; Newby, R. F.; Bourgeois, S. *Lac* Repressor—Operator Interaction. *J. Mol. Biol.* **1970**, *51* (2), 303–314. https://doi.org/10.1016/0022-2836(70)90144-0.
(59) Oehler, S.; Amouyal, M.; Kolkhof, P.; von Wilcken-Bergmann, B.; Müller-Hill, B. Quality and Position of the Three Lac Operators of E. Coli Define Efficiency of Repression. *EMBO J.* **1994**, *13* (14), 3348–3355. https://doi.org/10.1002/j.1460-2075.1994.tb06637.x.
(60) Borowiec, J. A.; Zhang, L.; Sasse-Dwight, S.; Gralla, J. D. DNA Supercoiling Promotes Formation of a Bent Repression Loop in *Lac* DNA. *J. Mol. Biol.* **1987**, *196* (1), 101–111. https://doi.org/10.1016/0022-2836(87)90513-4.
(61) Flashner, Y.; Gralla, J. D. Dual Mechanism of Repression at a Distance in the Lac Operon. *Proc. Natl. Acad. Sci.* **1988**, *85* (23), 8968–8972. https://doi.org/10.1073/pnas.85.23.8968.
(62) Sasse-Dwight, S.; Gralla, J. D. Probing Co-Operative DNA-Binding in Vivo: The Lac O1: O3 Interaction. *J. Mol. Biol.* **1988**, *202* (1), 107–119. https://doi.org/10.1016/0022-2836(88)90523-2.
(63) Oehler, S.; Eismann, E. R.; Krämer, H.; Müller-Hill, B. The Three Operators of the Lac Operon Cooperate in Repression. *EMBO J.* **1990**, *9* (4), 973–979.
(64) Gilbert, W.; Müller-Hill, B. The Lac Operator Is DNA. *Proc. Natl. Acad. Sci. U. S. A.* **1967**, *58* (6), 2415–2421. https://doi.org/10.1073/pnas.58.6.2415.
(65) Matthews, K. S. DNA Looping. *Microbiol. Rev.* **1992**, *56* (1), 123–136. https://doi.org/10.1128/mr.56.1.123-136.1992.
(66) Rutkauskas, D.; Zhan, H.; Matthews, K. S.; Pavone, F. S.; Vanzi, F. Tetramer Opening in LacI-Mediated DNA Looping. *Proc. Natl. Acad. Sci.* **2009**, *106* (39), 16627–16632. https://doi.org/10.1073/pnas.0904617106.
(67) Bell, C. E.; Lewis, M. A Closer View of the Conformation of the Lac Repressor Bound to Operator. *Nat. Struct. Biol.* **2000**, *7* (3), 209–214. https://doi.org/10.1038/73317.
(68) Daber, R.; Lewis, M. A Novel Molecular Switch. *J. Mol. Biol.* **2009**, *391* (4), 661–670. https://doi.org/10.1016/j.jmb.2009.06.039.
(69) Kalodimos, C. G.; Bonvin, A. M. J. J.; Salinas, R. K.; Wechselberger, R.; Boelens, R.; Kaptein, R. Plasticity in Protein–DNA Recognition: Lac Repressor Interacts with Its Natural Operator O1 through Alternative Conformations of Its DNA-binding Domain. *EMBO J.* **2002**, *21* (12), 2866–2876. https://doi.org/10.1093/emboj/cdf318.
(70) Slijper, M.; Boelens, R.; Davis, A. L.; Konings, R. N. H.; van der Marel, G. A.; van Boom, J. H.; Kaptein, R. Backbone and Side Chain Dynamics of Lac Repressor Headpiece (1−56) and Its Complex with DNA. *Biochemistry* **1997**, *36* (1), 249–254. https://doi.org/10.1021/bi961670d.
(71) Sartorius, J.; Lehming, N.; Kisters, B.; von Wilcken-Bergmann, B.; Müller-Hill, B. Lac Repressor Mutants with Double or Triple Exchanges in the Recognition Helix Bind Specifically to Lac Operator Variants with Multiple Exchanges. *EMBO J.* **1989**, *8* (4), 1265–1270. https://doi.org/10.1002/j.1460-2075.1989.tb03500.x.

(72) Lehming, N.; Sartorius, J.; Kisters-Woike, B. Mutant Lac Repressors with New Specificities Hint at Rules for Protein DNA Recognition.
(73) Lehming, N.; Sartorius, J.; Niemöller, M.; Genenger, G.; V Wilcken-Bergmann, B.; Müller-Hill, B. The Interaction of the Recognition Helix of Lac Repressor with Lac Operator. *EMBO J.* **1987**, *6* (10), 3145–3153. https://doi.org/10.1002/j.1460-2075.1987.tb02625.x.
(74) Lehming, N.; Sartorius, J.; Oehler, S.; Von Wilcken-Bergmann, B.; Müller-Hill, B. Recognition Helices of Lac and Lambda Repressor Are Oriented in Opposite Directions and Recognize Similar DNA Sequences. *Proc. Natl. Acad. Sci.* **1988**, *85* (21), 7947–7951. https://doi.org/10.1073/pnas.85.21.7947.
(75) Chuprina, V. P.; Rullmann, J. A. C.; Lamerichs, R. M. J. N.; van Boom, J. H.; Boelens, R.; Kaptein, R. Structure of the Complex of *Lac* Repressor Headpiece and an 11 Base-Pair Half-Operator Determined by Nuclear Magnetic Resonance Spectroscopy and Restrained Molecular Dynamics. *J. Mol. Biol.* **1993**, *234* (2), 446–462. https://doi.org/10.1006/jmbi.1993.1598.
(76) Zuo, Z.; Stormo, G. D. High-Resolution Specificity from DNA Sequencing Highlights Alternative Modes of Lac Repressor Binding. *Genetics* **2014**, *198* (3), 1329–1343. https://doi.org/10.1534/genetics.114.170100.
(77) Milk, L.; Daber, R.; Lewis, M. Functional Rules for Lac Repressor–Operator Associations and Implications for Protein–DNA Interactions. *Protein Sci. Publ. Protein Soc.* **2010**, *19* (6), 1162–1172. https://doi.org/10.1002/pro.389.
(78) Hsieh, M.; Brenowitz, M. Comparison of the DNA Association Kinetics of the Lac Repressor Tetramer, Its Dimeric Mutant LacI Adi, and the Native Dimeric Gal Repressor *. *J. Biol. Chem.* **1997**, *272* (35), 22092–22096. https://doi.org/10.1074/jbc.272.35.22092.
(79) Berg, O. G.; Winter, R. B.; Von Hippel, P. H. Diffusion-Driven Mechanisms of Protein Translocation on Nucleic Acids. 1. Models and Theory. *Biochemistry* **1981**, *20* (24), 6929–6948. https://doi.org/10.1021/bi00527a028.
(80) Winter, R. B.; Von Hippel, P. H. Diffusion-Driven Mechanisms of Protein Translocation on Nucleic Acids. 2. The Escherichia Coli Lac Repressor-Operator Interaction: Equilibrium Measurements. *Biochemistry* **1981**, *20* (24), 6948–6960. https://doi.org/10.1021/bi00527a029.
(81) Winter, R. B.; Berg, O. G.; Von Hippel, P. H. Diffusion-Driven Mechanisms of Protein Translocation on Nucleic Acids. 3. The Escherichia Coli Lac Repressor-Operator Interaction: Kinetic Measurements and Conclusions. *Biochemistry* **1981**, *20* (24), 6961–6977. https://doi.org/10.1021/bi00527a030.
(82) Wang, Y. M.; Austin, R. H.; Cox, E. C. Single Molecule Measurements of Repressor Protein 1D Diffusion on DNA. *Phys. Rev. Lett.* **2006**, *97* (4), 048302. https://doi.org/10.1103/PhysRevLett.97.048302.
(83) Tempestini, A.; Monico, C.; Gardini, L.; Vanzi, F.; Pavone, F. S.; Capitanio, M. Sliding of a Single Lac Repressor Protein along DNA Is Tuned by DNA Sequence and Molecular Switching. *Nucleic Acids Res.* **2018**, *46* (10), 5001–5011. https://doi.org/10.1093/nar/gky208.
(84) Spolar, R. S.; Record, M. T. Coupling of Local Folding to Site-Specific Binding of Proteins to DNA. *Science* **1994**, *263* (5148), 777–784. https://doi.org/10.1126/science.8303294.
(85) Kalodimos, C. G.; Boelens, R.; Kaptein, R. Toward an Integrated Model of Protein−DNA Recognition as Inferred from NMR Studies on the *L Ac* Repressor System. *Chem. Rev.* **2004**, *104* (8), 3567–3586. https://doi.org/10.1021/cr0304065.
(86) Kopke Salinas, R.; Folkers, G. E.; Bonvin, A. M. J. J.; Das, D.; Boelens, R.; Kaptein, R. Altered Specificity in DNA Binding by the Lac Repressor: A Mutant Lac Headpiece That Mimics the Gal Repressor. *Chembiochem Eur. J. Chem. Biol.* **2005**, *6* (9), 1628–1637. https://doi.org/10.1002/cbic.200500049.
(87) Shis, D. L.; Hussain, F.; Meinhardt, S.; Swint-Kruse, L.; Bennett, M. R. Modular, Multi-Input Transcriptional Logic Gating with Orthogonal LacI/GalR Family Chimeras. *ACS Synth. Biol.* **2014**, *3* (9), 645–651. https://doi.org/10.1021/sb500262f.

(88) Meyer, S.; Ramot, R.; Kishore Inampudi, K.; Luo, B.; Lin, C.; Amere, S.; Wilson, C. J. Engineering Alternate Cooperative-Communications in the Lactose Repressor Protein Scaffold. *Protein Eng. Des. Sel.* **2013**, *26* (6), 433–443. https://doi.org/10.1093/protein/gzt013.
(89) Swint-Kruse, L.; Elam, C. R.; Lin, J. W.; Wycuff, D. R.; Shive Matthews, K. Plasticity of Quaternary Structure: Twenty-Two Ways to Form a LacI Dimer. *Protein Sci. Publ. Protein Soc.* **2001**, *10* (2), 262–276. https://doi.org/10.1110/ps.35801.
(90) Jobe, A.; Bourgeois, S. *Lac* Repressor-Operator Interaction. *J. Mol. Biol.* **1972**, *69* (3), 397–408. https://doi.org/10.1016/0022-2836(72)90253-7.
(91) Barkley, M. D.; Riggs, A. D.; Jobe, A.; Bourgeois, S. Interaction of Effecting Ligands with Lac Repressor and Repressor-Operator Complex. *Biochemistry* **1975**, *14* (8), 1700–1712. https://doi.org/10.1021/bi00679a024.
(92) *MolView*. https://app.molview.com (accessed 2026-02-13).
(93) Wu, J.; Jiang, P.; Chen, W.; Xiong, D.; Huang, L.; Jia, J.; Chen, Y.; Jin, J.-M.; Tang, S.-Y. Design and Application of a Lactulose Biosensor. *Sci. Rep.* **2017**, *7* (1), 45994. https://doi.org/10.1038/srep45994.
(94) Taylor, N. D.; Garruss, A. S.; Moretti, R.; Chan, S.; Arbing, M. A.; Cascio, D.; Rogers, J. K.; Isaacs, F. J.; Kosuri, S.; Baker, D.; Fields, S.; Church, G. M.; Raman, S. Engineering an Allosteric Transcription Factor to Respond to New Ligands. *Nat. Methods* **2016**, *13* (2), 177–183. https://doi.org/10.1038/nmeth.3696.
(95) Bell, C. E.; Lewis, M. A Closer View of the Conformation of the Lac Repressor Bound to Operator. *Nat. Struct. Biol.* **2000**, *7* (3), 209–214. https://doi.org/10.1038/73317.
(96) Friedman, A. M.; Fischmann, T. O.; Steitz, T. A. Crystal Structure of Lac Repressor Core Tetramer and Its Implications for DNA Looping. *Science* **1995**, *268* (5218), 1721–1727. https://doi.org/10.1126/science.7792597.
(97) Xu, J.; Matthews, K. S. Flexibility in the Inducer Binding Region Is Crucial for Allostery in the Escherichia Coli Lactose Repressor. *Biochemistry* **2009**, *48* (22), 4988–4998. https://doi.org/10.1021/bi9002343.
(98) Zhan, H.; Sun, Z.; Matthews, K. S. Functional Impact of Polar and Acidic Substitutions in the Lactose Repressor Hydrophobic Monomer·Monomer Interface with a Buried Lysine. *Biochemistry* **2009**, *48* (6), 1305–1314. https://doi.org/10.1021/bi801357f.
(99) Swint-Kruse, L.; Zhan, H.; Matthews, K. S. Integrated Insights from Simulation, Experiment, and Mutational Analysis Yield New Details of LacI Function. *Biochemistry* **2005**, *44* (33), 11201–11213. https://doi.org/10.1021/bi050404+.
(100) Nichols, J. C.; Matthews, K. S. Combinatorial Mutations of *Lac* Repressor. *J. Biol. Chem.* **1997**, *272* (30), 18550–18557. https://doi.org/10.1074/jbc.272.30.18550.
(101) Chang, W. I.; Olson, J. S.; Matthews, K. S. Lysine 84 Is at the Subunit Interface of Lac Repressor Protein. *J. Biol. Chem.* **1993**, *268* (23), 17613–17622.
(102) Zhan, H.; Camargo, M.; Matthews, K. S. Positions 94−98 of the Lactose Repressor N-Subdomain Monomer−Monomer Interface Are Critical for Allosteric Communication. *Biochemistry* **2010**, *49* (39), 8636–8645. https://doi.org/10.1021/bi101106x.
(103) Richards, D. H.; Meyer, S.; Wilson, C. J. Fourteen Ways to Reroute Cooperative Communication in the Lactose Repressor: Engineering Regulatory Proteins with Alternate Repressive Functions. *ACS Synth. Biol.* **2017**, *6* (1), 6–12. https://doi.org/10.1021/acssynbio.6b00048.
(104) Groseclose, T. M.; Rondon, R. E.; Herde, Z. D.; Aldrete, C. A.; Wilson, C. J. Engineered Systems of Inducible Anti-Repressors for the next Generation of Biological Programming. *Nat. Commun.* **2020**, *11* (1), 4440. https://doi.org/10.1038/s41467-020-18302-1.
(105) Meyer, S.; Ramot, R.; Kishore Inampudi, K.; Luo, B.; Lin, C.; Amere, S.; Wilson, C. J. Engineering Alternate Cooperative-Communications in the Lactose Repressor Protein Scaffold. *Protein Eng. Des. Sel.* **2013**, *26* (6), 433–443. https://doi.org/10.1093/protein/gzt013.

(106) Swint-Kruse, L.; Elam, C. R.; Lin, J. W.; Wycuff, D. R.; Matthews, K. S. Plasticity of Quaternary Structure: Twenty-Two Ways to Form a LacI Dimer. *Protein Sci.* **2001**, *10* (2), 262–276. https://doi.org/10.1110/ps.35801.
(107) Spott, S. Die Dimerisierungsoberfläche des Lac Repressors aus Escherichia coli. Monographie, Universität Köln, Köln, 2022.
(108) Gerk, L. P.; Leven, O.; Müller-Hill, B. Strengthening the Dimerisation Interface of Lac Repressor Increases Its Thermostability by 40 Deg. C1. *J. Mol. Biol.* **2000**, *299* (3), 805–812. https://doi.org/10.1006/jmbi.2000.3706.
(109) Suckow, J.; Markiewicz, P.; Kleina, L. G.; Miller, J.; Kisters-Woike, B.; Müller-Hill, B. Genetic Studies of the Lac Repressor. XV: 4000 Single Amino Acid Substitutions and Analysis of the Resulting Phenotypes on the Basis of the Protein Structure. *J. Mol. Biol.* **1996**, *261* (4), 509–523. https://doi.org/10.1006/jmbi.1996.0479.
(110) Chakerian, A. E.; Matthews, K. S. Characterization of Mutations in Oligomerization Domain of Lac Repressor Protein. *J. Biol. Chem.* **1991**, *266* (33), 22206–22214. https://doi.org/10.1016/S0021-9258(18)54555-5.
(111) Daber, R.; Sharp, K.; Lewis, M. One Is Not Enough. *J. Mol. Biol.* **2009**, *392* (5), 1133–1144. https://doi.org/10.1016/j.jmb.2009.07.050.
(112) Klein, W.; Horlacher, R.; Boos, W. Molecular Analysis of treB Encoding the Escherichia Coli Enzyme II Specific for Trehalose. *J. Bacteriol.* **1995**, *177* (14), 4043–4052. https://doi.org/10.1128/jb.177.14.4043-4052.1995.
(113) Bürklen, L.; Schöck, F.; Dahl, M. K. Molecular Analysis of the Interaction between the Bacillus Subtilis Trehalose Repressor TreR and the Tre Operator. *Mol. Gen. Genet. MGG* **1998**, *260* (1), 48–55. https://doi.org/10.1007/s004380050869.
(114) Schöck, F.; Dahl, M. K. Expression of the Tre Operon of Bacillus Subtilis 168 Is Regulated by the Repressor TreR. *J. Bacteriol.* **1996**, *178* (15), 4576–4581.
(115) Hars, U.; Horlacher, R.; Boos, W.; Welte, W.; Diederichs, K. Crystal Structure of the Effector-Binding Domain of the Trehalose-Repressor of Escherichia Coli, a Member of the LacI Family, in Its Complexes with Inducer Trehalose-6-Phosphate and Noninducer Trehalose. *Protein Sci.* **1998**, *7* (12), 2511–2521. https://doi.org/10.1002/pro.5560071204.
(116) Tam, R.; Saier Jr, M. H. Structural, Functional, and Evolutionary Relationships among Extracellular Solute-Binding Receptors of Bacteria. *Microbiol. Rev.* **1993**, *57* (2), 320–346.
(117) Moussatova, A.; Kandt, C.; O'Mara, M. L.; Tieleman, D. P. ATP-Binding Cassette Transporters in *Escherichia Coli*. *Biochim. Biophys. Acta BBA - Biomembr.* **2008**, *1778* (9), 1757–1771. https://doi.org/10.1016/j.bbamem.2008.06.009.
(118) Oloo, E. O.; Kandt, C.; O'Mara, M. L.; Tieleman, D. P. Computer Simulations of ABC Transporter Components. *Biochem. Cell Biol. Biochim. Biol. Cell.* **2006**, *84* (6), 900–911. https://doi.org/10.1139/o06-182.
(119) Agarwal, S.; Dey, S.; Ghosh, B.; Biswas, M.; Dasgupta, J. Structure and Dynamics of Type III Periplasmic Proteins VcFhuD and VcHutB Reveal Molecular Basis of Their Distinctive Ligand Binding Properties. *Sci. Rep.* **2017**, *7* (1), 42812. https://doi.org/10.1038/srep42812.
(120) Felder, C. B.; Graul, R. C.; Lee, A. Y.; Merkle, H.-P.; Sadee, W. The Venus Flytrap of Periplasmic Binding Proteins: An Ancient Protein Module Present in Multiple Drug Receptors. *AAPS PharmSci* **1999**, *1* (2), 7–26. https://doi.org/10.1208/ps010202.
(121) Kröger, P.; Stiel, A. C.; Stüven, B.; Shanmugaratnam, S.; Schoch, S.; Wachten, D.; Höcker, B. A Fluorescent Biosensor for the Visualization of Agmatine. bioRxiv July 27, 2024, p 2024.07.27.605435. https://doi.org/10.1101/2024.07.27.605435.
(122) Kröger, P.; Shanmugaratnam, S.; Ferruz, N.; Schweimer, K.; Höcker, B. A Comprehensive Binding Study Illustrates Ligand Recognition in the Periplasmic Binding Protein PotF. *Structure* **2021**, *29* (5), 433–443.
(123) Vassylyev, D. G.; Tomitori, H.; Kashiwagi, K.; Morikawa, K.; Igarashi, K. Crystal Structure and Mutational Analysis of theEscherichia Coli Putrescine Receptor: STRUCTURAL

BASIS FOR SUBSTRATE SPECIFICITY *. *J. Biol. Chem.* **1998**, *273* (28), 17604–17609. https://doi.org/10.1074/jbc.273.28.17604.
(124) Gerstein, M.; Lesk, A. M.; Chothia, C. Structural Mechanisms for Domain Movements in Proteins. *Biochemistry* **1994**, *33* (22), 6739–6749. https://doi.org/10.1021/bi00188a001.
(125) Evenäs, J.; Malmendal, A.; Akke, M. Dynamics of the Transition between Open and Closed Conformations in a Calmodulin C-Terminal Domain Mutant. *Structure* **2001**, *9* (3), 185–195. https://doi.org/10.1016/S0969-2126(01)00575-5.
(126) Brostrom, C. O.; Wolff, D. J. Properties and Functions of Calmodulin. *Biochem. Pharmacol.* **1981**, *30* (12), 1395–1405. https://doi.org/10.1016/0006-2952(81)90358-0.
(127) Geller, D. H.; Henry, J. G.; Belch, J.; Schwartz, N. B. Co-Purification and Characterization of ATP-Sulfurylase and Adenosine-5'-Phosphosulfate Kinase from Rat Chondrosarcoma. *J. Biol. Chem.* **1987**, *262* (15), 7374–7382. https://doi.org/10.1016/S0021-9258(18)48247-6.
(128) MacRae, I. J.; Segel, I. H.; Fisher, A. J. Crystal Structure of ATP Sulfurylase from Penicillium Chrysogenum: Insights into the Allosteric Regulation of Sulfate Assimilation,. *Biochemistry* **2001**, *40* (23), 6795–6804. https://doi.org/10.1021/bi010367w.
(129) Bae, J.; Kim, J.; Choi, J.; Lee, H.; Koh, M. Split Proteins and Reassembly Modules for Biological Applications. *ChemBioChem* **2024**, *25* (10), e202400123. https://doi.org/10.1002/cbic.202400123.
(130) Blaszczak, E.; Lazarewicz, N.; Sudevan, A.; Wysocki, R.; Rabut, G. Protein-Fragment Complementation Assays for Large-Scale Analysis of Protein–Protein Interactions. *Biochem. Soc. Trans.* **2021**, *49* (3), 1337–1348. https://doi.org/10.1042/BST20201058.
(131) Shekhawat, S. S.; Ghosh, I. Split-Protein Systems: Beyond Binary Protein-Protein Interactions. *Curr. Opin. Chem. Biol.* **2011**, *15* (6), 789–797. https://doi.org/10.1016/j.cbpa.2011.10.014.
(132) Crosson, S.; Moffat, K. Structure of a Flavin-Binding Plant Photoreceptor Domain: Insights into Light-Mediated Signal Transduction. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (6), 2995–3000. https://doi.org/10.1073/pnas.051520298.
(133) Sheets, M. B.; Wong, W. W.; Dunlop, M. J. Light-Inducible Recombinases for Bacterial Optogenetics. *ACS Synth. Biol.* **2020**, *9* (2), 227–235. https://doi.org/10.1021/acssynbio.9b00395.
(134) Baumschlager, A.; Aoki, S. K.; Khammash, M. Dynamic Blue Light-Inducible T7 RNA Polymerases (Opto-T7RNAPs) for Precise Spatiotemporal Gene Expression Control. *ACS Synth. Biol.* **2017**, *6* (11), 2157–2167. https://doi.org/10.1021/acssynbio.7b00169.
(135) Seifert, S.; Ehrt, C.; Lückfeldt, L.; Lubeck, M.; Schramm, F.; Brakmann, S. Optical Control of Transcription: Genetically Encoded Photoswitchable Variants of T7 RNA Polymerase. *ChemBioChem* **2019**, *20* (22), 2813–2817. https://doi.org/10.1002/cbic.201900298.
(136) Wang, Y.; Barnett, S. F. H.; Le, S.; Guo, Z.; Zhong, X.; Kanchanawong, P.; Yan, J. Label-Free Single-Molecule Quantification of Rapamycin-Induced FKBP–FRB Dimerization for Direct Control of Cellular Mechanotransduction. *Nano Lett.* **2019**, *19* (10), 7514–7525. https://doi.org/10.1021/acs.nanolett.9b03364.
(137) Naik, R. R.; Kirkpatrick, S. M.; Stone, M. O. The Thermostability of an α-Helical Coiled-Coil Protein and Its Potential Use in Sensor Applications. *Biosens. Bioelectron.* **2001**, *16* (9), 1051–1057. https://doi.org/10.1016/S0956-5663(01)00226-3.
(138) Nihongaki, Y.; Otabe, T.; Ueda, Y.; Sato, M. A Split CRISPR–Cpf1 Platform for Inducible Genome Editing and Gene Activation. *Nat. Chem. Biol.* **2019**, *15* (9), 882–888. https://doi.org/10.1038/s41589-019-0338-y.
(139) Dirnberger, D.; Unsin, G.; Schlenker, S.; Reichel, C. A Small-Molecule-Protein Interaction System with Split-Ubiquitin as Sensor. *Chembiochem Eur. J. Chem. Biol.* **2006**, *7* (6), 936–942. https://doi.org/10.1002/cbic.200500544.

(140) Gray, D. C.; Mahrus, S.; Wells, J. A. Activation of Specific Apoptotic Caspases with an Engineered Small-Molecule-Activated Protease. *Cell* **2010**, *142* (4), 637–646. https://doi.org/10.1016/j.cell.2010.07.014.
(141) Snapp, E. L. Fluorescent Proteins: A Cell Biologist's User Guide. *Trends Cell Biol.* **2009**, *19* (11), 649–655. https://doi.org/10.1016/j.tcb.2009.08.002.
(142) Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W. W.; Prasher, D. C. Green Fluorescent Protein as a Marker for Gene Expression. *Science* **1994**, *263* (5148), 802–805. https://doi.org/10.1126/science.8303295.
(143) Ormö, M.; Cubitt, A. B.; Kallio, K.; Gross, L. A.; Tsien, R. Y.; Remington, S. J. Crystal Structure of the Aequorea Victoria Green Fluorescent Protein. *Science* **1996**, *273* (5280), 1392–1395. https://doi.org/10.1126/science.273.5280.1392.
(144) Tsien, R. Y. The Green Fluorescent Protein. *Annu. Rev. Biochem.* **1998**, *67*, 509–544. https://doi.org/10.1146/annurev.biochem.67.1.509.
(145) Shimomura, O.; Johnson, F. H.; Saiga, Y. Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan, Aequorea. *J. Cell. Comp. Physiol.* **1962**, *59*, 223–239. https://doi.org/10.1002/jcp.1030590302.
(146) Arpino, J. A. J.; Rizkallah, P. J.; Jones, D. D. Crystal Structure of Enhanced Green Fluorescent Protein to 1.35 Å Resolution Reveals Alternative Conformations for Glu222. *PLOS ONE* **2012**, *7* (10), e47132. https://doi.org/10.1371/journal.pone.0047132.
(147) Heim, R.; Cubitt, A. B.; Tsien, R. Y. Improved Green Fluorescence. *Nature* **1995**, *373* (6516), 663–664. https://doi.org/10.1038/373663b0.
(148) Shu, X.; Shaner, N. C.; Yarbrough, C. A.; Tsien, R. Y.; Remington, S. J. Novel Chromophores and Buried Charges Control Color in mFruits. *Biochemistry* **2006**, *45* (32), 9639–9647. https://doi.org/10.1021/bi060773l.
(149) Koraïchi, F.; Gence, R.; Bouchenot, C.; Grosjean, S.; Lajoie-Mazenc, I.; Favre, G.; Cabantous, S. High-Content Tripartite Split-GFP Cell-Based Assays to Screen for Modulators of Small GTPase Activation. *J. Cell Sci.* **2018**, *131* (1), jcs210419. https://doi.org/10.1242/jcs.210419.
(150) Groseclose, T. M.; Kober, E.; Zupancic, J.; McLelland, C. K.; Lujan, L. A.; Kunde, Y. A.; Mozden, S. C.; Velappan, N.; Lillo, A. M.; Solomon, E.; Yoder, J.; Waldo, G. S.; Nguyen, H. B. Engineering a New Tripartite Split-ccGFP System from Corynactis Californica for Detecting Protein–Protein Interactions. *Sci. Rep.* **2025**, *15* (1), 40798. https://doi.org/10.1038/s41598-025-24657-6.
(151) Cabantous, S.; Terwilliger, T. C.; Waldo, G. S. Protein Tagging and Detection with Engineered Self-Assembling Fragments of Green Fluorescent Protein. *Nat. Biotechnol.* **2005**, *23* (1), 102–107. https://doi.org/10.1038/nbt1044.
(152) Köker, T.; Fernandez, A.; Pinaud, F. Characterization of Split Fluorescent Protein Variants and Quantitative Analyses of Their Self-Assembly Process. *Sci. Rep.* **2018**, *8* (1), 5344. https://doi.org/10.1038/s41598-018-23625-7.
(153) Pereira, M.; Tomé, D.; Domingues, A. S.; Varanda, A. S.; Paulo, C.; Santos, M. A. S.; Soares, A. R. A Fluorescence-Based Sensor Assay That Monitors General Protein Aggregation in Human Cells. *Biotechnol. J.* **2018**, *13* (4), e1700676. https://doi.org/10.1002/biot.201700676.
(154) Cabantous, S.; Waldo, G. S. In Vivo and in Vitro Protein Solubility Assays Using Split GFP. *Nat. Methods* **2006**, *3* (10), 845–854. https://doi.org/10.1038/nmeth932.
(155) Chun, W.; Waldo, G. S.; Johnson, G. V. W. Split GFP Complementation Assay for Quantitative Measurement of Tau Aggregation in Situ. *Methods Mol. Biol.* **2011**, *670*, 109–123. https://doi.org/10.1007/978-1-60761-744-0_9.
(156) Kerppola, T. K. Design and Implementation of Bimolecular Fluorescence Complementation (BiFC) Assays for the Visualization of Protein Interactions in Living Cells. *Nat. Protoc.* **2006**, *1* (3), 1278–1286. https://doi.org/10.1038/nprot.2006.201.
(157) Kerppola, T. K. BIMOLECULAR FLUORESCENCE COMPLEMENTATION (BiFC) ANALYSIS AS A PROBE OF PROTEIN INTERACTIONS IN LIVING CELLS. *Annu.*

Rev. Biophys. **2008**, *37*, 465–487. https://doi.org/10.1146/annurev.biophys.37.032807.125842.
(158) Deng, Q.; Wang, D.; Xiang, X.; Gao, X.; Hardwidge, P. R.; Kaushik, R. S.; Wolff, T.; Chakravarty, S.; Li, F. Application of a Split Luciferase Complementation Assay for the Detection of Viral Protein-Protein Interactions. *J. Virol. Methods* **2011**, *176* (1–2), 108–111. https://doi.org/10.1016/j.jviromet.2011.04.028.
(159) Kamiyama, D.; Sekine, S.; Barsi-Rhyne, B.; Hu, J.; Chen, B.; Gilbert, L. A.; Ishikawa, H.; Leonetti, M. D.; Marshall, W. F.; Weissman, J. S.; Huang, B. Versatile Protein Tagging in Cells with Split Fluorescent Protein. *Nat. Commun.* **2016**, *7* (1), 11046. https://doi.org/10.1038/ncomms11046.
(160) Magliery, T. J.; Wilson, C. G. M.; Pan, W.; Mishler, D.; Ghosh, I.; Hamilton, A. D.; Regan, L. Detecting Protein-Protein Interactions with a Green Fluorescent Protein Fragment Reassembly Trap: Scope and Mechanism. *J. Am. Chem. Soc.* **2005**, *127* (1), 146–157. https://doi.org/10.1021/ja046699g.
(161) Ghosh, I.; Hamilton, A. D.; Regan, L. Antiparallel Leucine Zipper-Directed Protein Reassembly: Application to the Green Fluorescent Protein. *J. Am. Chem. Soc.* **2000**, *122* (23), 5658–5659. https://doi.org/10.1021/ja994421w.
(162) Wilson, C. G. M.; Magliery, T. J.; Regan, L. Detecting Protein-Protein Interactions with GFP-Fragment Reassembly. *Nat. Methods* **2004**, *1* (3), 255–262. https://doi.org/10.1038/nmeth1204-255.
(163) Hu, C.-D.; Chinenov, Y.; Kerppola, T. K. Visualization of Interactions among bZIP and Rel Family Proteins in Living Cells Using Bimolecular Fluorescence Complementation. *Mol. Cell* **2002**, *9* (4), 789–798. https://doi.org/10.1016/s1097-2765(02)00496-3.
(164) Nakagawa, C.; Inahata, K.; Nishimura, S.; Sugimoto, K. Improvement of a Venus-Based Bimolecular Fluorescence Complementation Assay to Visualize bFos-bJun Interaction in Living Cells. *Biosci. Biotechnol. Biochem.* **2011**, *75* (7), 1399–1401. https://doi.org/10.1271/bbb.110189.
(165) Kodama, Y.; Hu, C.-D. An Improved Bimolecular Fluorescence Complementation Assay with a High Signal-To-Noise Ratio. *BioTechniques* **2010**, *49* (5), 793–805. https://doi.org/10.2144/000113519.
(166) Feng, S.; Sekine, S.; Pessino, V.; Li, H.; Leonetti, M. D.; Huang, B. Improved Split Fluorescent Proteins for Endogenous Protein Labeling. *Nat. Commun.* **2017**, *8* (1), 370. https://doi.org/10.1038/s41467-017-00494-8.
(167) Luker, K. E.; Smith, M. C. P.; Luker, G. D.; Gammon, S. T.; Piwnica-Worms, H.; Piwnica-Worms, D. Kinetics of Regulated Protein–Protein Interactions Revealed with Firefly Luciferase Complementation Imaging in Cells and Living Animals. *Proc. Natl. Acad. Sci.* **2004**, *101* (33), 12288–12293. https://doi.org/10.1073/pnas.0404041101.
(168) Furman, J. L.; Mok, P.-W.; Shen, S.; Stains, C. I.; Ghosh, I. A Turn-on Split-Luciferase Sensor for the Direct Detection of Poly(ADP-Ribose) as a Marker for DNA Repair and Cell Death. *Chem. Commun.* **2010**, *47* (1), 397–399. https://doi.org/10.1039/C0CC02229B.
(169) Martell, J. D.; Yamagata, M.; Deerinck, T. J.; Phan, S.; Kwa, C. G.; Ellisman, M. H.; Sanes, J. R.; Ting, A. Y. A Split Horseradish Peroxidase for the Detection of Intercellular Protein–Protein Interactions and Sensitive Visualization of Synapses. *Nat. Biotechnol.* **2016**, *34* (7), 774–780. https://doi.org/10.1038/nbt.3563.
(170) Krainer, F. W.; Glieder, A. An Updated View on Horseradish Peroxidases: Recombinant Production and Biotechnological Applications. *Appl. Microbiol. Biotechnol.* **2015**, *99* (4), 1611–1625. https://doi.org/10.1007/s00253-014-6346-7.
(171) Galarneau, A.; Primeau, M.; Trudeau, L.-E.; Michnick, S. W. β-Lactamase Protein Fragment Complementation Assays as in Vivo and in Vitro Sensors of Protein–Protein Interactions. *Nat. Biotechnol.* **2002**, *20* (6), 619–622. https://doi.org/10.1038/nbt0602-619.
(172) Dolberg, T. B.; Meger, A. T.; Boucher, J. D.; Corcoran, W. K.; Schauer, E. E.; Prybutok, A. N.; Raman, S.; Leonard, J. N. Computation-Guided Optimization of Split Protein

Systems. *Nat. Chem. Biol.* **2021**, *17* (5), 531–539. https://doi.org/10.1038/s41589-020-00729-8.
(173) De Las Rivas, J.; Fontanillo, C. Protein-Protein Interactions Essentials: Key Concepts to Building and Analyzing Interactome Networks. *PLoS Comput. Biol.* **2010**, *6* (6), e1000807. https://doi.org/10.1371/journal.pcbi.1000807.
(174) Vidal, M.; Cusick, M. E.; Barabási, A.-L. Interactome Networks and Human Disease. *Cell* **2011**, *144* (6), 986–998. https://doi.org/10.1016/j.cell.2011.02.016.
(175) Barabási, A.-L.; Gulbahce, N.; Loscalzo, J. Network Medicine: A Network-Based Approach to Human Disease. *Nat. Rev. Genet.* **2011**, *12* (1), 56–68. https://doi.org/10.1038/nrg2918.
(176) Guarracino, D. A.; Bullock, B. N.; Arora, P. S. Protein-Protein Interactions in Transcription: A Fertile Ground for Helix Mimetics. *Biopolymers* **2011**, *95* (1), 1–7. https://doi.org/10.1002/bip.21546.
(177) Moldovan, G.-L.; Pfander, B.; Jentsch, S. PCNA, the Maestro of the Replication Fork. *Cell* **2007**, *129* (4), 665–679. https://doi.org/10.1016/j.cell.2007.05.003.
(178) Lee, M. J.; Yaffe, M. B. Protein Regulation in Signal Transduction. *Cold Spring Harb. Perspect. Biol.* **2016**, *8* (6), a005918. https://doi.org/10.1101/cshperspect.a005918.
(179) Sun, A.-Q.; Balasubramaniyan, N.; Xu, K.; Liu, C. J.; Ponamgi, V. M.; Liu, H.; Suchy, F. J. Protein-Protein Interactions and Membrane Localization of the Human Organic Solute Transporter. *Am. J. Physiol. Gastrointest. Liver Physiol.* **2007**, *292* (6), G1586-1593. https://doi.org/10.1152/ajpgi.00457.2006.
(180) Jiang, Y.; Thienpont, B.; Sapuru, V.; Hite, R. K.; Dittman, J. S.; Sturgis, J. N.; Scheuring, S. Membrane-Mediated Protein Interactions Drive Membrane Protein Organization. *Nat. Commun.* **2022**, *13* (1), 7373. https://doi.org/10.1038/s41467-022-35202-8.
(181) Westermarck, J.; Ivaska, J.; Corthals, G. L. Identification of Protein Interactions Involved in Cellular Signaling. *Mol. Cell. Proteomics MCP* **2013**, *12* (7), 1752–1763. https://doi.org/10.1074/mcp.R113.027771.
(182) Kashtan, N.; Alon, U. Spontaneous Evolution of Modularity and Network Motifs. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (39), 13773–13778. https://doi.org/10.1073/pnas.0503610102.
(183) Fan, R.; Aranko, A. S. Catcher/Tag Toolbox: Biomolecular Click-Reactions For Protein Engineering Beyond Genetics. *ChemBioChem* **2024**, *25* (1), e202300600. https://doi.org/10.1002/cbic.202300600.
(184) Agouridas, V.; El Mahdi, O.; Diemer, V.; Cargoët, M.; Monbaliu, J.-C. M.; Melnyk, O. Native Chemical Ligation and Extended Methods: Mechanisms, Catalysis, Scope, and Limitations. *Chem. Rev.* **2019**, *119* (12), 7328–7443. https://doi.org/10.1021/acs.chemrev.8b00712.
(185) Pihl, R.; Zheng, Q.; David, Y. Nature-Inspired Protein Ligation and Its Applications. *Nat. Rev. Chem.* **2023**, *7* (4), 234–255. https://doi.org/10.1038/s41570-023-00468-z.
(186) Morgan, H. E.; Turnbull, W. B.; Webb, M. E. Challenges in the Use of Sortase and Other Peptide Ligases for Site-Specific Protein Modification. *Chem. Soc. Rev.* **2022**, *51* (10), 4121–4145. https://doi.org/10.1039/D0CS01148G.
(187) Freund, C.; Schwarzer, D. Engineered Sortases in Peptide and Protein Chemistry. *Chembiochem Eur. J. Chem. Biol.* **2021**, *22* (8), 1347–1356. https://doi.org/10.1002/cbic.202000745.
(188) Schneider, H.; Deweid, L.; Avrutina, O.; Kolmar, H. Recent Progress in Transglutaminase-Mediated Assembly of Antibody-Drug Conjugates. *Anal. Biochem.* **2020**, *595*, 113615. https://doi.org/10.1016/j.ab.2020.113615.
(189) Aranko, A. S.; Wlodawer, A.; Iwaï, H. Nature's Recipe for Splitting Inteins. *Protein Eng. Des. Sel.* **2014**, *27* (8), 263–271. https://doi.org/10.1093/protein/gzu028.
(190) Milczek, E. M. Commercial Applications for Enzyme-Mediated Protein Conjugation: New Developments in Enzymatic Processes to Deliver Functionalized Proteins on the Commercial Scale. *Chem. Rev.* **2018**, *118* (1), 119–141. https://doi.org/10.1021/acs.chemrev.6b00832.

(191) Züger, S.; Iwai, H. Intein-Based Biosynthetic Incorporation of Unlabeled Protein Tags into Isotopically Labeled Proteins for NMR Studies. *Nat. Biotechnol.* **2005**, *23* (6), 736–740. https://doi.org/10.1038/nbt1097.
(192) Zakeri, B.; Fierer, J. O.; Celik, E.; Chittock, E. C.; Schwarz-Linek, U.; Moy, V. T.; Howarth, M. Peptide Tag Forming a Rapid Covalent Bond to a Protein, through Engineering a Bacterial Adhesin. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (12), E690-697. https://doi.org/10.1073/pnas.1115485109.
(193) Shaik, M. M.; Maccagni, A.; Tourcier, G.; Guilmi, A. M. D.; Dessen, A. Structural Basis of Pilus Anchoring by the Ancillary Pilin RrgC of Streptococcus Pneumoniae*. *J. Biol. Chem.* **2014**, *289* (24), 16988–16997. https://doi.org/10.1074/jbc.M114.555854.
(194) Li, L.; Fierer, J. O.; Rapoport, T. A.; Howarth, M. Structural Analysis and Optimization of the Covalent Association between SpyCatcher and a Peptide Tag. *J. Mol. Biol.* **2014**, *426* (2), 309–317. https://doi.org/10.1016/j.jmb.2013.10.021.
(195) Izoré, T.; Contreras-Martel, C.; Mortaji, L. E.; Manzano, C.; Terrasse, R.; Vernet, T.; Guilmi, A. M. D.; Dessen, A. Structural Basis of Host Cell Recognition by the Pilus Adhesin from Streptococcus Pneumoniae. *Structure* **2010**, *18* (1), 106–115. https://doi.org/10.1016/j.str.2009.10.019.
(196) Fan, R.; Hakanpää, J.; Elfving, K.; Taberman, H.; Linder, M. B.; Aranko, A. S. Biomolecular Click Reactions Using a Minimal pH-Activated Catcher/Tag Pair for Producing Native-Sized Spider-Silk Proteins. *Angew. Chem. Int. Ed.* **2023**, *62* (11), e202216371. https://doi.org/10.1002/anie.202216371.
(197) Keeble, A. H.; Yadav, V. K.; Ferla, M. P.; Bauer, C. C.; Chuntharpursat-Bon, E.; Huang, J.; Bon, R. S.; Howarth, M. DogCatcher Allows Loop-Friendly Protein-Protein Ligation. *Cell Chem. Biol.* **2022**, *29* (2), 339-350.e10. https://doi.org/10.1016/j.chembiol.2021.07.005.
(198) Veggiani, G.; Nakamura, T.; Brenner, M. D.; Gayet, R. V.; Yan, J.; Robinson, C. V.; Howarth, M. Programmable Polyproteams Built Using Twin Peptide Superglues. *Proc. Natl. Acad. Sci.* **2016**, *113* (5), 1202–1207. https://doi.org/10.1073/pnas.1519214113.
(199) Buldun, C. M.; Jean, J. X.; Bedford, M. R.; Howarth, M. SnoopLigase Catalyzes Peptide–Peptide Locking and Enables Solid-Phase Conjugate Isolation. *J. Am. Chem. Soc.* **2018**, *140* (8), 3008–3018. https://doi.org/10.1021/jacs.7b13237.
(200) Wu, X.-L.; Liu, Y.; Liu, D.; Sun, F.; Zhang, W.-B. An Intrinsically Disordered Peptide-Peptide Stapler for Highly Efficient Protein Ligation Both in Vivo and in Vitro. *J. Am. Chem. Soc.* **2018**, *140* (50), 17474–17483. https://doi.org/10.1021/jacs.8b08250.
(201) Hagan, R. M.; Björnsson, R.; McMahon, S. A.; Schomburg, B.; Braithwaite, V.; Bühl, M.; Naismith, J. H.; Schwarz-Linek, U. NMR Spectroscopic and Theoretical Analysis of a Spontaneously Formed Lys-Asp Isopeptide Bond. *Angew. Chem.* **2010**, *49* (45), 8421–8425. https://doi.org/10.1002/anie.201004340.
(202) Larson, M. R.; Rajashankar, K. R.; Crowley, P. J.; Kelly, C.; Mitchell, T. J.; Brady, L. J.; Deivanayagam, C. Crystal Structure of the C-Terminal Region of Streptococcus Mutans Antigen I/II and Characterization of Salivary Agglutinin Adherence Domains♦. *J. Biol. Chem.* **2011**, *286* (24), 21657–21666. https://doi.org/10.1074/jbc.M111.231100.
(203) Hu, X.; Hu, H.; Melvin, J. A.; Clancy, K. W.; McCafferty, D. G.; Yang, W. Autocatalytic Intramolecular Isopeptide Bond Formation in Gram-Positive Bacterial Pili: A QM/MM Simulation. *J. Am. Chem. Soc.* **2011**, *133* (3), 478–485. https://doi.org/10.1021/ja107513t.
(204) Reddington, S. C.; Howarth, M. Secrets of a Covalent Interaction for Biomaterials and Biotechnology: SpyTag and SpyCatcher. *Curr. Opin. Chem. Biol.* **2015**, *29*, 94–99. https://doi.org/10.1016/j.cbpa.2015.10.002.
(205) Peschke, T.; Bitterwolf, P.; Gallus, S.; Hu, Y.; Oelschlaeger, C.; Willenbacher, N.; Rabe, K. S.; Niemeyer, C. M. Self-Assembling All-Enzyme Hydrogels for Flow Biocatalysis. *Angew. Chem.* **2018**, *57* (52), 17028–17032. https://doi.org/10.1002/anie.201810331.
(206) Sun, F.; Zhang, W.-B.; Mahdavi, A.; Arnold, F. H.; Tirrell, D. A. Synthesis of Bioactive Protein Hydrogels by Genetically Encoded SpyTag-SpyCatcher Chemistry. *Proc. Natl. Acad. Sci.* **2014**, *111* (31), 11269–11274. https://doi.org/10.1073/pnas.1401291111.

(207) Anderson, G. P.; Liu, J. L.; Shriver-Lake, L. C.; Zabetakis, D.; Sugiharto, V. A.; Chen, H.-W.; Lee, C.-R.; Defang, G. N.; Wu, S.-J. L.; Venkateswaran, N.; Goldman, E. R. Oriented Immobilization of Single-Domain Antibodies Using SpyTag/SpyCatcher Yields Improved Limits of Detection. *Anal. Chem.* **2019**, *91* (15), 9424–9429. https://doi.org/10.1021/acs.analchem.9b02096.

(208) Ke, X.; Zhang, Y.; Zheng, F.; Liu, Y.; Zheng, Z.; Xu, Y.; Wang, H. SpyCatcher–SpyTag Mediated in Situ Labelling of Progeny Baculovirus with Quantum Dots for Tracking Viral Infection in Living Cells. *Chem. Commun.* **2018**, *54* (10), 1189–1192. https://doi.org/10.1039/C7CC08880A.

(209) Khairil Anuar, I. N. A.; Banerjee, A.; Keeble, A. H.; Carella, A.; Nikov, G. I.; Howarth, M. Spy&Go Purification of SpyTag-Proteins Using Pseudo-SpyCatcher to Access an Oligomerization Toolbox. *Nat. Commun.* **2019**, *10* (1), 1734. https://doi.org/10.1038/s41467-019-09678-w.

(210) Bedbrook, C. N.; Kato, M.; Ravindra Kumar, S.; Lakshmanan, A.; Nath, R. D.; Sun, F.; Sternberg, P. W.; Arnold, F. H.; Gradinaru, V. Genetically Encoded Spy Peptide Fusion System to Detect Plasma Membrane-Localized Proteins In Vivo. *Chem. Biol.* **2015**, *22* (8), 1108–1121. https://doi.org/10.1016/j.chembiol.2015.06.020.

(211)Fairhead, M.; Veggiani, G.; Lever, M.; Yan, J.; Mesner, D.; Robinson, C. V.; Dushek, O.; van der Merwe, P. A.; Howarth, M. SpyAvidin Hubs Enable Precise and Ultrastable Orthogonal Nanoassembly. *J. Am. Chem. Soc.* **2014**, *136* (35), 12355–12363. https://doi.org/10.1021/ja505584f.

(212) Nguyen, P. Q.; Botyanszki, Z.; Tay, P. K. R.; Joshi, N. S. Programmable Biofilm-Based Materials from Engineered Curli Nanofibres. *Nat. Commun.* **2014**, *5*, 4945. https://doi.org/10.1038/ncomms5945.

(213) Chen, A. Y.; Deng, Z.; Billings, A. N.; Seker, U. O. S.; Lu, M. Y.; Citorik, R. J.; Zakeri, B.; Lu, T. K. Synthesis and Patterning of Tunable Multiscale Materials with Engineered Cells. *Nat. Mater.* **2014**, *13* (5), 515–523. https://doi.org/10.1038/nmat3912.

(214) Nilsson, P.; Robertson, N. O.; Gustafsson, N.; Davies, R. B.; Liew, C. W.; Lyons, A.; Eichhorn, R.; Niman, C. S.; Blab, G. A.; Bromley, E. H. C.; Whitten, A. E.; Duff, A. P.; Beech, J. P.; Jönsson, P.; Böcking, T.; Höcker, B.; Woolfson, D. N.; Forde, N. R.; Linke, H.; Curmi, P. M. G. Tumbleweed: An Artificial Motor Protein That Walks along a DNA Track. bioRxiv March 12, 2025, p 2025.03.07.641123. https://doi.org/10.1101/2025.03.07.641123.

(215) Korendovych, I. V.; DeGrado, W. F. De Novo Protein Design, a Retrospective. *Q. Rev. Biophys.* **2020**, *53*, e3. https://doi.org/10.1017/S0033583519000131.

(216) Woolfson, D. N. A Brief History of *De Novo* Protein Design: Minimal, Rational, and Computational. *J. Mol. Biol.* **2021**, *433* (20), 167160. https://doi.org/10.1016/j.jmb.2021.167160.

(217) Abriata, L. A. The Nobel Prize in Chemistry: Past, Present, and Future of AI in Biology. *Commun. Biol.* **2024**, *7* (1), 1409. https://doi.org/10.1038/s42003-024-07113-5.

(218) Meiler, J.; Schoeder, C. Nobelpreis für Chemie 2024: Proteine falten und entwerfen. *Nachrichten Aus Chem.* **2024**, *72* (11), 8–12. https://doi.org/10.1002/nadc.20244146113.

(219) Lutz, S.; Iamurri, S. M. Protein Engineering: Past, Present, and Future. *Methods Mol. Biol.* **2018**, *1685*, 1–12. https://doi.org/10.1007/978-1-4939-7366-8_1.

(220) Regan, L.; DeGrado, W. F. Characterization of a Helical Protein Designed from First Principles. *Science* **1988**, *241* (4868), 976–978. https://doi.org/10.1126/science.3043666.

(221) Goldenzweig, A.; Fleishman, S. J. Principles of Protein Stability and Their Application in Computational Design. *Annu. Rev. Biochem.* **2018**, *87*, 105–129. https://doi.org/10.1146/annurev-biochem-062917-012102.

(222) Levinthal, C. How to Fold Graciously. *Mossbauer Spectrosc. Biol. Syst.* **1969**, *67*, 22–24.

(223) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Funnels, Pathways, and the Energy Landscape of Protein Folding: A Synthesis. *Proteins Struct. Funct. Bioinforma.* **1995**, *21* (3), 167–195. https://doi.org/10.1002/prot.340210302.

(224) Gutte, B.; Däumigen, M.; Wittschieber, E. Design, Synthesis and Characterisation of a 34-Residue Polypeptide That Interacts with Nucleic Acids. *Nature* **1979**, *281* (5733), 650–655. https://doi.org/10.1038/281650a0.
(225) Moser, R.; Thomas, R. M.; Gutte, B. An Artificial Crystalline DDT-Binding Polypeptide. *FEBS Lett.* **1983**, *157* (2), 247–251. https://doi.org/10.1016/0014-5793(83)80555-9.
(226) Pasternak, A.; Kaplan, J.; Lear, J. D.; Degrado, W. F. Proton and Metal Ion-Dependent Assembly of a Model Diiron Protein. *Protein Sci.* **2001**, *10* (5), 958–969. https://doi.org/10.1110/ps.52101.
(227) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4* (3), 435–447. https://doi.org/10.1021/ct700301q.
(228) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* **2006**, *65* (3), 712–725. https://doi.org/10.1002/prot.21123.
(229) Leaver-Fay, A.; Tyka, M.; Lewis, S. M.; Lange, O. F.; Thompson, J.; Jacak, R.; Kaufman, K.; Renfrew, P. D.; Smith, C. A.; Sheffler, W.; Davis, I. W.; Cooper, S.; Treuille, A.; Mandell, D. J.; Richter, F.; Ban, Y.-E. A.; Fleishman, S. J.; Corn, J. E.; Kim, D. E.; Lyskov, S.; Berrondo, M.; Mentzer, S.; Popović, Z.; Havranek, J. J.; Karanicolas, J.; Das, R.; Meiler, J.; Kortemme, T.; Gray, J. J.; Kuhlman, B.; Baker, D.; Bradley, P. ROSETTA3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules. *Methods Enzymol.* **2011**, *487*, 545–574. https://doi.org/10.1016/B978-0-12-381270-4.00019-6.
(230) Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science* **2003**, *302* (5649), 1364–1368. https://doi.org/10.1126/science.1089427.
(231) Rocklin, G. J.; Chidyausiku, T. M.; Goreshnik, I.; Ford, A.; Houliston, S.; Lemak, A.; Carter, L.; Ravichandran, R.; Mulligan, V. K.; Chevalier, A.; Arrowsmith, C. H.; Baker, D. Global Analysis of Protein Folding Using Massively Parallel Design, Synthesis, and Testing. *Science* **2017**, *357* (6347), 168–175. https://doi.org/10.1126/science.aan0693.
(232) Bednar, D.; Beerens, K.; Sebestova, E.; Bendl, J.; Khare, S.; Chaloupkova, R.; Prokop, Z.; Brezovsky, J.; Baker, D.; Damborsky, J. FireProt: Energy- and Evolution-Based Computational Design of Thermostable Multiple-Point Mutants. *PLOS Comput. Biol.* **2015**, *11* (11), e1004556. https://doi.org/10.1371/journal.pcbi.1004556.
(233) Huang, P.-S.; Feldmeier, K.; Parmeggiani, F.; Fernandez Velasco, D. A.; Höcker, B.; Baker, D. De Novo Design of a Four-Fold Symmetric TIM-Barrel Protein with Atomic-Level Accuracy. *Nat. Chem. Biol.* **2016**, *12* (1), 29–34. https://doi.org/10.1038/nchembio.1966.
(234) King, N. P.; Sheffler, W.; Sawaya, M. R.; Vollmar, B. S.; Sumida, J. P.; André, I.; Gonen, T.; Yeates, T. O.; Baker, D. Computational Design of Self-Assembling Protein Nanomaterials with Atomic Level Accuracy. *Science* **2012**, *336* (6085), 1171–1174. https://doi.org/10.1126/science.1219364.
(235) Lin, Y.-R.; Koga, N.; Tatsumi-Koga, R.; Liu, G.; Clouser, A. F.; Montelione, G. T.; Baker, D. Control over Overall Shape and Size in de Novo Designed Proteins. *Proc. Natl. Acad. Sci.* **2015**, *112* (40), E5478–E5485. https://doi.org/10.1073/pnas.1509508112.
(236) Marcos, E.; Chidyausiku, T. M.; McShan, A. C.; Evangelidis, T.; Nerli, S.; Carter, L.; Nivón, L. G.; Davis, A.; Oberdorfer, G.; Tripsianes, K.; Sgourakis, N. G.; Baker, D. De Novo Design of a Non-Local β-Sheet Protein with High Stability and Accuracy. *Nat. Struct. Mol. Biol.* **2018**, *25* (11), 1028–1034. https://doi.org/10.1038/s41594-018-0141-6.
(237) Röthlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D. Kemp Elimination Catalysts by Computational Enzyme Design. *Nature* **2008**, *453* (7192), 190–195. https://doi.org/10.1038/nature06879.
(238) Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; St Clair, J. L.; Gallaher, J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L.; Houk, K. N.; Michael, F. E.; Baker, D. Computational Design of an Enzyme Catalyst for a Stereoselective

Bimolecular Diels-Alder Reaction. *Science* **2010**, *329* (5989), 309–313. https://doi.org/10.1126/science.1190239.
(239) Chevalier, A.; Silva, D.-A.; Rocklin, G. J.; Hicks, D. R.; Vergara, R.; Murapa, P.; Bernard, S. M.; Zhang, L.; Lam, K.-H.; Yao, G.; Bahl, C. D.; Miyashita, S.-I.; Goreshnik, I.; Fuller, J. T.; Koday, M. T.; Jenkins, C. M.; Colvin, T.; Carter, L.; Bohn, A.; Bryan, C. M.; Fernández-Velasco, D. A.; Stewart, L.; Dong, M.; Huang, X.; Jin, R.; Wilson, I. A.; Fuller, D. H.; Baker, D. Massively Parallel de Novo Protein Design for Targeted Therapeutics. *Nature* **2017**, *550* (7674), 74–79. https://doi.org/10.1038/nature23912.
(240) Tinberg, C. E.; Khare, S. D.; Dou, J.; Doyle, L.; Nelson, J. W.; Schena, A.; Jankowski, W.; Kalodimos, C. G.; Johnsson, K.; Stoddard, B. L.; Baker, D. Computational Design of Ligand-Binding Proteins with High Affinity and Selectivity. *Nature* **2013**, *501* (7466), 212–216. https://doi.org/10.1038/nature12443.
(241) Hallen, M. A.; Martin, J. W.; Ojewole, A.; Jou, J. D.; Lowegard, A. U.; Frenkel, M. S.; Gainza, P.; Nisonoff, H. M.; Mukund, A.; Wang, S. OSPREY 3.0: Open-source Protein Redesign for You, with Powerful New Features. *J. Comput. Chem.* **2018**, *39* (30), 2494–2507.
(242) Chen, C.-Y.; Georgiev, I.; Anderson, A. C.; Donald, B. R. Computational Structure-Based Redesign of Enzyme Activity. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (10), 3764–3769. https://doi.org/10.1073/pnas.0900266106.
(243) Reeve, S. M.; Gainza, P.; Frey, K. M.; Georgiev, I.; Donald, B. R.; Anderson, A. C. Protein Design Algorithms Predict Viable Resistance to an Experimental Antifolate. *Proc. Natl. Acad. Sci.* **2015**, *112* (3), 749–754.
(244) Wang, S.; Reeve, S. M.; Holt, G. T.; Ojewole, A. A.; Frenkel, M. S.; Gainza, P.; Keshipeddy, S.; Fowler, V. G.; Wright, D. L.; Donald, B. R. Chiral Evasion and Stereospecific Antifolate Resistance in Staphylococcus Aureus. *PLOS Comput. Biol.* **2022**, *18* (2), e1009855. https://doi.org/10.1371/journal.pcbi.1009855.
(245) Broom, A.; Rakotoharisoa, R. V.; Thompson, M. C.; Zarifi, N.; Nguyen, E.; Mukhametzhanov, N.; Liu, L.; Fraser, J. S.; Chica, R. A. Ensemble-Based Enzyme Design Can Recapitulate the Effects of Laboratory Directed Evolution in Silico. *Nat. Commun.* **2020**, *11* (1), 4808.
(246) Legault, S.; P. Fraser-Halberg, D.; L. McAnelly, R.; G. Eason, M.; C. Thompson, M.; A. Chica, R. Generation of Bright Monomeric Red Fluorescent Proteins via Computational Design of Enhanced Chromophore Packing. *Chem. Sci.* **2022**, *13* (5), 1408–1418. https://doi.org/10.1039/D1SC05088E.
(247) Beck, J.; Smith, B. J.; Zarifi, N.; Freund, E.; Chica, R. A.; Höcker, B. Customizing the Structure of a Minimal TIM Barrel to Craft a De Novo Enzyme. bioRxiv January 29, 2025, p 2025.01.28.635154. https://doi.org/10.1101/2025.01.28.635154.
(248) Kortemme, T. *De Novo* Protein Design—From New Structures to Programmable Functions. *Cell* **2024**, *187* (3), 526–544. https://doi.org/10.1016/j.cell.2023.12.028.
(249) Ferruz, N.; Heinzinger, M.; Akdel, M.; Goncearenco, A.; Naef, L.; Dallago, C. From Sequence to Function through Structure: Deep Learning for Protein Design. *Comput. Struct. Biotechnol. J.* **2023**, *21*, 238–250. https://doi.org/10.1016/j.csbj.2022.11.014.
(250) Ofer, D.; Brandes, N.; Linial, M. The Language of Proteins: NLP, Machine Learning & Protein Sequences. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 1750–1758. https://doi.org/10.1016/j.csbj.2021.03.022.
(251) Ferruz, N.; Höcker, B. Controllable Protein Design with Language Models. *Nat. Mach. Intell.* **2022**, *4* (6), 521–532. https://doi.org/10.1038/s42256-022-00499-z.
(252) Dauparas, J.; Anishchenko, I.; Bennett, N.; Bai, H.; Ragotte, R. J.; Milles, L. F.; Wicky, B. I. M.; Courbet, A.; de Haas, R. J.; Bethel, N.; Leung, P. J. Y.; Huddy, T. F.; Pellock, S.; Tischer, D.; Chan, F.; Koepnick, B.; Nguyen, H.; Kang, A.; Sankaran, B.; Bera, A. K.; King, N. P.; Baker, D. Robust Deep Learning–Based Protein Sequence Design Using ProteinMPNN. *Science* **2022**, *378* (6615), 49–56. https://doi.org/10.1126/science.add2187.
(253) Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G. R.; Wang, J.; Cong, Q.; Kinch, L. N.; Schaeffer, R. D.; Millán, C.; Park, H.; Adams, C.; Glassman,

C. R.; DeGiovanni, A.; Pereira, J. H.; Rodrigues, A. V.; van Dijk, A. A.; Ebrecht, A. C.; Opperman, D. J.; Sagmeister, T.; Buhlheller, C.; Pavkov-Keller, T.; Rathinaswamy, M. K.; Dalwadi, U.; Yip, C. K.; Burke, J. E.; Garcia, K. C.; Grishin, N. V.; Adams, P. D.; Read, R. J.; Baker, D. Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network. *Science* **2021**, *373* (6557), 871–876. https://doi.org/10.1126/science.abj8754.

(254) Anishchenko, I.; Pellock, S. J.; Chidyausiku, T. M.; Ramelot, T. A.; Ovchinnikov, S.; Hao, J.; Bafna, K.; Norn, C.; Kang, A.; Bera, A. K.; DiMaio, F.; Carter, L.; Chow, C. M.; Montelione, G. T.; Baker, D. De Novo Protein Design by Deep Network Hallucination. *Nature* **2021**, *600* (7889), 547–552. https://doi.org/10.1038/s41586-021-04184-w.

(255) Ferruz, N.; Schmidt, S.; Höcker, B. ProtGPT2 Is a Deep Unsupervised Language Model for Protein Design. *Nat. Commun.* **2022**, *13* (1), 4348. https://doi.org/10.1038/s41467-022-32007-7.

(256) Wang, J.; Lisanza, S.; Juergens, D.; Tischer, D.; Watson, J. L.; Castro, K. M.; Ragotte, R.; Saragovi, A.; Milles, L. F.; Baek, M.; Anishchenko, I.; Yang, W.; Hicks, D. R.; Expòsit, M.; Schlichthaerle, T.; Chun, J.-H.; Dauparas, J.; Bennett, N.; Wicky, B. I. M.; Muenks, A.; DiMaio, F.; Correia, B.; Ovchinnikov, S.; Baker, D. Scaffolding Protein Functional Sites Using Deep Learning. *Science* **2022**, *377* (6604), 387–394. https://doi.org/10.1126/science.abn2100.

(257) Watson, J. L.; Juergens, D.; Bennett, N. R.; Trippe, B. L.; Yim, J.; Eisenach, H. E.; Ahern, W.; Borst, A. J.; Ragotte, R. J.; Milles, L. F.; Wicky, B. I. M.; Hanikel, N.; Pellock, S. J.; Courbet, A.; Sheffler, W.; Wang, J.; Venkatesh, P.; Sappington, I.; Torres, S. V.; Lauko, A.; De Bortoli, V.; Mathieu, E.; Ovchinnikov, S.; Barzilay, R.; Jaakkola, T. S.; DiMaio, F.; Baek, M.; Baker, D. De Novo Design of Protein Structure and Function with RFdiffusion. *Nature* **2023**, *620* (7976), 1089–1100. https://doi.org/10.1038/s41586-023-06415-8.

(258) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.

(259) Chai Discovery; Boitreaud, J.; Dent, J.; McPartlon, M.; Meier, J.; Reis, V.; Rogozhnikov, A.; Wu, K. Chai-1: Decoding the Molecular Interactions of Life. Synthetic Biology October 11, 2024. https://doi.org/10.1101/2024.10.10.615955.

(260) Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; Bodenstein, S. W.; Evans, D. A.; Hung, C.-C.; O'Neill, M.; Reiman, D.; Tunyasuvunakool, K.; Wu, Z.; Žemgulytė, A.; Arvaniti, E.; Beattie, C.; Bertolli, O.; Bridgland, A.; Cherepanov, A.; Congreve, M.; Cowen-Rivers, A. I.; Cowie, A.; Figurnov, M.; Fuchs, F. B.; Gladman, H.; Jain, R.; Khan, Y. A.; Low, C. M. R.; Perlin, K.; Potapenko, A.; Savy, P.; Singh, S.; Stecula, A.; Thillaisundaram, A.; Tong, C.; Yakneen, S.; Zhong, E. D.; Zielinski, M.; Žídek, A.; Bapst, V.; Kohli, P.; Jaderberg, M.; Hassabis, D.; Jumper, J. M. Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3. *Nature* **2024**, *630* (8016), 493–500. https://doi.org/10.1038/s41586-024-07487-w.

(261) Beck, J.; Shanmugaratnam, S.; Höcker, B. Diversifying de Novo TIM Barrels by Hallucination. *Protein Sci. Publ. Protein Soc.* **2024**, *33* (6), e5001. https://doi.org/10.1002/pro.5001.

(262) Kim, D. E.; Watson, J. L.; Juergens, D.; Majumder, S.; Sonigra, R.; Gerben, S. R.; Kang, A.; Bera, A. K.; Li, X.; Baker, D. Parametrically Guided Design of Beta Barrels and Transmembrane Nanopores Using Deep Learning. *bioRxiv* **2025**, 2024.07.22.604663. https://doi.org/10.1101/2024.07.22.604663.

(263) Torres, S. V.; Leung, P. J. Y.; Lutz, I. D.; Venkatesh, P.; Watson, J. L.; Hink, F.; Huynh, H.-H.; Yeh, A. H.-W.; Bennett, N. R.; Hoofnagle, A. N.; Huang, E.; Expòsit, M.; Lee, G.

R.; Korkmaz, E. N.; Nivala, J.; Rogers, J. M.; Baker, D. De Novo Design of High-Affinity Protein Binders to Bioactive Helical Peptides.
(264) Bennett, N. R.; Watson, J. L.; Ragotte, R. J.; Borst, A. J.; See, D. L.; Weidle, C.; Biswas, R.; Yu, Y.; Shrock, E. L.; Ault, R.; Leung, P. J. Y.; Huang, B.; Goreshnik, I.; Tam, J.; Carr, K. D.; Singer, B.; Criswell, C.; Wicky, B. I. M.; Vafeados, D.; Garcia Sanchez, M.; Kim, H. M.; Vázquez Torres, S.; Chan, S.; Sun, S. M.; Spear, T. T.; Sun, Y.; O'Reilly, K.; Maris, J. M.; Sgourakis, N. G.; Melnyk, R. A.; Liu, C. C.; Baker, D. Atomically Accurate de Novo Design of Antibodies with RFdiffusion. *Nature* **2026**, *649* (8095), 183–193. https://doi.org/10.1038/s41586-025-09721-5.
(265) Kırboğa, K. K. Dynamics-Informed RFdiffusion for de Novo Design of Binders against the KRAS Conformational Ensemble. *J. Biomol. Struct. Dyn.* **2025**, 1–16. https://doi.org/10.1080/07391102.2025.2609688.
(266) Rankovic, S.; Carr, K. D.; Decarreau, J.; Skotheim, R.; Kibler, R. D.; Ols, S.; Lee, S.; Chun, J.-H.; Tooley, M. R.; Dauparas, J.; Eisenach, H. E.; Glögl, M.; Weidle, C.; Borst, A. J.; Baker, D.; King, N. P. Computational Design of Bifaceted Protein Nanomaterials. *Nat. Mater.* **2025**, *24* (10), 1635–1643. https://doi.org/10.1038/s41563-025-02295-7.
(267) Ahern, W.; Yim, J.; Tischer, D.; Salike, S.; Woodbury, S. M.; Kim, D.; Kalvet, I.; Kipnis, Y.; Coventry, B.; Altae-Tran, H. R.; Bauer, M. S.; Barzilay, R.; Jaakkola, T. S.; Krishna, R.; Baker, D. Atom-Level Enzyme Active Site Scaffolding Using RFdiffusion2. *Nat. Methods* **2026**, *23* (1), 96–105. https://doi.org/10.1038/s41592-025-02975-x.
(268) Kim, D.; Woodbury, S. M.; Ahern, W.; Tischer, D.; Kang, A.; Joyce, E.; Bera, A. K.; Hanikel, N.; Salike, S.; Krishna, R.; Yim, J.; Pellock, S. J.; Lauko, A.; Kalvet, I.; Hilvert, D.; Baker, D. Computational Design of Metallohydrolases. *Nature* **2026**, *649* (8095), 246–253. https://doi.org/10.1038/s41586-025-09746-w.
(269) Haley, O. C.; Harding, S.; Sen, T. Z.; Woodhouse, M. R.; Kim, H.-S.; Andorf, C. Application of RFdiffusion to Predict Interspecies Protein-Protein Interactions between Fungal Pathogens and Cereal Crops. bioRxiv September 19, 2024, p 2024.09.17.613523. https://doi.org/10.1101/2024.09.17.613523.
(270) Guo, P.; Noji, H.; Yengo, C. M.; Zhao, Z.; Grainge, I. Biological Nanomotors with a Revolution, Linear, or Rotation Motion Mechanism. *Microbiol. Mol. Biol. Rev. MMBR* **2016**, *80* (1), 161–186. https://doi.org/10.1128/MMBR.00056-15.
(271) Schwartz, C.; De Donatis, G. M.; Zhang, H.; Fang, H.; Guo, P. Revolution Rather than Rotation of AAA+ Hexameric Phi29 Nanomotor for Viral dsDNA Packaging without Coiling. *Virology* **2013**, *443* (1), 28–39. https://doi.org/10.1016/j.virol.2013.04.019.
(272) Guo, P.; Schwartz, C.; Haak, J.; Zhao, Z. Discovery of a New Motion Mechanism of Biomotors Similar to the Earth Revolving around the Sun without Rotation. *Virology* **2013**, *446* (1), 133–143. https://doi.org/10.1016/j.virol.2013.07.025.
(273) Yoshida, M.; Muneyuki, E.; Hisabori, T. ATP Synthase — a Marvellous Rotary Engine of the Cell. *Nat. Rev. Mol. Cell Biol.* **2001**, *2* (9), 669–677. https://doi.org/10.1038/35089509.
(274) DeRosier, D. J. The Turn of the Screw: The Bacterial Flagellar Motor. *Cell* **1998**, *93* (1), 17–20. https://doi.org/10.1016/S0092-8674(00)81141-1.
(275) Steinmetz, E. J.; Platt, T. Evidence Supporting a Tethered Tracking Model for Helicase Activity of Escherichia Coli Rho Factor. *Proc. Natl. Acad. Sci.* **1994**, *91* (4), 1401–1405. https://doi.org/10.1073/pnas.91.4.1401.
(276) Fairman-Williams, M. E.; Guenther, U.-P.; Jankowsky, E. SF1 and SF2 Helicases: Family Matters. *Curr. Opin. Struct. Biol.* **2010**, *20* (3), 313–324. https://doi.org/10.1016/j.sbi.2010.03.011.
(277) Itsathitphaisarn, O.; Wing, R. A.; Eliason, W. K.; Wang, J.; Steitz, T. A. The Hexameric Helicase DnaB Adopts a Nonplanar Conformation during Translocation. *Cell* **2012**, *151* (2), 267–277. https://doi.org/10.1016/j.cell.2012.09.014.
(278) Wang, W.; Cao, L.; Wang, C.; Gigant, B.; Knossow, M. Kinesin, 30 Years Later: Recent Insights from Structural Studies. *Protein Sci.* **2015**, *24* (7), 1047–1056. https://doi.org/10.1002/pro.2697.

(279) Yildiz, A.; Forkey, J. N.; McKinney, S. A.; Ha, T.; Goldman, Y. E.; Selvin, P. R. Myosin V Walks Hand-over-Hand: Single Fluorophore Imaging with 1.5-Nm Localization. *Science* **2003**, *300* (5628), 2061–2065. https://doi.org/10.1126/science.1084398.
(280) Howard, J. Protein Power Strokes. *Curr. Biol.* **2006**, *16* (14), R517–R519. https://doi.org/10.1016/j.cub.2006.06.045.
(281) Brown, A. I.; Sivak, D. A. Theory of Nonequilibrium Free Energy Transduction by Molecular Machines. *Chem. Rev.* **2020**, *120* (1), 434–459. https://doi.org/10.1021/acs.chemrev.9b00254.
(282) Astumian, R. D. Thermodynamics and Kinetics of Molecular Motors. *Biophys. J.* **2010**, *98* (11), 2401–2409. https://doi.org/10.1016/j.bpj.2010.02.040.
(283) Astumian, R. D. Irrelevance of the Power Stroke for the Directionality, Stopping Force, and Optimal Efficiency of Chemically Driven Molecular Machines. *Biophys. J.* **2015**, *108* (2), 291–303. https://doi.org/10.1016/j.bpj.2014.11.3459.
(284) Budaitis, B. G.; Jariwala, S.; Reinemann, D. N.; Schimert, K. I.; Scarabelli, G.; Grant, B. J.; Sept, D.; Lang, M. J.; Verhey, K. J. Neck Linker Docking Is Critical for Kinesin-1 Force Generation in Cells but at a Cost to Motor Speed and Processivity. *eLife* **2019**, *8*, e44146. https://doi.org/10.7554/eLife.44146.
(285) Purcell, T. J.; Morris, C.; Spudich, J. A.; Sweeney, H. L. Role of the Lever Arm in the Processive Stepping of Myosin V. *Proc. Natl. Acad. Sci.* **2002**, *99* (22), 14159–14164. https://doi.org/10.1073/pnas.182539599.
(286) Peskin, C. S.; Odell, G. M.; Oster, G. F. Cellular Motions and Thermal Fluctuations: The Brownian Ratchet. *Biophys. J.* **1993**, *65* (1), 316–324. https://doi.org/10.1016/S0006-3495(93)81035-X.
(287) Hwang, W.; Karplus, M. Structural Basis for Power Stroke vs. Brownian Ratchet Mechanisms of Motor Proteins. *Proc. Natl. Acad. Sci.* **2019**, *116* (40), 19777–19785. https://doi.org/10.1073/pnas.1818589116.
(288) Wagoner, J. A.; Dill, K. A. Molecular Motors: Power Strokes Outperform Brownian Ratchets. *J. Phys. Chem. B* **2016**, *120* (26), 6327–6336. https://doi.org/10.1021/acs.jpcb.6b02776.
(289) Qian, J.; Dunlap, D.; Finzi, L. Basic Mechanisms and Kinetics of Pause-Interspersed Transcript Elongation. *Nucleic Acids Res.* **2021**, *49* (1), 15–24. https://doi.org/10.1093/nar/gkaa1182.
(290) Brueckner, F.; Ortiz, J.; Cramer, P. A Movie of the RNA Polymerase Nucleotide Addition Cycle. *Curr. Opin. Struct. Biol.* **2009**, *19* (3), 294–299. https://doi.org/10.1016/j.sbi.2009.04.005.
(291) Gomez-Marin, A.; Sancho, J. M. Ratchet, Pawl and Spring Brownian Motor. *Phys. Nonlinear Phenom.* **2006**, *216* (1), 214–219. https://doi.org/10.1016/j.physd.2005.12.015.
(292) Frank, J.; Gonzalez, R. L. Structure and Dynamics of a Processive Brownian Motor: The Translating Ribosome. *Annu. Rev. Biochem.* **2010**, *79*, 381–412. https://doi.org/10.1146/annurev-biochem-060408-173330.
(293) Sweeney, H. L.; Holzbaur, E. L. F. Motor Proteins. *Cold Spring Harb. Perspect. Biol.* **2018**, *10* (5), a021931. https://doi.org/10.1101/cshperspect.a021931.
(294) Yildiz, A. How Molecular Motors Move. *Science* **2006**, *311* (5762), 792–793. https://doi.org/10.1126/science.1125068a.
(295) Yildiz, A.; Tomishige, M.; Vale, R. D.; Selvin, P. R. Kinesin Walks Hand-over-Hand. *Science* **2004**, *303* (5658), 676–678. https://doi.org/10.1126/science.1093753.
(296) Velankar, S. S.; Soultanas, P.; Dillingham, M. S.; Subramanya, H. S.; Wigley, D. B. Crystal Structures of Complexes of PcrA DNA Helicase with a DNA Substrate Indicate an Inchworm Mechanism. *Cell* **1999**, *97* (1), 75–84. https://doi.org/10.1016/S0092-8674(00)80716-3.
(297) Raney, K. D.; Byrd, A. K.; Aarattuthodiyil, S. Structure and Mechanisms of SF1 DNA Helicases. *Adv. Exp. Med. Biol.* **2013**, *767*, 10.1007/978-1-4614-5037-5_2. https://doi.org/10.1007/978-1-4614-5037-5_2.

(298) Sweeney, H. L.; Houdusse, A. Structural and Functional Insights into the Myosin Motor Mechanism. *Annu. Rev. Biophys.* **2010**, *39*, 539–557. https://doi.org/10.1146/annurev.biophys.050708.133751.
(299) Bailey, S.; Eliason, W. K.; Steitz, T. A. Structure of Hexameric DnaB Helicase and Its Complex with a Domain of DnaG Primase. *Science* **2007**, *318* (5849), 459–463. https://doi.org/10.1126/science.1147353.
(300) Reyes-Lamothe, R.; Sherratt, D. J.; Leake, M. C. Stoichiometry and Architecture of Active DNA Replication Machinery in Escherichia Coli. *Science* **2010**, *328* (5977), 498–501. https://doi.org/10.1126/science.1185757.
(301) Kozielski, F.; Sack, S.; Marx, A.; Thormählen, M.; Schönbrunn, E.; Biou, V.; Thompson, A.; Mandelkow, E.-M.; Mandelkow, E. The Crystal Structure of Dimeric Kinesin and Implications for Microtubule-Dependent Motility. *Cell* **1997**, *91* (7), 985–994. https://doi.org/10.1016/S0092-8674(00)80489-4.
(302) Baumann, B. A. J.; Taylor, D. W.; Huang, Z.; Tama, F.; Fagnant, P. M.; Trybus, K. M.; Taylor, K. A. Phosphorylated Smooth Muscle Heavy Meromyosin Shows an Open Conformation Linked to Activation. *J. Mol. Biol.* **2012**, *415* (2), 274–287. https://doi.org/10.1016/j.jmb.2011.10.047.
(303) Zhang, K.; Foster, H. E.; Rondelet, A.; Lacey, S. E.; Bahi-Buisson, N.; Bird, A. W.; Carter, A. P. Cryo-EM Reveals How Human Cytoplasmic Dynein Is Auto-Inhibited and Activated. *Cell* **2017**, *169* (7), 1303-1314.e18. https://doi.org/10.1016/j.cell.2017.05.025.
(304) Arias-Palomo, E.; Puri, N.; Murray, V. L. O.; Yan, Q.; Berger, J. M. Physical Basis for the Loading of a Bacterial Replicative Helicase onto DNA. *Mol. Cell* **2019**, *74* (1), 173-184.e4. https://doi.org/10.1016/j.molcel.2019.01.023.
(305) Nebenführ, A.; Dixit, R. Kinesins and Myosins: Molecular Motors That Coordinate Cellular Functions in Plants. *Annu. Rev. Plant Biol.* **2018**, *69*, 329–361. https://doi.org/10.1146/annurev-arplant-042817-040024.
(306) Roberts, A. J.; Kon, T.; Knight, P. J.; Sutoh, K.; Burgess, S. A. Functions and Mechanics of Dynein Motor Proteins. *Nat. Rev. Mol. Cell Biol.* **2013**, *14* (11), 713–726. https://doi.org/10.1038/nrm3667.
(307) Bazin, A.; Cherrier, M. V.; Gutsche, I.; Timmins, J.; Terradot, L. Structure and Primase-Mediated Activation of a Bacterial Dodecameric Replicative Helicase. *Nucleic Acids Res.* **2015**, *43* (17), 8564–8576. https://doi.org/10.1093/nar/gkv792.
(308) Vale, R. D.; Milligan, R. A. The Way Things Move: Looking Under the Hood of Molecular Motor Proteins. *Science* **2000**, *288* (5463), 88–95. https://doi.org/10.1126/science.288.5463.88.
(309) Golcuk, M.; Yilmaz, S. Z.; Yildiz, A.; Gur, M. The Mechanism and Energetics of the Dynein Priming Stroke. *Structure* **2024**, *32* (5), 603-610.e4. https://doi.org/10.1016/j.str.2024.02.003.
(310) Enemark, E. J.; Joshua-Tor, L. Mechanism of DNA Translocation in a Replicative Hexameric Helicase. *Nature* **2006**, *442* (7100), 270–275. https://doi.org/10.1038/nature04943.
(311)Donmez, I.; Patel, S. S. Mechanisms of a Ring Shaped Helicase. *Nucleic Acids Res.* **2006**, *34* (15), 4216–4224. https://doi.org/10.1093/nar/gkl508.
(312) Higuchi, H.; Endow, S. A. Directionality and Processivity of Molecular Motors. *Curr. Opin. Cell Biol.* **2002**, *14* (1), 50–57. https://doi.org/10.1016/S0955-0674(01)00293-9.
(313) Astumian, R. D. Interpreting Molecular Ratchets, Directionality, and Kinetic Asymmetry through the Work of Curie, Einstein, and Lewis. *Commun. Phys.* **2025**, *8* (1), 365. https://doi.org/10.1038/s42005-025-02303-x.
(314) Kapitein, L. C.; Hoogenraad, C. C. Which Way to Go? Cytoskeletal Organization and Polarized Transport in Neurons. *Mol. Cell. Neurosci.* **2011**, *46* (1), 9–20. https://doi.org/10.1016/j.mcn.2010.08.015.
(315) Mickolajczyk, K. J.; Deffenbaugh, N. C.; Ortega Arroyo, J.; Andrecka, J.; Kukura, P.; Hancock, W. O. Kinetics of Nucleotide-Dependent Structural Transitions in the Kinesin-

1 Hydrolysis Cycle. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (52), E7186–E7193. https://doi.org/10.1073/pnas.1517638112.
(316) Niitani, Y.; Matsuzaki, K.; Jonsson, E.; Vale, R. D.; Tomishige, M. Kinetic Regulation of Kinesin's Two Motor Domains Coordinates Its Stepping along Microtubules. *eLife* **2025**, *14*, RP106228. https://doi.org/10.7554/eLife.106228.
(317) Hancock, W. O. The Kinesin-1 Chemomechanical Cycle: Stepping Toward a Consensus. *Biophys. J.* **2016**, *110* (6), 1216–1225. https://doi.org/10.1016/j.bpj.2016.02.025.
(318) Astumian, R. D.; Bier, M. Mechanochemical Coupling of the Motion of Molecular Motors to ATP Hydrolysis. *Biophys. J.* **1996**, *70* (2), 637–653. https://doi.org/10.1016/S0006-3495(96)79605-4.
(319) RP, F. The Feynman Lectures on Physics, 1963, *1*, 46. https://cir.nii.ac.jp/crid/1370285712509472388 (accessed 2026-02-19).
(320) Guydosh, N. R.; Block, S. M. Backsteps Induced by Nucleotide Analogs Suggest the Front Head of Kinesin Is Gated by Strain. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (21), 8054–8059. https://doi.org/10.1073/pnas.0600931103.
(321) Costa, A.; Renault, L.; Swuec, P.; Petojevic, T.; Pesavento, J. J.; Ilves, I.; MacLellan-Gibson, K.; Fleck, R. A.; Botchan, M. R.; Berger, J. M. DNA Binding Polarity, Dimerization, and ATPase Ring Remodeling in the CMG Helicase of the Eukaryotic Replisome. *eLife* **2014**, *3*, e03273. https://doi.org/10.7554/eLife.03273.
(322) O'Donnell, M.; Li, H. The Ring Shaped Hexameric Helicases That Function at DNA Replication Forks. *Nat. Struct. Mol. Biol.* **2018**, *25* (2), 122–130. https://doi.org/10.1038/s41594-018-0024-x.
(323) Bray, D. *Cell Movements: From Molecules to Motility*; Garland Science, 2000.
(324) Howard, J. *Mechanics of Motor Proteins and the Cytoskeleton*, Nachdr.; Sinauer: Sunderland, Mass, 2001.
(325) de Cuevas, M.; Tao, T.; Goldstein, L. S. Evidence That the Stalk of Drosophila Kinesin Heavy Chain Is an Alpha-Helical Coiled Coil. *J. Cell Biol.* **1992**, *116* (4), 957–965. https://doi.org/10.1083/jcb.116.4.957.
(326) Lu, Q.; Ye, F.; Wei, Z.; Wen, Z.; Zhang, M. Antiparallel Coiled-Coil–Mediated Dimerization of Myosin X. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (43), 17388–17393. https://doi.org/10.1073/pnas.1208642109.
(327) Nicholas, M. P.; Höök, P.; Brenner, S.; Wynne, C. L.; Vallee, R. B.; Gennerich, A. Control of Cytoplasmic Dynein Force Production and Processivity by Its C-Terminal Domain. *Nat. Commun.* **2015**, *6* (1), 6206. https://doi.org/10.1038/ncomms7206.
(328) Bi, L.; Qin, Z.; Wang, T.; Li, Y.; Jia, X.; Zhang, X.; Hou, X.-M.; Modesti, M.; Xi, X.-G.; Sun, B. The Convergence of Head-on DNA Unwinding Forks Induces Helicase Oligomerization and Activity Transition. *Proc. Natl. Acad. Sci. U. S. A.* **2022**, *119* (23), e2116462119. https://doi.org/10.1073/pnas.2116462119.
(329) Hyeon, C.; Onuchic, J. N. Internal Strain Regulates the Nucleotide Binding Site of the Kinesin Leading Head. *Proc. Natl. Acad. Sci.* **2007**, *104* (7), 2175–2180. https://doi.org/10.1073/pnas.0610939104.
(330) Kunwar, A.; Vershinin, M.; Xu, J.; Gross, S. P. Stepping, Strain Gating, and an Unexpected Force-Velocity Curve for Multiple-Motor Based Transport. *Curr. Biol.* **2008**, *18* (16), 1173–1183. https://doi.org/10.1016/j.cub.2008.07.027.
(331) Case, R. B.; Rice, S.; Hart, C. L.; Ly, B.; Vale, R. D. Role of the Kinesin Neck Linker and Catalytic Core in Microtubule-Based Motility. *Curr. Biol. CB* **2000**, *10* (3), 157–160. https://doi.org/10.1016/s0960-9822(00)00316-x.
(332) Sun, Y.; Goldman, Y. E. Lever-Arm Mechanics of Processive Myosins. *Biophys. J.* **2011**, *101* (1), 1–11. https://doi.org/10.1016/j.bpj.2011.05.026.
(333) Guydosh, N. R.; Block, S. M. Direct Observation of the Binding State of the Kinesin Head to the Microtubule. *Nature* **2009**, *461* (7260), 125–128. https://doi.org/10.1038/nature08259.
(334) Sun, B.; Johnson, D. S.; Patel, G.; Smith, B. Y.; Pandey, M.; Patel, S. S.; Wang, M. D. ATP-Induced Helicase Slippage Reveals Highly Coordinated Subunits. *Nature* **2011**, *478* (7367), 132–135. https://doi.org/10.1038/nature10409.

(335) Schnitzer, M. J.; Visscher, K.; Block, S. M. Force Production by Single Kinesin Motors. *Nat. Cell Biol.* **2000**, *2* (10), 718–723. https://doi.org/10.1038/35036345.
(336) Mehta, A. Myosin Learns to Walk. *J. Cell Sci.* **2001**, *114* (11), 1981–1998. https://doi.org/10.1242/jcs.114.11.1981.
(337) Hirokawa, N. Kinesin and Dynein Superfamily Proteins and the Mechanism of Organelle Transport. *Science* **1998**, *279* (5350), 519–526. https://doi.org/10.1126/science.279.5350.519.
(338) Matson, S. W. Escherichia Coli Helicase II (urvD Gene Product) Translocates Unidirectionally in a 3' to 5' Direction. *J. Biol. Chem.* **1986**, *261* (22), 10169–10175. https://doi.org/10.1016/S0021-9258(18)67506-4.
(339) Yu, J.; Ha, T.; Schulten, K. Structure-Based Model of the Stepping Motor of PcrA Helicase. *Biophys. J.* **2006**, *91* (6), 2097–2114. https://doi.org/10.1529/biophysj.106.088203.
(340) Ray, B. K.; Lawson, T. G.; Kramer, J. C.; Cladaras, M. H.; Grifo, J. A.; Abramson, R. D.; Merrick, W. C.; Thach, R. E. ATP-Dependent Unwinding of Messenger RNA Structure by Eukaryotic Initiation Factors. *J. Biol. Chem.* **1985**, *260* (12), 7651–7658.
(341) Siddiqui, S. M.; Sauer, R. T.; Baker, T. A. Role of the Processing Pore of the ClpX AAA+ ATPase in the Recognition and Engagement of Specific Protein Substrates. *Genes Dev.* **2004**, *18* (4), 369–374. https://doi.org/10.1101/gad.1170304.
(342) Yao, B.; Sun, H.; Yang, L.; Wang, S.; Liu, X. Recent Progress in Light-Driven Molecular Shuttles. *Front. Chem.* **2022**, *9*, 832735. https://doi.org/10.3389/fchem.2021.832735.
(343) Borsley, S.; Leigh, D. A.; Roberts, B. M. W. Molecular Ratchets and Kinetic Asymmetry: Giving Chemistry Direction. *Angew. Chem. Int. Ed.* **2024**, *63* (23), e202400495. https://doi.org/10.1002/anie.202400495.
(344) Abbasi-Pérez, D.; Sang, H.; Pérez-García, L.; Floris, A.; Amabilino, D. B.; Raval, R.; Recio, J. M.; Kantorovich, L. Controlling the Preferential Motion of Chiral Molecular Walkers on a Surface †Electronic Supplementary Information (ESI) Available. See DOI: 10.1039/C9sc01135h. *Chem. Sci.* **2019**, *10* (23), 5864–5874. https://doi.org/10.1039/c9sc01135h.
(345) Valero, J.; Škugor, M. Mechanisms, Methods of Tracking and Applications of DNA Walkers: A Review. *ChemPhysChem* **2020**, *21* (17), 1971–1988. https://doi.org/10.1002/cphc.202000235.
(346) Song, L.; Zhuge, Y.; Zuo, X.; Li, M.; Wang, F. DNA Walkers for Biosensing Development. *Adv. Sci.* **2022**, *9* (18), 2200327. https://doi.org/10.1002/advs.202200327.
(347) Korosec, C. S.; Unksov, I. N.; Surendiran, P.; Lyttleton, R.; Curmi, P. M. G.; Angstmann, C. N.; Eichhorn, R.; Linke, H.; Forde, N. R. Motility of an Autonomous Protein-Based Artificial Motor That Operates via a Burnt-Bridge Principle. *Nat. Commun.* **2024**, *15* (1), 1511. https://doi.org/10.1038/s41467-024-45570-y.
(348) Thubagere, A. J.; Li, W.; Johnson, R. F.; Chen, Z.; Doroudi, S.; Lee, Y. L.; Izatt, G.; Wittman, S.; Srinivas, N.; Woods, D.; Winfree, E.; Qian, L. A Cargo-Sorting DNA Robot. *Science* **2017**, *357* (6356), eaan6558. https://doi.org/10.1126/science.aan6558.
(349) Lund, K.; Manzo, A. J.; Dabby, N.; Michelotti, N.; Johnson-Buck, A.; Nangreave, J.; Taylor, S.; Pei, R.; Stojanovic, M. N.; Walter, N. G.; Winfree, E.; Yan, H. Molecular Robots Guided by Prescriptive Landscapes. *Nature* **2010**, *465* (7295), 206–210. https://doi.org/10.1038/nature09012.
(350) Mornet, D.; Bonet, A.; Audemard, E.; Bonicel, J. Functional Sequences of the Myosin Head. *J. Muscle Res. Cell Motil.* **1989**, *10* (1), 10–24. https://doi.org/10.1007/BF01739853.
(351) Xu, M.; Tang, D. Recent Advances in DNA Walker Machines and Their Applications Coupled with Signal Amplification Strategies: A Critical Review. *Anal. Chim. Acta* **2021**, *1171*, 338523. https://doi.org/10.1016/j.aca.2021.338523.
(352) Bromley, E. H. C.; Kuwada, N. J.; Zuckermann, M. J.; Donadini, R.; Samii, L.; Blab, G. A.; Gemmen, G. J.; Lopez, B. J.; Curmi, P. M. G.; Forde, N. R.; Woolfson, D. N.; Linke, H. The Tumbleweed: Towards a Synthetic Proteinmotor. *HFSP J.* **2009**, *3* (3), 204–212. https://doi.org/10.2976/1.3111282.

(353) Yin, P.; Turberfield, A. J.; Reif, J. H. Designs of Autonomous Unidirectional Walking DNA Devices. In *DNA Computing*; Ferretti, C., Mauri, G., Zandron, C., Eds.; Springer: Berlin, Heidelberg, 2005; pp 410–425. https://doi.org/10.1007/11493785_36.
(354) A. Watson, M.; L. Cockroft, S. Man-Made Molecular Machines: Membrane Bound. *Chem. Soc. Rev.* **2016**, *45* (22), 6118–6129. https://doi.org/10.1039/C5CS00874C.
(355) Omabegho, T.; Sha, R.; Seeman, N. C. A Bipedal DNA Brownian Motor with Coordinated Legs. *Science* **2009**, *324* (5923), 67–71. https://doi.org/10.1126/science.1170336.
(356) Yao, D.; Bhadra, S.; Xiong, E.; Liang, H.; Ellington, A. D.; Jung, C. Dynamic Programming of a DNA Walker Controlled by Protons. *ACS Nano* **2020**, *14* (4), 4007–4013. https://doi.org/10.1021/acsnano.9b08166.
(357) Ashton, P. R.; Ballardini, R.; Balzani, V.; Baxter, I.; Credi, A.; Fyfe, M. C. T.; Gandolfi, M. T.; Gómez-López, M.; Martínez-Díaz, M.-V.; Piersanti, A.; Spencer, N.; Stoddart, J. F.; Venturi, M.; White, A. J. P.; Williams, D. J. Acid−Base Controllable Molecular Shuttles. *J. Am. Chem. Soc.* **1998**, *120* (46), 11932–11942. https://doi.org/10.1021/ja982167m.
(358) Livoreil, A.; Sauvage, J.-P.; Armaroli, N.; Balzani, V.; Flamigni, L.; Ventura, B. Electrochemically and Photochemically Driven Ring Motions in a Disymmetrical Copper [2]-Catenate. *J. Am. Chem. Soc.* **1997**, *119* (50), 12114–12124. https://doi.org/10.1021/ja9720826.
(359) Zhu, M.; Zhong, X.; Deng, H.; Huang, L.; Yuan, R.; Yuan, Y. Dependent Signal Quenching and Enhancing Triggered by Bipedal DNA Walker for Ultrasensitive Photoelectrochemical Biosensor. *Biosens. Bioelectron.* **2019**, *143*, 111618. https://doi.org/10.1016/j.bios.2019.111618.
(360) Škugor, M.; Valero, J.; Murayama, K.; Centola, M.; Asanuma, H.; Famulok, M. Orthogonally Photocontrolled Non-Autonomous DNA Walker. *Angew. Chem.* **2019**, *58* (21), 6948–6951. https://doi.org/10.1002/anie.201901272.
(361) Wang, Z.-G.; Elbaz, J.; Willner, I. DNA Machines: Bipedal Walker and Stepper. *Nano Lett.* **2011**, *11* (1), 304–309. https://doi.org/10.1021/nl104088s.
(362) Wilson, M. R.; Solà, J.; Carlone, A.; Goldup, S. M.; Lebrasseur, N.; Leigh, D. A. An Autonomous Chemically Fuelled Small-Molecule Motor. *Nature* **2016**, *534* (7606), 235–240. https://doi.org/10.1038/nature18013.
(363) You, M.; Chen, Y.; Zhang, X.; Liu, H.; Wang, R.; Wang, K.; Williams, K. R.; Tan, W. An Autonomous and Controllable Light-Driven DNA Walking Device. *Angew. Chem.* **2012**, *124* (10), 2507–2510. https://doi.org/10.1002/ange.201107733.
(364) Koumura, N.; Zijlstra, R. W. J.; van Delden, R. A.; Harada, N.; Feringa, B. L. Light-Driven Monodirectional Molecular Rotor. *Nature* **1999**, *401* (6749), 152–155. https://doi.org/10.1038/43646.
(365) Tian, Y.; He, Y.; Chen, Y.; Yin, P.; Mao, C. A DNAzyme That Walks Processively and Autonomously along a One-Dimensional Track. *Angew. Chem. Int. Ed.* **2005**, *44* (28), 4355–4358. https://doi.org/10.1002/anie.200500703.
(366) Leigh, D. A. Genesis of the Nanomachines: The 2016 Nobel Prize in Chemistry. *Angew. Chem. Int. Ed.* **2016**, *55* (47).
(367) D. Astumian, R. How Molecular Motors Work – Insights from the Molecular Machinist's Toolbox: The Nobel Prize in Chemistry 2016. *Chem. Sci.* **2017**, *8* (2), 840–845. https://doi.org/10.1039/C6SC04806D.
(368) Gleick, J. *Genius: The Life and Science of Richard Feynman*; Vintage, 1993.
(369) *Nobel Prize in Chemistry 2016*. NobelPrize.org. https://www.nobelprize.org/prizes/chemistry/2016/press-release/ (accessed 2026-02-25).
(370) Bissell, R. A.; Córdova, E.; Kaifer, A. E.; Stoddart, J. F. A Chemically and Electrochemically Switchable Molecular Shuttle. *Nature* **1994**, *369* (6476), 133–137. https://doi.org/10.1038/369133a0.
(371) Blanco, V.; Carlone, A.; Hänni, K. D.; Leigh, D. A.; Lewandowski, B. A Rotaxane-Based Switchable Organocatalyst. **2012**. https://doi.org/10.1002/anie.201201364.

(372) Cheng, C.; McGonigal, P. R.; Schneebeli, S. T.; Li, H.; Vermeulen, N. A.; Ke, C.; Stoddart, J. F. An Artificial Molecular Pump. *Nat. Nanotechnol.* **2015**, *10* (6), 547–553. https://doi.org/10.1038/nnano.2015.96.
(373) Serreli, V.; Lee, C.-F.; Kay, E. R.; Leigh, D. A. A Molecular Information Ratchet. *Nature* **2007**, *445* (7127), 523–527. https://doi.org/10.1038/nature05452.
(374) Shirai, Y.; Osgood, A. J.; Zhao, Y.; Kelly, K. F.; Tour, J. M. Directional Control in Thermally Driven Single-Molecule Nanocars. *Nano Lett.* **2005**, *5* (11), 2330–2334. https://doi.org/10.1021/nl051915k.
(375) Gimzewski, J. K.; Joachim, C.; Schlittler, R. R.; Langlais, V.; Tang, H.; Johannsen, I. Rotation of a Single Molecule Within a Supramolecular Bearing. *Science* **1998**, *281* (5376), 531–533. https://doi.org/10.1126/science.281.5376.531.
(376) Nilsson, P. F.; Zink, A.; Laprévote, O. M. C.; Korosec, C. S.; Gustafsson, N.; Robertson, N. O.; Forde, N. R.; Eichhorn, R.; Höcker, B.; Curmi, P. M. G.; Linke, H. Walking by Design: How to Build Artificial Molecular Motors Made of Proteins. *Nano Futur.* **2024**, *8* (4), 042501. https://doi.org/10.1088/2399-1984/ad7d7e.
(377) Feringa, B. L. The Art of Building Small: From Molecular Switches to Motors (Nobel Lecture). *Angew. Chem. Int. Ed.* **2017**, *56* (37), 11060–11078. https://doi.org/10.1002/anie.201702979.
(378) Zhang, J.; Osgood, A.; Shirai, Y.; Morin, J.-F.; Sasaki, T.; Tour, J. M.; Kelly, K. F. Investigating the Motion of Molecular Machines on Surfaces by STM: The Nanocar and Beyond; IEEE, 2007; pp 243–246.
(379) von Delius, M.; Geertsema, E. M.; Leigh, D. A. A Synthetic Small Molecule That Can Walk down a Track. *Nat. Chem.* **2010**, *2* (2), 96–101. https://doi.org/10.1038/nchem.481.
(380) Zhu, K.; O'Keefe, C. A.; Vukotic, V. N.; Schurko, R. W.; Loeb, S. J. A Molecular Shuttle That Operates inside a Metal–Organic Framework. *Nat. Chem.* **2015**, *7* (6), 514–519. https://doi.org/10.1038/nchem.2258.
(381) Rothemund, P. W. K. Folding DNA to Create Nanoscale Shapes and Patterns. *Nature* **2006**, *440* (7082), 297–302. https://doi.org/10.1038/nature04586.
(382) Seeman, N. C.; Sleiman, H. F. DNA Nanotechnology. *Nat. Rev. Mater.* **2017**, *3* (1), 17068. https://doi.org/10.1038/natrevmats.2017.68.
(383) Simmel, F. C. DNA Nanotechnology Out of Equilibrium. In *Visions of DNA Nanotechnology at 40 for the Next 40 : A Tribute to Nadrian C. Seeman*; Jonoska, N., Winfree, E., Eds.; Springer Nature: Singapore, 2023; pp 17–29. https://doi.org/10.1007/978-981-19-9891-1_2.
(384) Hou, R.; Loh, I. Y.; Li, H.; Wang, Z. Mechanical-Kinetic Modeling of a Molecular Walker from a Modular Design Principle. *Phys. Rev. Appl.* **2017**, *7* (2), 024020.
(385) Simmel, F. C.; Yurke, B.; Singh, H. R. Principles and Applications of Nucleic Acid Strand Displacement Reactions. *Chem. Rev.* **2019**, *119* (10), 6326–6369. https://doi.org/10.1021/acs.chemrev.8b00580.
(386) Astumian, R.; Mukherjee, S.; Warshel, A. The Physics and Physical Chemistry of Molecular Machines. *Chemphyschem Eur. J. Chem. Phys. Phys. Chem.* **2016**, *17* (12), 1719–1741. https://doi.org/10.1002/cphc.201600184.
(387) Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. The Self-Assembly and Dynamic Structure of Cytoskeletal Filaments. In *Molecular Biology of the Cell. 4th edition*; Garland Science, 2002.
(388) Cooper, G. M. Structure and Organization of Actin Filaments. In *The Cell: A Molecular Approach. 2nd edition*; Sinauer Associates, 2000.
(389) Gennerich, A.; Vale, R. D. Walking the Walk: How Kinesin and Dynein Coordinate Their Steps. *Curr. Opin. Cell Biol.* **2009**, *21* (1), 59–67. https://doi.org/10.1016/j.ceb.2008.12.002.
(390) Antal, T.; Krapivsky, P. L. ``Burnt-Bridge'' Mechanism of Molecular Motor Motion. *Phys. Rev. E* **2005**, *72* (4), 046104. https://doi.org/10.1103/PhysRevE.72.046104.
(391) Astumian, R. D. Thermodynamics and Kinetics of a Brownian Motor. *science* **1997**, *276* (5314), 917–922.

(392) Machinek, R. R. F.; Ouldridge, T. E.; Haley, N. E. C.; Bath, J.; Turberfield, A. J. Programmable Energy Landscapes for Kinetic Control of DNA Strand Displacement. *Nat. Commun.* **2014**, *5* (1), 5324. https://doi.org/10.1038/ncomms6324.
(393) Sherman, W. B.; Seeman, N. C. A Precisely Controlled DNA Biped Walking Device. *Nano Lett.* **2004**, *4* (7), 1203–1207. https://doi.org/10.1021/nl049527q.
(394) Shin, J.-S.; Pierce, N. A. A Synthetic DNA Walker for Molecular Transport. *J. Am. Chem. Soc.* **2004**, *126* (35), 10834–10835. https://doi.org/10.1021/ja047543j.
(395) Yehl, K.; Mugler, A.; Vivek, S.; Liu, Y.; Zhang, Y.; Fan, M.; Weeks, E. R.; Salaita, K. High-Speed DNA-Based Rolling Motors Powered by RNase H. *Nat. Nanotechnol.* **2016**, *11* (2), 184–190. https://doi.org/10.1038/nnano.2015.259.
(396) Yao, X.; Wang, L.; Luo, J.; Zhang, C.; Li, Y. Multipedal DNA Walker: Engineering Strategy, Biosensing Application and Future Perspectives. *ACS Sens.* **2024**, *9* (11), 5748–5762. https://doi.org/10.1021/acssensors.4c02251.
(397) *Consolidator Grants 2025: Examples of projects*. ERC. https://erc.europa.eu/news-events/news/consolidator-grants-2025-examples-projects (accessed 2026-02-23).
(398) Siti, W.; Too, H. L.; Anderson, T.; Liu, X. R.; Loh, I. Y.; Wang, Z. Autonomous DNA Molecular Motor Tailor-Designed to Navigate DNA Origami Surface for Fast Complex Motion and Advanced Nanorobotics. *Sci. Adv.* **2023**, *9* (38), eadi8444. https://doi.org/10.1126/sciadv.adi8444.
(399) Anderson, T.; Wu, W.; Sirbu, O.; Tong, K.; Siti, W.; Rui Liu, X.; Kou, B.; Murayama, K.; Asanuma, H.; Wang, Z. A Light-Powered Single-Stranded DNA Molecular Motor with Colour-Selective Single-Step Control. *Angew. Chem. Int. Ed.* **2024**, *63* (32), e202405250. https://doi.org/10.1002/anie.202405250.
(400) Hu, X.; Zhao, X.; Loh, I. Y.; Yan, J.; Wang, Z. Single-Molecule Mechanical Study of an Autonomous Artificial Translational Molecular Motor beyond Bridge-Burning Design. *Nanoscale* **2021**, *13* (31), 13195–13207. https://doi.org/10.1039/D1NR02296B.
(401) Tomov, T. E.; Tsukanov, R.; Liber, M.; Masoud, R.; Plavner, N.; Nir, E. Rational Design of DNA Motors: Fuel Optimization through Single-Molecule Fluorescence. *J. Am. Chem. Soc.* **2013**, *135* (32), 11935–11941. https://doi.org/10.1021/ja4048416.
(402) Jung, C.; Allen, P. B.; Ellington, A. D. A Stochastic DNA Walker That Traverses a Microparticle Surface. *Nat. Nanotechnol.* **2016**, *11* (2), 157–163. https://doi.org/10.1038/nnano.2015.246.
(403) Bath, J.; Green, S. J.; Allen, K. E.; Turberfield, A. J. Mechanism for a Directional, Processive, and Reversible DNA Motor. *Small* **2009**, *5* (13), 1513–1516. https://doi.org/10.1002/smll.200900078.
(404) Tomov, T. E.; Tsukanov, R.; Glick, Y.; Berger, Y.; Liber, M.; Avrahami, D.; Gerber, D.; Nir, E. DNA Bipedal Motor Achieves a Large Number of Steps Due to Operation Using Microfluidics-Based Interface. *ACS Nano* **2017**, *11* (4), 4002–4008. https://doi.org/10.1021/acsnano.7b00547.
(405) Piranej, S.; Bazrafshan, A.; Salaita, K. Chemical-to-Mechanical Molecular Computation Using DNA-Based Motors with Onboard Logic. *Nat. Nanotechnol.* **2022**, *17* (5), 514–523. https://doi.org/10.1038/s41565-022-01080-w.
(406) Zhang, L.; Piranej, S.; Namazi, A.; Narum, S.; Salaita, K. “Turbo-Charged” DNA Motors with Optimized Sequence Enable Single-Molecule Nucleic Acid Sensing. *Angew. Chem. Int. Ed.* **2024**, *63* (13), e202316851. https://doi.org/10.1002/anie.202316851.
(407) Harashima, T.; Otomo, A.; Iino, R. Rational Engineering of DNA-Nanoparticle Motor with High Speed and Processivity Comparable to Motor Proteins. *Nat. Commun.* **2025**, *16* (1), 729. https://doi.org/10.1038/s41467-025-56036-0.
(408) Piranej, S.; Ogasawara, H.; Zhang, L.; Jackson, K.; Bazrafshan, A.; Salaita, K. On-Demand Photoactivation of DNA-Based Motor Motion. *ACS Nano* **2025**, *19* (5), 5363–5375. https://doi.org/10.1021/acsnano.4c13068.
(409) Bazrafshan, A.; Kyriazi, M.-E.; Holt, B. A.; Deng, W.; Piranej, S.; Su, H.; Hu, Y.; El-Sagheer, A. H.; Brown, T.; Kwong, G. A.; Kanaras, A. G.; Salaita, K. DNA Gold Nanoparticle Motors Demonstrate Processive Motion with Bursts of Speed Up to 50

Nm Per Second. *ACS Nano* **2021**, *15* (5), 8427–8438. https://doi.org/10.1021/acsnano.0c10658.
(410) Zhan, P.; Jahnke, K.; Liu, N.; Göpfrich, K. Functional DNA-Based Cytoskeletons for Synthetic Cells. *Nat. Chem.* **2022**, *14* (8), 958–963. https://doi.org/10.1038/s41557-022-00945-w.
(411)Cha, T.-G.; Pan, J.; Chen, H.; Robinson, H. N.; Li, X.; Mao, C.; Choi, J. H. Design Principles of DNA Enzyme-Based Walkers: Translocation Kinetics and Photoregulation. *J. Am. Chem. Soc.* **2015**, *137* (29), 9429–9437. https://doi.org/10.1021/jacs.5b05522.
(412) Pan, J.; Du, Y.; Qiu, H.; Upton, L. R.; Li, F.; Choi, J. H. Mimicking Chemotactic Cell Migration with DNA Programmable Synthetic Vesicles. *Nano Lett.* **2019**, *19* (12), 9138–9144. https://doi.org/10.1021/acs.nanolett.9b04428.
(413) Cha, T.-G.; Pan, J.; Chen, H.; Salgado, J.; Li, X.; Mao, C.; Choi, J. H. A Synthetic DNA Motor That Transports Nanoparticles along Carbon Nanotubes. *Nat. Nanotechnol.* **2014**, *9* (1), 39–43. https://doi.org/10.1038/nnano.2013.257.
(414) Peng, H.; Li, X.-F.; Zhang, H.; Le, X. C. A microRNA-Initiated DNAzyme Motor Operating in Living Cells. *Nat. Commun.* **2017**, *8* (1), 14378. https://doi.org/10.1038/ncomms14378.
(415) Zhou, C.; Duan, X.; Liu, N. A Plasmonic Nanorod That Walks on DNA Origami. *Nat. Commun.* **2015**, *6* (1), 8102. https://doi.org/10.1038/ncomms9102.
(416) Tominaga, M.; Kojima, H.; Yokota, E.; Orii, H.; Nakamori, R.; Katayama, E.; Anson, M.; Shimmen, T.; Oiwa, K. Higher Plant Myosin XI Moves Processively on Actin with 35 Nm Steps at High Velocity. *EMBO J.* **2003**, *22* (6), 1263–1272.
(417) Soppina, V.; Norris, S. R.; Dizaji, A. S.; Kortus, M.; Veatch, S.; Peckham, M.; Verhey, K. J. Dimerization of Mammalian Kinesin-3 Motors Results in Superprocessive Motion. *Proc. Natl. Acad. Sci.* **2014**, *111* (15), 5562–5567. https://doi.org/10.1073/pnas.1400759111.
(418) Verma, V.; Wadsworth, P.; Maresca, T. J. Human Dynein-Dynactin Is a Fast Processive Motor in Living Cells. *eLife* **2025**, *13*. https://doi.org/10.7554/eLife.94963.2.
(419) Imai, H.; Shima, T.; Sutoh, K.; Walker, M. L.; Knight, P. J.; Kon, T.; Burgess, S. A. Direct Observation Shows Superposition and Large Scale Flexibility within Cytoplasmic Dynein Motors Moving along Microtubules. *Nat. Commun.* **2015**, *6* (1), 8179. https://doi.org/10.1038/ncomms9179.
(420) Wirth, J. O.; Scheiderer, L.; Engelhardt, T.; Engelhardt, J.; Matthias, J.; Hell, S. W. MINFLUX Dissects the Unimpeded Walking of Kinesin-1. *Science* **2023**, *379* (6636), 1004–1010. https://doi.org/10.1126/science.ade2650.
(421) Ali, M. Y.; Uemura, S.; Adachi, K.; Itoh, H.; Kinosita, K.; Ishiwata, S. Myosin V Is a Left-Handed Spiral Motor on the Right-Handed Actin Helix. *Nat. Struct. Biol.* **2002**, *9* (6), 464–467. https://doi.org/10.1038/nsb803.
(422) Reck-Peterson, S. L.; Yildiz, A.; Carter, A. P.; Gennerich, A.; Zhang, N.; Vale, R. D. Single-Molecule Analysis of Dynein Processivity and Stepping Behavior. *Cell* **2006**, *126* (2), 335–348. https://doi.org/10.1016/j.cell.2006.05.046.
(423) Melli, L.; Billington, N.; Sun, S. A.; Bird, J. E.; Nagy, A.; Friedman, T. B.; Takagi, Y.; Sellers, J. R. Bipolar Filaments of Human Nonmuscle Myosin 2-A and 2-B Have Distinct Motile and Mechanical Properties. *Elife* **2018**, *7*, e32871.
(424) Norstrom, M. F.; Smithback, P. A.; Rock, R. S. Unconventional Processive Mechanics of Non-Muscle Myosin IIB*. *J. Biol. Chem.* **2010**, *285* (34), 26326–26334. https://doi.org/10.1074/jbc.M110.123851.
(425) Torisawa, T.; Ichikawa, M.; Furuta, A.; Saito, K.; Oiwa, K.; Kojima, H.; Toyoshima, Y. Y.; Furuta, K. Autoinhibition and Cooperative Activation Mechanisms of Cytoplasmic Dynein. *Nat. Cell Biol.* **2014**, *16* (11), 1118–1124. https://doi.org/10.1038/ncb3048.
(426) Li, J.; Johnson-Buck, A.; Yang, Y. R.; Shih, W. M.; Yan, H.; Walter, N. G. Exploring the Speed Limit of Toehold Exchange with a Cartwheeling DNA Acrobat. *Nat. Nanotechnol.* **2018**, *13* (8), 723–729. https://doi.org/10.1038/s41565-018-0130-2.

(427) Zhang, D. Y.; Winfree, E. Control of DNA Strand Displacement Kinetics Using Toehold Exchange. *J. Am. Chem. Soc.* **2009**, *131* (47), 17303–17314. https://doi.org/10.1021/ja906987s.
(428) Chen, J.; Matthews, K. S. Deletion of Lactose Repressor Carboxyl-Terminal Domain Affects Tetramer Formation. *J. Biol. Chem.* **1992**, *267* (20), 13843–13850.
(429) Chen, X.; Zaro, J. L.; Shen, W.-C. Fusion Protein Linkers: Property, Design and Functionality. *Adv. Drug Deliv. Rev.* **2013**, *65* (10), 1357–1369. https://doi.org/10.1016/j.addr.2012.09.039.
(430) Lieser, R. M.; Yur, D.; Sullivan, M. O.; Chen, W. Site-Specific Bioconjugation Approaches for Enhanced Delivery of Protein Therapeutics and Protein Drug Carriers. *Bioconjug. Chem.* **2020**, *31* (10), 2272–2282. https://doi.org/10.1021/acs.bioconjchem.0c00456.
(431) Reddy Chichili, V. P.; Kumar, V.; Sivaraman, J. Linkers in the Structural Biology of Protein–Protein Interactions. *Protein Sci. Publ. Protein Soc.* **2013**, *22* (2), 153–167. https://doi.org/10.1002/pro.2206.
(432) O'Gorman, R. B.; Ferguson, L.; Betz, J. L.; Sadler, J. R.; Matthews, K. S. Determination of the Ligand-Binding Characteristics of Several Tight-Binding Mutants of the Lactose Repressor Protein. *Biochim. Biophys. Acta BBA - Nucleic Acids Protein Synth.* **1981**, *653* (2), 236–247. https://doi.org/10.1016/0005-2787(81)90159-3.
(433) Spotts, R. O.; Chakerian, A. E.; Matthews, K. S. Arginine 197 of Lac Repressor Contributes Significant Energy to Inducer Binding. Confirmation of Homology to Periplasmic Sugar Binding Proteins. *J. Biol. Chem.* **1991**, *266* (34), 22998–23002. https://doi.org/10.1016/S0021-9258(18)54453-7.
(434) Michel, F.; Romero-Romero, S.; Höcker, B. Retracing the Evolution of a Modern Periplasmic Binding Protein. *Protein Sci. Publ. Protein Soc.* **2023**, *32* (11), e4793. https://doi.org/10.1002/pro.4793.
(435) Kalodimos, C. G.; Folkers, G. E.; Boelens, R.; Kaptein, R. Strong DNA Binding by Covalently Linked Dimeric Lac Headpiece: Evidence for the Crucial Role of the Hinge Helices. *Proc. Natl. Acad. Sci.* **2001**, *98* (11), 6039–6044. https://doi.org/10.1073/pnas.101129898.
(436) Hersey, A. N.; Kay, V. E.; Lee, S.; Realff, M. J.; Wilson, C. J. Engineering Allosteric Transcription Factors Guided by the LacI Topology. *Cell Syst.* **2023**, *14* (8), 645–655. https://doi.org/10.1016/j.cels.2023.04.008.
(437) *Study of strong to ultratight protein interactions using differential scanning calorimetry*. ACS Publications. https://pubs.acs.org/doi/pdf/10.1021/bi00481a024 (accessed 2026-05-06).
(438) González, M. Differential Scanning Calorimetry as a Tool to Estimate Binding Parameters in Multiligand Binding Proteins. *Anal. Biochem.* **2006**. https://doi.org/10.1016/J.AB.2005.12.029.
(439) Khan, M. F.; Rahman, M. M.; Xin, Y.; Mustafa, A.; Smith, B. J.; Ottemann, K. M.; Roujeinikova, A. Determination of Protein–Ligand Binding Affinities by Thermal Shift Assay. *ACS Pharmacol. Transl. Sci.* **2024**, *7* (10), 3096–3107. https://doi.org/10.1021/acsptsci.4c00293.
(440) Meinhardt, S.; Swint-Kruse, L. Experimental Identification of Specificity Determinants in the Domain Linker of a LacI/GalR Protein: Bioinformatics-Based Predictions Generate True Positives and False Negatives. *Proteins Struct. Funct. Bioinforma.* **2008**, *73* (4), 941–957. https://doi.org/10.1002/prot.22121.
(441) Dong, F.; Spott, S.; Zimmermann, O.; Kisters-Woike, B.; Müller-Hill, B.; Barker, A. Dimerisation Mutants of Lac Repressor. I. a Monomeric Mutant, L251A, That Binds *Lac* Operator DNA as a Dimer1. *J. Mol. Biol.* **1999**, *290* (3), 653–666. https://doi.org/10.1006/jmbi.1999.2902.
(442) Daly, T. J.; Matthews, K. S. Characterization and Modification of a Monomeric Mutant of the Lactose Repressor Protein. *Biochemistry* **1986**, *25* (19), 5474–5478. https://doi.org/10.1021/bi00367a019.
(443) Spott, S.; Dong, F.; Kisters-Woike, B.; Müller-Hill, B. Dimerisation Mutants of Lac Repressor. II. A Single Amino Acid Substitution, D278L, Changes the Specificity of

Dimerisation1. *J. Mol. Biol.* **2000**, *296* (2), 673–684. https://doi.org/10.1006/jmbi.1999.3469.
(444) Kröger, P.; Shanmugaratnam, S.; Scheib, U.; Höcker, B. Fine-Tuning Spermidine Binding Modes in the Putrescine Binding Protein PotF. *J. Biol. Chem.* **2021**, *297* (6).
(445) Scherer, M.; Kriegel, M.; Höcker, B.; Fleishman, S. J. Optimizing Stability in Dynamic Small-Molecule Binding Proteins. *J. Am. Chem. Soc.* **2026**, *148* (1), 1911–1918. https://doi.org/10.1021/jacs.5c19571.
(446) Zimmerman, S. B.; Trach, S. O. Estimation of Macromolecule Concentrations and Excluded Volume Effects for the Cytoplasm of Escherichia Coli. *J. Mol. Biol.* **1991**, *222* (3), 599–620. https://doi.org/10.1016/0022-2836(91)90499-v.
(447) Sarkar, M.; Li, C.; Pielak, G. J. Soft Interactions and Crowding. *Biophys. Rev.* **2013**, *5* (2), 187–194. https://doi.org/10.1007/s12551-013-0104-4.
(448) Monteith, W. B.; Cohen, R. D.; Smith, A. E.; Guzman-Cisneros, E.; Pielak, G. J. Quinary Structure Modulates Protein Stability in Cells. *Proc. Natl. Acad. Sci.* **2015**, *112* (6), 1739–1742. https://doi.org/10.1073/pnas.1417415112.
(449) Guseman, A. J.; Pielak, G. J. Cosolute and Crowding Effects on a Side-By-Side Protein Dimer. *Biochemistry* **2017**, *56* (7), 971–976. https://doi.org/10.1021/acs.biochem.6b01251.
(450) Smith, A. E.; Zhou, L. Z.; Gorensek, A. H.; Senske, M.; Pielak, G. J. In-Cell Thermodynamics and a New Role for Protein Surfaces. *Proc. Natl. Acad. Sci.* **2016**, *113* (7), 1725–1730. https://doi.org/10.1073/pnas.1518620113.
(451) Speer, S. L.; Zheng, W.; Jiang, X.; Chu, I.-T.; Guseman, A. J.; Liu, M.; Pielak, G. J.; Li, C. The Intracellular Environment Affects Protein–Protein Interactions. *Proc. Natl. Acad. Sci.* **2021**, *118* (11), e2019918118. https://doi.org/10.1073/pnas.2019918118.
(452) Zhang, N.; An, L.; Li, J.; Liu, Z.; Yao, L. Quinary Interactions Weaken the Electric Field Generated by Protein Side-Chain Charges in the Cell-like Environment. *J. Am. Chem. Soc.* **2017**, *139* (2), 647–654. https://doi.org/10.1021/jacs.6b11058.
(453) Kyne, C.; Jordon, K.; Filoti, D. I.; Laue, T. M.; Crowley, P. B. Protein Charge Determination and Implications for Interactions in Cell Extracts. *Protein Sci.* **2017**, *26* (2), 258–267. https://doi.org/10.1002/pro.3077.
(454) Magliery, T. J.; Wilson, C. G. M.; Pan, W.; Mishler, D.; Ghosh, I.; Hamilton, A. D.; Regan, L. Detecting Protein−Protein Interactions with a Green Fluorescent Protein Fragment Reassembly Trap: Scope and Mechanism. *J. Am. Chem. Soc.* **2005**, *127* (1), 146–157. https://doi.org/10.1021/ja046699g.
(455) Köker, T.; Fernandez, A.; Pinaud, F. Characterization of Split Fluorescent Protein Variants and Quantitative Analyses of Their Self-Assembly Process. *Sci. Rep.* **2018**, *8*, 5344. https://doi.org/10.1038/s41598-018-23625-7.
(456) Hatlem, D.; Trunk, T.; Linke, D.; Leo, J. C. Catching a SPY: Using the SpyCatcher-SpyTag and Related Systems for Labeling and Localizing Bacterial Proteins. *Int. J. Mol. Sci.* **2019**, *20* (9), 2129. https://doi.org/10.3390/ijms20092129.
(457) Liu, C.; Wu, K.; Choi, H.; Han, H. L.; Zhang, X.; Watson, J. L.; Ahn, G.; Zhang, J. Z.; Shijo, S.; Good, L. L.; Fischer, C. M.; Bera, A. K.; Kang, A.; Brackenbrough, E.; Coventry, B.; Hick, D. R.; Qamar, S.; Li, X.; Decarreau, J.; Gerben, S. R.; Yang, W.; Goreshnik, I.; Vafeados, D.; Wang, X.; Lamb, M.; Murray, A.; Kenny, S.; Bauer, M. S.; Hoofnagle, A. N.; Zhu, P.; Knowles, T. P. J.; Baker, D. Diffusing Protein Binders to Intrinsically Disordered Proteins. *Nature* **2025**, *644* (8077), 809–817. https://doi.org/10.1038/s41586-025-09248-9.
(458) Vázquez Torres, S.; Leung, P. J. Y.; Venkatesh, P.; Lutz, I. D.; Hink, F.; Huynh, H.-H.; Becker, J.; Yeh, A. H.-W.; Juergens, D.; Bennett, N. R.; Hoofnagle, A. N.; Huang, E.; MacCoss, M. J.; Expòsit, M.; Lee, G. R.; Bera, A. K.; Kang, A.; De La Cruz, J.; Levine, P. M.; Li, X.; Lamb, M.; Gerben, S. R.; Murray, A.; Heine, P.; Korkmaz, E. N.; Nivala, J.; Stewart, L.; Watson, J. L.; Rogers, J. M.; Baker, D. De Novo Design of High-Affinity Binders of Bioactive Helical Peptides. *Nature* **2024**, *626* (7998), 435–442. https://doi.org/10.1038/s41586-023-06953-1.

(459) Sappington, I.; Toul, M.; Lee, D. S.; Robinson, S. A.; Goreshnik, I.; McCurdy, C.; Chan, T. C.; Buchholz, N.; Huang, B.; Vafeados, D.; Garcia-Sanchez, M.; Roullier, N.; Glögl, M.; Kim, C. J.; Watson, J. L.; Torres, S. V.; Verschueren, K. H. G.; Verstraete, K.; Hinck, C. S.; Benard-Valle, M.; Coventry, B.; Sims, J. N.; Ahn, G.; Wang, X.; Hinck, A. P.; Jenkins, T. P.; Ruohola-Baker, H.; Banik, S. M.; Savvides, S. N.; Baker, D. Improved Protein Binder Design Using β-Pairing Targeted RFdiffusion. *Nat. Commun.* **2026**, *17* (1), 1101. https://doi.org/10.1038/s41467-025-67866-3.

(460) Pillai, A.; Idris, A.; Philomin, A.; Weidle, C.; Skotheim, R.; Leung, P. J. Y.; Broerman, A.; Demakis, C.; Borst, A. J.; Praetorius, F.; Baker, D. De Novo Design of Allosterically Switchable Protein Assemblies. *Nature* **2024**, *632* (8026), 911–920. https://doi.org/10.1038/s41586-024-07813-2.

(461) Gao, K.; Zhang, X.; Nie, J.; Meng, H.; Zhang, W.; Tian, B.; Liu, X. De Novo Design of a Fusion Protein Tool for GPCR Research. *Proc. Natl. Acad. Sci.* **2025**, *122* (29), e2422360122. https://doi.org/10.1073/pnas.2422360122.

(462) Salgado, H.; Santos-Zavaleta, A.; Gama-Castro, S.; Millán-Zárate, D.; Díaz-Peredo, E.; Sánchez-Solano, F.; Pérez-Rueda, E.; Bonavides-Martínez, C.; Collado-Vides, J. RegulonDB (Version 3.2): Transcriptional Regulation and Operon Organization in Escherichia Coli K-12. *Nucleic Acids Res.* **2001**, *29* (1), 72–74. https://doi.org/10.1093/nar/29.1.72.

(463) AbuOun, M.; Suthers, P. F.; Jones, G. I.; Carter, B. R.; Saunders, M. P.; Maranas, C. D.; Woodward, M. J.; Anjum, M. F. Genome Scale Reconstruction of a Salmonella Metabolic Model: Comparison of Similarity and Differences with a Commensal Escherichia Coli Strain. *J. Biol. Chem.* **2009**, *284* (43), 29480–29488. https://doi.org/10.1074/jbc.M109.005868.

(464) Morabbi Heravi, K.; Watzlawick, H.; Altenbuchner, J. The melREDCA Operon Encodes a Utilization System for the Raffinose Family of Oligosaccharides in Bacillus Subtilis. *J. Bacteriol.* **2019**, *201* (15), e00109-19. https://doi.org/10.1128/JB.00109-19.

(465) Reeve, E. C. R. The Lactose System in *Klebsiella Aerogenes* V9A: 3. Specific Repression of the *Lac* Operon by Melibiose and Raffinose. *Genet. Res.* **1973**, *22* (2), 217–221. https://doi.org/10.1017/S0016672300013021.

(466) van Leeuwen, H. C.; Strating, M. J.; Rensen, M.; de Laat, W.; van der Vliet, P. C. Linker Length and Composition Influence the Flexibility of Oct-1 DNA Binding. *EMBO J.* **1997**, *16* (8), 2043–2053. https://doi.org/10.1093/emboj/16.8.2043.

(467) Varming, A. K.; Rasmussen, K. K.; Zong, Z.; Thulstrup, P. W.; Kilstrup, M.; Lo Leggio, L. Flexible Linker Modulates the Binding Affinity of the TP901-1 CI Phage Repressor to DNA. *FEBS J.* **2022**, *289* (4), 1135–1148. https://doi.org/10.1111/febs.16238.

(468) Lewis, M. The *Lac* Repressor. *C. R. Biol.* **2005**, *328* (6), 521–548. https://doi.org/10.1016/j.crvi.2005.04.004.

(469) Zhan, H.; Swint-Kruse, L.; Matthews, K. S. Extrinsic Interactions Dominate Helical Propensity in Coupled Binding and Folding of Lactose Repressor Protein Hinge Helix. *Biochemistry* **2006**, *45* (18), 5896–5906. https://doi.org/10.1021/bi052619p.

(470) Falcon, C. M.; Matthews, K. S. Glycine Insertion in the Hinge Region of Lactose Repressor Protein Alters DNA Binding. *J. Biol. Chem.* **1999**, *274* (43), 30849–30857. https://doi.org/10.1074/jbc.274.43.30849.

(471) Weinstein, J. J.; Goldenzweig, A.; Hoch, S.; Fleishman, S. J. PROSS 2: A New Server for the Design of Stable and Highly Expressed Protein Variants. *Bioinformatics* **2021**, *37* (1), 123–125. https://doi.org/10.1093/bioinformatics/btaa1071.

(472) Dieckhaus, H.; Brocidiacono, M.; Randolph, N. Z.; Kuhlman, B. Transfer Learning to Leverage Larger Datasets for Improved Prediction of Protein Stability Changes. *Proc. Natl. Acad. Sci.* **2024**, *121* (6), e2314853121. https://doi.org/10.1073/pnas.2314853121.

(473) Pucci, F.; Rooman, M. Improved Insights into Protein Thermal Stability: From the Molecular to the Structurome Scale. *Philos. Transact. A Math. Phys. Eng. Sci.* **2016**, *374* (2080), 20160141. https://doi.org/10.1098/rsta.2016.0141.

(474) Goldenzweig, A.; Fleishman, S. J. Principles of Protein Stability and Their Application in Computational Design. *Annu. Rev. Biochem.* **2018**, *87*, 105–129. https://doi.org/10.1146/annurev-biochem-062917-012102.
(475) Tokuriki, N.; Stricher, F.; Schymkowitz, J.; Serrano, L.; Tawfik, D. S. The Stability Effects of Protein Mutations Appear to Be Universally Distributed. *J. Mol. Biol.* **2007**, *369* (5), 1318–1332. https://doi.org/10.1016/j.jmb.2007.03.069.
(476) Bahadur, R. P.; Zacharias, M. The Interface of Protein-Protein Complexes: Analysis of Contacts and Prediction of Interactions. *Cell. Mol. Life Sci. CMLS* **2008**, *65* (7–8), 1059–1072. https://doi.org/10.1007/s00018-007-7451-x.
(477) Bell, C. E.; Barry, J.; Matthews, K. S.; Lewis, M. Structure of a Variant of Lac Repressor with Increased Thermostability and Decreased Affinity for Operator. *J. Mol. Biol.* **2001**, *313* (1), 99–109. https://doi.org/10.1006/jmbi.2001.5041.
(478) Savino, S.; Desmet, T.; Franceus, J. Insertions and Deletions in Protein Evolution and Engineering. *Biotechnol. Adv.* **2022**, *60*, 108010. https://doi.org/10.1016/j.biotechadv.2022.108010.
(479) Yang, Y.; Braga, M. V.; Dean, M. D. Insertion–Deletion Events Are Depleted in Protein Regions with Predicted Secondary Structure. *Genome Biol. Evol.* **2024**, *16* (5), evae093. https://doi.org/10.1093/gbe/evae093.
(480) Reddy Chichili, V. P.; Kumar, V.; Sivaraman, J. Linkers in the Structural Biology of Protein–Protein Interactions. *Protein Sci. Publ. Protein Soc.* **2013**, *22* (2), 153–167. https://doi.org/10.1002/pro.2206.
(481) Keeble, A. H.; Turkki, P.; Stokes, S.; Khairil Anuar, I. N. A.; Rahikainen, R.; Hytönen, V. P.; Howarth, M. Approaching Infinite Affinity through Engineering of Peptide–Protein Interaction. *Proc. Natl. Acad. Sci.* **2019**, *116* (52), 26523–26533. https://doi.org/10.1073/pnas.1909653116.
(482) Ma, X.; Yi, L.; Li, J.; Ma, J.; Li, W.; Wang, L.; Liu, C.; Li, B.; Deng, N.; Kang, W.; Xue, C. Engineering a Self-Assembled Protein Cage for Targeted Dual Functionalization. *Nano Lett.* **2024**, *24* (30), 9237–9244. https://doi.org/10.1021/acs.nanolett.4c01693.
(483) Raupach, J.; Housria, Y.; Hein, S.; Schnierle, B.; Durmaz, H.; Zahn, T.; Hildt, E. Comparative Analysis of a Flexible Novel Vaccine Platform Based on Cell Permeable Chimeric HBV Capsids for Mucosal Vaccination. *Vaccine* **2026**, *69*, 127987. https://doi.org/10.1016/j.vaccine.2025.127987.
(484) Dicks, M. D. J.; Rose, L. M.; Russell, R. A.; Bowman, L. A. H.; Graham, C.; Jimenez-Guardeño, J. M.; Doores, K. J.; Malim, M. H.; Draper, S. J.; Howarth, M.; Biswas, S. Modular Capsid Decoration Boosts Adenovirus Vaccine-Induced Humoral Immunity against SARS-CoV-2. *Mol. Ther.* **2022**, *30* (12), 3639–3657. https://doi.org/10.1016/j.ymthe.2022.08.002.
(485) Cai, M.; Ni, Z.; Sun, Z.; Li, X.; Qi, Y.; Wang, L.; He, C.; Li, C. The SpyTag/SpyCatcher System: Precise Regulation of Covalent Conjugation and Expansion of Application Scenarios. *Biotechnol. J.* **2025**, *20* (10), e70131. https://doi.org/10.1002/biot.70131.
(486) LLC, G. S. *pET-21b(+) Sequence and Map*. https://www.snapgene.com/plasmids/pet_and_duet_vectors_(novagen)/pET-21b(%2B) (accessed 2026-05-07).
(487) LLC, G. S. *pETDuet-1 Sequence and Map*. https://www.snapgene.com/plasmids/pet_and_duet_vectors_(novagen)/pETDuet-1 (accessed 2026-05-07).
(488) *Addgene: Vector Database - pBAD/His A*. https://www.addgene.org/vector-database/1827/ (accessed 2026-05-07).
(489) *DNASU Plasmid | Detailed Vector Information: pCDF-BAD*. DNASU Plasmids. https://dnasu.org/DNASU/GetVectorDetail.do?vectorid=648 (accessed 2026-05-07).
(490) QuickC. *QuickChange Primer Design*. https://www.agilent.com/store/primerDesignProgram.jsp?srsltid=AfmBOopHnVPabP7BUkldkk9oU86OKqzqo7juDGbvL91Q3QEDiP2wcm27 (accessed 2026-01-27).

(491) *Oligo Analysis Tool | Eurofins Genomics US*. https://eurofinsgenomics.com/en/resources/tools/oligo-analysis (accessed 2026-01-27).
(492) *_d4029_d4030_dna_clean_concentrator-100_kit.pdf*. https://files.zymoresearch.com/protocols/_d4029_d4030_dna_clean_concentrator-100_kit.pdf?_gl=1*1jz49a3*_gcl_au*OTc5MzcyMjUzLjE3NjgzMTY1OTA (accessed 2026-01-27).
(493) MACHERY-NAGEL GmbH & Co. KG. *Instruction-NucleoSpin-Gel-and-PCR-Clean-up*. https://www.mn-net.com/media/pdf/02/1a/74/Instruction-NucleoSpin-Gel-and-PCR-Clean-up.pdf (accessed 2026-01-27).
(494) Gallagher, S. R. Quantitation of DNA and RNA with Absorption and Fluorescence Spectroscopy. *Curr. Protoc. Mol. Biol.* **2011**, *93* (1), A.3D.1-A.3D.14. https://doi.org/10.1002/0471142727.mba03ds93.
(495) Qamar, W.; Khan, M. R.; Arafah, A. Optimization of Conditions to Extract High Quality DNA for PCR Analysis from Whole Blood Using SDS-Proteinase K Method. *Saudi J. Biol. Sci.* **2017**, *24* (7), 1465–1469. https://doi.org/10.1016/j.sjbs.2016.09.016.
(496) MACHERY-NAGEL GmbH & Co. KG. *Instruction-NucleoSpin-Plasmid*. https://www.mn-net.com/media/pdf/45/51/02/Instruction-NucleoSpin-Plasmid.pdf (accessed 2026-01-27).
(497) Beer. Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten. *Ann. Phys.* **1852**, *162* (5), 78–88. https://doi.org/10.1002/andp.18521620505.
(498) *Expasy - ProtParam*. https://web.expasy.org/protparam/ (accessed 2026-03-18).
(499) Többens, D. M.; Zander, S. KMC-2: An X-Ray Beamline with Dedicated Diffraction and XAS Endstations at BESSY II. *J. Large-Scale Res. Facil. JLSRF* **2016**, *2*, A49–A49. https://doi.org/10.17815/jlsrf-2-65.
(500) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser Crystallographic Software. *J. Appl. Crystallogr.* **2007**, *40* (Pt 4), 658–674. https://doi.org/10.1107/S0021889807021206.
(501) Yang, Z.; Zeng, X.; Zhao, Y.; Chen, R. AlphaFold2 and Its Applications in the Fields of Biology and Medicine. *Signal Transduct. Target. Ther.* **2023**, *8* (1), 115. https://doi.org/10.1038/s41392-023-01381-z.
(502) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66* (Pt 4), 486–501. https://doi.org/10.1107/S0907444910007493.
(503) Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D. Towards Automated Crystallographic Structure Refinement with Phenix.Refine. *Acta Crystallogr. D Biol. Crystallogr.* **2012**, *68* (Pt 4), 352–367. https://doi.org/10.1107/S0907444912001308.
(504) Cock, P. J. A.; Antao, T.; Chang, J. T.; Chapman, B. A.; Cox, C. J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B.; de Hoon, M. J. L. Biopython: Freely Available Python Tools for Computational Molecular Biology and Bioinformatics. *Bioinformatics* **2009**, *25* (11), 1422–1423. https://doi.org/10.1093/bioinformatics/btp163.
(505) Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold: Making Protein Folding Accessible to All. *Nat. Methods* **2022**, *19* (6), 679–682. https://doi.org/10.1038/s41592-022-01488-1.
(506) Nishikawa, K.; Ooi, T.; Isogai, Y.; Saitô, N. Tertiary Structure of Proteins. I. Representation and Computation of the Conformations. *J. Phys. Soc. Jpn.* **1972**, *32* (5), 1331–1337. https://doi.org/10.1143/JPSJ.32.1331.

7 Supplementary Information

7.1 Additional information about constructs

Table 37. Summary of the purification steps performed for each construct.

Construct	Purification Step 1	Purification Step 2	Purification Step 3
LacI-A1	Ni^{2+}-IMAC	SEC	-
LacI-B1	Ni^{2+}-IMAC	SEC	-
LacI-A2	Strep-Tactin-AC	SEC	-
LacI-B2	Ni^{2+}-IMAC	SEC	-
LacI-B3	Ni^{2+}-IMAC	SEC	-
LacI-splitGFP	Ni^{2+}-IMAC	rvIMAC	SEC
LacI*-SpyCatcher	Strep-Tactin-AC	SEC	
LacI-A2-SpyTag	Ni^{2+}-IMAC	SEC	
Het-SpyTagxCatcher	Strep-Tactin-AC	Ni^{2+}-IMAC	SEC
LacI-Lev1	Ni^{2+}-IMAC	rvIMAC	SEC
LacI-Lev2	Ni^{2+}-IMAC	rvIMAC	SEC
PBP1	Ni^{2+}-IMAC	rvIMAC	SEC
PBP2	Ni^{2+}-IMAC	rvIMAC	SEC
LacI-Link1	Strep-Tactin-AC	SEC	
LacI-Link2	Strep-Tactin-AC	SEC	
LacI-Link3	Strep-Tactin-AC	SEC	
LacI-Link4	Strep-Tactin-AC	SEC	
LacI-Link5	Strep-Tactin-AC	SEC	
Het-Link1-SpyTagxCatcher	Strep-Tactin-AC	Ni^{2+}-IMAC	SEC
Het-Link2-SpyTagxCatcher	Strep-Tactin-AC	Ni^{2+}-IMAC	SEC
LLhT	Strep-Tactin-AC	SEC	
LLhF	Strep-Tactin-AC	SEC	
LLhT-Lev1	Ni^{2+}-IMAC	SEC	
LLhT-Lev7	Ni^{2+}-IMAC	SEC	
LLhT-Lev8	Ni^{2+}-IMAC	SEC	
LLhT-Link6	Ni^{2+}-IMAC	SEC	
LLhT-Link7	Ni^{2+}-IMAC	SEC	
LLhT-271A	Strep-Tactin-AC	SEC	
LLhT-272A	Strep-Tactin-AC	SEC	
LLhT-272S	Strep-Tactin-AC	SEC	
LLhT-268A	Strep-Tactin-AC	SEC	
LLhT-244A	Strep-Tactin-AC	SEC	
LLhT-4Mut	Strep-Tactin-AC	SEC	
LLhT-6Mut	Strep-Tactin-AC	SEC	

Construct	Purification Step 1	Purification Step 2	Purification Step 3
LLhT-6Mut-Link1-SpyCatcher-DogTag LLhT-6Mut-DogCatcher			
LacI-A2-Link1-SpyCatcher-DogTag	Ni^{2+}-IMAC	SEC	
LacI*-DogCatcher	Ni^{2+}-IMAC	SEC	
Het-Link1-DogTagxCatcher	Ni^{2+}-IMAC	SEC	
CM-PotF	Ni^{2+}-IMAC	SEC	

Table 38. Overview of the analysed protein constructs and their calculated molecular properties. The table lists the theoretical molecular weight (MW), extinction coefficient, isoelectric point (pI), and amino acid number for each construct.

Construct	MW [Da]	extinction coefficient [M^{-1} cm^{-1}]	pI	Nr. AS
LacI	36546	23045	6.34	339
LacI-A1	40476	34170	5.74	376
LacI-B1	36547	21555	6.49	339
LacI-A2	37410	25565	5.77	349
LacI-B2	36625	32555	6.49	339
LacI-B3	36553	27055	6.67	339
LacI-splitGFP	106054	89115	5.87	983
LacI-A2-CleavedGFP	37423	21555	5.63	349
LacI-B2-CleavedGFP	36954	34045	5.82	343
LacI*-SpyCatcher	48235	38515	5.62	451
LacI-A2-SpyTag	39454	23045	6.04	370
Het-SpyTagxCatcher	87689	61560	5.83	821
LacI-Lev1	62154	42080	6.21	571
without splitGFP	39050	24535	5.83	366
LacI-Lev2	61943	42080	6.29	570
without splitGFP	38839	24535	6.00	365
PBP1	55606	36120	6.02	511
without splitGFP	32502	18575	5.37	306
PBP2	55395	36120	6.09	510
without splitGFP	32291	18575	5.52	305
LacI-Link1	44105	33015	5.91	404
LacI-Link2	44843	34505	5.56	410
LacI-Link3	43985	34505	5.35	407
LacI-Link4	43800	33015	5.87	413
LacI-Link5	44717	33015	5.43	417
LacI-A2-Link1-SpyTag	44652	24535	6.11	414

Construct	MW [Da]	extinction coefficient [M^{-1} cm^{-1}]	pI	Nr. AS
Het-Link1-SpyTagxCatcher	93626	64540	5.86	871
LacI-A2-Link2-SpyTag	45390	26025	5.83	420
Het-Link2-SpyTagxCatcher	92887	63050	5.72	865
LacI-Gentio	59613	42080	6.24	644
without splitGFP	35715	23045	6.07	333
LLhF	40049	41495	5.72	352
LLhT	36477	32110	6.56	332
LLhT-Lev1	39084	26610	6.16	359
LLhT-Lev7	37040	26360	6.43	337
LLhT-Lev8	36098	24870	6.43	328
LLhT-Link6	46590	36580	6.19	430
LLhT-Link7	48845	45185	6.39	446
LLhT-271A	36401	32110	6.56	322
LLhT-272A	36435	32110	6.56	322
LLhT-272S	36451	32110	6.56	322
LLhT-268A	36435	32110	6.56	322
LLhT-244A	36435	32110	6.56	322
LLhT-4Mut	36275	32110	6.56	322
LLhT-6Mut	36158	32110	6.56	322
LLhT-6Mut-Lev7-SpyCatcher-DogTag	48825	39560	5.95	450
LLhT-6Mut-DogCatcher	46413	31080	5.95	428
LacI-A2-Link1-SpyCatcher-DogTag	54253	34505	5.84	504
LacI*-DogCatcher	48878	29005	6.17	452
Het-Link1-DogTagxCatcher	103131	63510	6.00	956
CM-PotF	55591	68425	5.82	503

Table 39. Overview of the constructs and corresponding amino acid sequences. DNA-binding domains of LacI- and LLhT-derived constructs are indicated in blue, while effector-binding domains are shown in red. TEV protease cleavage sites, where present, are shown in orange, and purification tags, either His_6 tag or Strep-II tag, are indicated in purple. Linker regions connecting functional domains and the LacI hinge helix are shown in black. Distinct design features are colour coded as follows: SpyTag and SpyCatcher are highlighted in grey, DogTag and DogCatcher in pink, splitGFP elements in dark green, introduced point mutations in light green, and designed extensions generated using RFdiffusion and/or ProteinMPNN in yellow.

Construct	AS Sequence
LacI	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVV SMVERSGVEACKAAVHNLLAQRVSGLIINYPLDDQDAIAVEAACTNVP ALFLDVSDQTPINSIIFSHEDGTRLGVEHLVALGHQQIALLAGPLSSVSA RLRLAGWHKYLTRNQIQPIAEREGDWSAMSGFQQTMQMLNEGIVPTA MLVANDQMALGAMRAITESGLRVGADISVVGYDDTEDSSCYIPPLTTIK QDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVKRKTTLALEHHHH HH
LacI-R197A	HMKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELN YIPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVV VSMVERSGVEACKAAVHNLLAQRVSGLIINYPLDDQDAIAVEAACTNV PALFLDVSDQTPINSIIFSHEDGTRLGVEHLVALGHQQIALLAGPLSSVS ARLALAGWHKYLTRNQIQPIAEREGDWSAMSGFQQTMQMLNEGIVP TAMLVANDQMALGAMRAITESGLRVGADISVVGYDDTEDSSCYIPPLT TIKQDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVKRKTTLALEHH HHHH
LacI-A1	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVV SMVERSGVEACKAAVHNLLAQRVSGLIINYPLDDQDAIAVEAACTNVP ALFLDVSDQTPINSIIFSHEDGTRLGVEHLVALGHQQIALLAGPLSSVSA RLRLAGWHKYLTRNQIQPIAEREGDWSAMSGFQQTMQMLNEGIVPTA MLVANDQMALGAMRAITESGLRVGADISVVGYDDTEDSSCAIPPLTTIK QDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVKRKTTLALEHHHH HH
LacI-B1	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVV SMVERSGVEACKAAVHNLLAQRVSGLIINYPLDDQDAIAVEAACTNVP ALFLDVSDQTPINSIIFSHEDGTRLGVEHLVALGHQQIALLAGPLSSVSA RLRLAGWHKYLTRNQIQPIAEREGDWSAMSGFQQTMQMLNEGIVPTA MLVANDQMAMGAMRAITESGLRVGADISVVGYDDTENSSCFIPPLTTI KQDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVKRKTTLALEHHH HHH
LacI-A2	MKPVTLYDVAEYAGVSTATVSNVVNQASHVSAKTREKVEAAMAELNYI PNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVVS MVERSGVEACKAAVHNLLAQRVSGLIINYPLDDQDAIAVEAACTNVPA LFLDVSDQTPINSIIFSHEDGTRLGVEHLVALGHQQIALLAGPLSSVSAR LRLAGWHKYLTRNQIQPIAEREGDWSAMSGFQQTMQMLNEGIVPTA MLVANDQMALGAMRAITESGLRVGADISVVGYDDTEDSSCSIPPLTTIK QDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVKRKTTLALELEVLF QGPWSHPQFEK

Construct	AS Sequence
LacI-B2	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVV SMVERSGVEACKAAVHNLLAQRVSGLIINYPLDDQDAIAVEAACTNVP ALFLDVSDQTPINSIIFSHEDGTRLGVEHLVALGHQQIALLAGPLSSVSA RLRLAGWHKYLTRNQIQPIAEREGDWSAMSGFQQTMQMLNEGIVPTA MLVANDQMAKGAMWAITESGLRVGADISVVGYDDTETSSQWIPPLTTI KQDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVKRKTTLALEHHH HHH
LacI-B3	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVV SMVERSGVEACKAAVHNLLAQRVSGLIINYPLDDQDAIAVEAACTNVP ALFLDVSDQTPINSIIFSHEDGTRLGVEHLVALGHQQIALLAGPLSSVSA RLRLAGWHKYLTRNQIQPIAEREGDWSAMSGFQQTMQMLNEGIVPTA MLVANDQMAKGAMKAITESGLRVGADISVVGYDDTESSSQWIPPLTTI KQDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVKRKTTLALEHHH HHH
LacI-splitGFP	MEFHMKPVTLYDVAEYAGVSTATVSNVVNQASHVSAKTREKVEAAMA ELNYIPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGAS VVVSMVERSGVEACKAAVHNLLAQRVSGLIINYPLDDQDAIAVEAACT NVPALFLDVSDQTPINSIIFSHEDGTRLGVEHLVALGHQQIALLAGPLSS VSARLRLAGWHKYLTRNQIQPIAEREGDWSAMSGFQQTMQMLNEGI VPTAMLVANDQMALGAMRAITESGLRVGADISVVGYDDTEDSSCSIPP LTTIKQDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVKRKTTLALEE LKLGGENLYFQGGGSTMGYTMDLPDNHYLSTQTILLKDLNGTGVGSG GGSHMGGGSGSGGGSGGGSTSEKRDHMVLLEYVTAAGITDASWSH PQFEK MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVV SMVERSGVEACKAAVHNLLAQRVSGLIINYPLDDQDAIAVEAACTNVP ALFLDVSDQTPINSIIFSHEDGTRLGVEHLVALGHQQIALLAGPLSSVSA RLRLAGWHKYLTRNQIQPIAEREGDWSAMSGFQQTMQMLNEGIVPTA MLVANDQMAKGAMWAITESGLRVGADISVVGYDDTETSSQWIPPLTTI KQDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVKRKTTLALELEG GENLYFQGGGSRKGEELFTGVVPILIELDGDVNGHKFFVRGEGEGDA TIGKLSLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFK SAMPEGYVQERTIYFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKED GNILGHKLEYNFNSHKVYITADKQNNGIKANFTIRHNVEDGSVQLADH YQQNTPIGDGPVLLPHHHHHH

Construct	AS Sequence
	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVV SMVERSGVEACKAAVHNLLAQRVSGLIINYPLDDQDAIAVEAACTNVP ALFLDVSDQTPINSIIFSHEDGTRLGVEHLVALGHQQIALLAGPLSSVSA RLRLAGWHKYLTRNQIQPIAEREGDWSAMSGFQQTMQMLNEGIVPTA MLVANDQMALGAMRAITESGLRVGADISVVGYDDTEDSSCYIPPLTTIK QDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVKRKTTLALEELKLG GENLYFQGGGSTMGYTMDLPDNHYLSTQTILLKDLNGTGVGSGGGS HMGGGSGSGGGSGGGSTSEKRDHMVLLEYVTAAGITDASWSHPQF EK
WT-splitGFP	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVV SMVERSGVEACKAAVHNLLAQRVSGLIINYPLDDQDAIAVEAACTNVP ALFLDVSDQTPINSIIFSHEDGTRLGVEHLVALGHQQIALLAGPLSSVSA RLRLAGWHKYLTRNQIQPIAEREGDWSAMSGFQQTMQMLNEGIVPTA MLVANDQMALGAMRAITESGLRVGADISVVGYDDTEDSSCYIPPLTTIK QDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVKRKTTLALEGGEN LYFQGGGSRKGEELFTGVVPILIELDGDVNGHKFFVRGEGEGDATIGK LSLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMP EGYVQERTIYFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILG HKLEYNFNSHKVYITADKQNNGIKANFTIRHNVEDGSVQLADHYQQNT PIGDGPVLLPHHHHHH
	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVV SMVERSGVEACKAAVHNLLAQRVSGLIINYPLDDQDAIAVEAACTNVP ALFLDVSDQTPINSIIFSHEDGTRLGVEHLVALGHQQIALLAGPLSSVSA RLRLAGWHKYLTRNQIQPIAEREGDWSAMSGFQQTMQMLNEGIVPTA MLVANDQMALGAMRAITESGLRVGADISVVGYDDTEDSSCAIPPLTTIK QDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVKRKTTLALEELKLG GENLYFQGGGSTMGYTMDLPDNHYLSTQTILLKDLNGTGVGSGGGS HMGGGSGSGGGSGGGSTSEKRDHMVLLEYVTAAGITDASWSHPQF EK
A1-splitGFP	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVV SMVERSGVEACKAAVHNLLAQRVSGLIINYPLDDQDAIAVEAACTNVP ALFLDVSDQTPINSIIFSHEDGTRLGVEHLVALGHQQIALLAGPLSSVSA RLRLAGWHKYLTRNQIQPIAEREGDWSAMSGFQQTMQMLNEGIVPTA MLVANDQMALGAMRAITESGLRVGADISVVGYDDTEDSSCAIPPLTTIK QDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVKRKTTLALEGGEN LYFQGGGSRKGEELFTGVVPILIELDGDVNGHKFFVRGEGEGDATIGK LSLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMP EGYVQERTIYFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILG HKLEYNFNSHKVYITADKQNNGIKANFTIRHNVEDGSVQLADHYQQNT PIGDGPVLLPHHHHHH

Construct	AS Sequence
	MLEDAKEIPYAELLGILSAQPTWDRSNGFHSVVDQYPEFKMVAQQSA EFDRDTAYKVTEQILQAHPEIKAIWCGNDAMALGAMKACEAAGRTDIYI FGFDGAEDVINAIKEGKQIVATIMQFPKLMARLAVEWADQYLRGERSF PEIVPVTVELVTRENIDKYTAYGRKELKLGGENLYFQGGGSTMGYTMD LPDNHYLSTQTILLKDLNGTGVGSGGGSHMGGGSGSGGGSGGGST SEKRDHMVLLEYVTAAGITDASWSHPQFEK
RBS-splitGFP	MLEKGKMAIVISTLNNPWFVVLAETAKQRAEQLGYEATIFDSQNDTAK ESAHFDAIIAAGYDAIIFNPTDADGSIANVKRAKEAGIPVFCVDRGINAR GLAVAQIYSDNYYGGVLMGEYFVKFLKEKDIGGENLYFQGGGSRKGE ELFTGVVPILIELDGDVNGHKFFVRGEGEGDATIGKLSLKFICTTGKLP VPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIYFK DDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHK VYITADKQNNGIKANFTIRHNVEDGSVQLADHYQQNTPIGDGPVLLPH HHHHH
LacI*-SpyCatcher	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVV SMVERSGVEACKAAVHNLLAQRVSGLIINYPLDDQDAIAVEAACTNVP ALFLDVSDQTPINSIIFSHEDGTRLGVEHLVALGHQQIALLAGPLSSVSA RLRLAGWHKYLTRNQIQPIAEREGDWSAMSGFQQTMQMLNEGIVPTA MLVANDQMALGAMRAITESGLRVGADISVVGYDDTEDSSCSIPPLTTIK QDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVKRKTTLALEGSGS GSGSDSATHIKFSKRDEDGKELAGATMELRDSSGKTISTWISDGQVK DFYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKATKGDAH ILEGSLEVLFQGPWSHPQFEK
LacI-A2-SpyTag	MKPVTLYDVAEYAGVSTATVSNVVNQASHVSAKTREKVEAAMAELNYI PNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVVS MVERSGVEACKAAVHNLLAQRVSGLIINYPLDDQDAIAVEAACTNVPA LFLDVSDQTPINSIIFSHEDGTRLGVEHLVALGHQQIALLAGPLSSVSAR LRLAGWHKYLTRNQIQPIAEREGDWSAMSGFQQTMQMLNEGIVPTA MLVANDQMALGAMRAITESGLRVGADISVVGYDDTEDSSCSIPPLTTIK QDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVKRKTTLALEGSGS GSGSAHIVMVDAYKPTLEGENLYFQGGHHHHHH
LacI-Lev1	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVV SEVSGRLAADREVLALTDEAARAAALAAVRAANTAACKAAVHNLLAQR VSGLIINYPLDDQDAIAVEAACTNVPALFLDVSDQTPINSIIFSHEDGTRL GVEHLVALGHQQIALLAGPLSSVSARLRLAGWHKYLTRNQIQPIAERE GDWSAMSGFQQTMQMLNEGIVPTAMLVANDQMALGAMRAITESGLR VGADISVVGYDDTEDSSCYIPPLTTIKQDFRLLGQTSVDRLLQLSQGQ AVKGNQLLPVSLVKRKTTLLEGGENLYFQGGGSRKGEELFTGVVPILIE LDGDVNGHKFFVRGEGEGDATIGKLSLKFICTTGKLPVPWPTLVTTLT YGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIYFKDDGTYKTRAE VKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHKVYITADKQNNG IKANFTIRHNVEDGSVQLADHYQQNTPIGDGPVLLPHHHHHH

Construct	AS Sequence
LacI-Lev2	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVV SMVERSGVEACKAAVHNLLAQRVSGLIINYPLDDQDAIAVEAACTNVP ALFLDVSDQTPINSIIFSHEDGTRLGVEHLVALGHQQIALLAGPLSSVSA RLRLAGWHKYLTRNQIQPIAEREGDWSAMSGFQQTMQMLNEGIVPTA MLVANDQMALGAMRAITESGLRVGADISVVGYDDTEDSSCYIPPLTTIK QDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVKRKTTLGADKAAR DAAAAKIRAAAAAAELAALEGGENLYFQGGGSRKGEELFTGVVPILIEL DGDVNGHKFFVRGEGEGDATIGKLSLKFICTTGKLPVPWPTLVTTLTY GVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIYFKDDGTYKTRAEV KFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHKVYITADKQNNGI KANFTIRHNVEDGSVQLADHYQQNTPIGDGPVLLPHHHHHH
LacI-Lev3	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVV SEVDFSVLVEMRKKGASLEELWEKKTEMVIAAVHNLLAQRVSGLIINYP LDDQDAIAVEAACTNVPALFLDVSDQTPINSIIFSHEDGTRLGVEHLVAL GHQQIALLAGPLSSVSARLRLAGWHKYLTRNQIQPIAEREGDWSAMS GFQQTMQMLNEGIVPTAMLVANDQMALGAMRAITESGLRVGADISVV GYDDTEDSSCYIPPLTTIKQDFRLLGQTSVDRLLQLSQGQAVKGNQLL PVSLVKRKTTLLEGGENLYFQGGGSRKGEELFTGVVPILIELDGDVNG HKFFVRGEGEGDATIGKLSLKFICTTGKLPVPWPTLVTTLTYGVQCFS RYPDHMKRHDFFKSAMPEGYVQERTIYFKDDGTYKTRAEVKFEGDTL VNRIELKGIDFKEDGNILGHKLEYNFNSHKVYITADKQNNGIKANFTIRH NVEDGSVQLADHYQQNTPIGDGPVLLPHHHHHH
LacI-Lev4	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVV SEAPAVVLAPTDTPEQIAAARAARTAACKAAVHNLLAQRVSGLIINYPL DDQDAIAVEAACTNVPALFLDVSDQTPINSIIFSHEDGTRLGVEHLVAL GHQQIALLAGPLSSVSARLRLAGWHKYLTRNQIQPIAEREGDWSAMS GFQQTMQMLNEGIVPTAMLVANDQMALGAMRAITESGLRVGADISVV GYDDTEDSSCYIPPLTTIKQDFRLLGQTSVDRLLQLSQGQAVKGNQLL PVSLVKRKTTLLEGGENLYFQGGGSRKGEELFTGVVPILIELDGDVNG HKFFVRGEGEGDATIGKLSLKFICTTGKLPVPWPTLVTTLTYGVQCFS RYPDHMKRHDFFKSAMPEGYVQERTIYFKDDGTYKTRAEVKFEGDTL VNRIELKGIDFKEDGNILGHKLEYNFNSHKVYITADKQNNGIKANFTIRH NVEDGSVQLADHYQQNTPIGDGPVLLPHHHHHH
LacI-Lev5	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVV SEVSGRLAADREVLALTDEAARAAALAAVRAANTAACKAAVHNLLAQR VSGLIINYPLDDQDAIAVEAACTNVPALFLDVSDQTPINSIIFSHEDGTRL GVEHLVALGHQQIALLAGPLSSVSARLRLAGWHKYLTRNQIQPIAERE GDWSAMSGFQQTMQMLNEGIVPTAMLVANDQMALGAMRAITESGLR VGADISVVGYDDTEDSSCYIPPLTTIKQDFRLLGQTSVDRLLQLSQGQ AVKGNQLLPVSLVKRKTTLLEGGENLYFQGGGSRKGEELFTGVVPILIE LDGDVNGHKFFVRGEGEGDATIGKLSLKFICTTGKLPVPWPTLVTTLT YGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIYFKDDGTYKTRAE VKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHKVYITADKQNNG IKANFTIRHNVEDGSVQLADHYQQNTPIGDGPVLLPHHHHHH

Construct	AS Sequence
LacI-Lev6	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVV SEVTVRTPFSTLERVTEREGGRLRTRFAGLEALRAEALAACKAAVHNL LAQRVSGLIINYPLDDQDAIAVEAACTNVPALFLDVSDQTPINSIIFSHED GTRLGVEHLVALGHQQIALLAGPLSSVSARLRLAGWHKYLTRNQIQPI AEREGDWSAMSGFQQTMQMLNEGIVPTAMLVANDQMALGAMRAITE SGLRVGADISVVGYDDTEDSSCYIPPLTTIKQDFRLLGQTSVDRLLQLS QGQAVKGNQLLPVSLVKRKTTLLEGGENLYFQGGGSRKGEELFTGVV PILIELDGDVNGHKFFVRGEGEGDATIGKLSLKFICTTGKLPVPWPTLV TTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIYFKDDGTYKT RAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHKVYITADKQ NNGIKANFTIRHNVEDGSVQLADHYQQNTPIGDGPVLLPHHHHHH
PBP1	MLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVVSEVSGRLAADREV LALTDEAARAAALAAVRAANTAACKAAVHNLLAQRVSGLIINYPLDDQD AIAVEAACTNVPALFLDVSDQTPINSIIFSHEDGTRLGVEHLVALGHQQI ALLAGPLSSVSARLRLAGWHKYLTRNQIQPIAEREGDWSAMSGFQQT MQMLNEGIVPTAMLVANDQMALGAMRAITESGLRVGADISVVGYDDT EDSSCYIPPLTTIKQDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVK RKTTLLEGGENLYFQGGGSRKGEELFTGVVPILIELDGDVNGHKFFVR GEGEGDATIGKLSLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHM KRHDFFKSAMPEGYVQERTIYFKDDGTYKTRAEVKFEGDTLVNRIELK GIDFKEDGNILGHKLEYNFNSHKVYITADKQNNGIKANFTIRHNVEDGS VQLADHYQQNTPIGDGPVLLPHHHHHH
PBP2	MLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVVSMVERSGVEACKA AVHNLLAQRVSGLIINYPLDDQDAIAVEAACTNVPALFLDVSDQTPINSII FSHEDGTRLGVEHLVALGHQQIALLAGPLSSVSARLRLAGWHKYLTRN QIQPIAEREGDWSAMSGFQQTMQMLNEGIVPTAMLVANDQMALGAM RAITESGLRVGADISVVGYDDTEDSSCYIPPLTTIKQDFRLLGQTSVDR LLQLSQGQAVKGNQLLPVSLVKRKTTLGADKAARDAAAAKIRAAAAAA ELAALEGGENLYFQGGGSRKGEELFTGVVPILIELDGDVNGHKFFVRG EGEGDATIGKLSLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMK RHDFFKSAMPEGYVQERTIYFKDDGTYKTRAEVKFEGDTLVNRIELKG IDFKEDGNILGHKLEYNFNSHKVYITADKQNNGIKANFTIRHNVEDGSV QLADHYQQNTPIGDGPVLLPHHHHHH
LacI-Link1	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVV SEVEFDRAGEVEMRKELEKLKKKDIPATYEFTDGKHYITNEPDERKKIE ERIAALEAAIREARTAACKAAVHNLLAQRVSGLIINYPLDDQDAIAVEAA CTNVPALFLDVSDQTPINSIIFSHEDGTRLGVEHLVALGHQQIALLAGPL SSVSARLRLAGWHKYLTRNQIQPIAEREGDWSAMSGFQQTMQMLNE GIVPTAMLVANDQMALGAMRAITESGLRVGADISVVGYDDTEDSSCYI PPLTTIKQDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVKRKTTLE NLYFQGGGSWSHPQFEK
LacI-Link2	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVV SEVEYDLAAVEENIRIRDELLRPDIPATYEFTDGKHYITNEPIPPKDPEE LLLLQRRLEEALAADRRRRTAACKAAVHNLLAQRVSGLIINYPLDDQDA IAVEAACTNVPALFLDVSDQTPINSIIFSHEDGTRLGVEHLVALGHQQIA LLAGPLSSVSARLRLAGWHKYLTRNQIQPIAEREGDWSAMSGFQQTM QMLNEGIVPTAMLVANDQMALGAMRAITESGLRVGADISVVGYDDTE DSSCYIPPLTTIKQDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVK RKTTLENLYFQGGGSWSHPQFEK

Construct	AS Sequence
LacI-Link3	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVV SEVAEDFAALDAIRGIRRELEALPEDIPATYEFTDGKHYITNEPVEEPAR LQARYEAALAAAEAARTAACKAAVHNLLAQRVSGLIINYPLDDQDAIAV EAACTNVPALFLDVSDQTPINSIIFSHEDGTRLGVEHLVALGHQQIALLA GPLSSVSARLRLAGWHKYLTRNQIQPIAEREGDWSAMSGFQQTMQM LNEGIVPTAMLVANDQMALGAMRAITESGLRVGADISVVGYDDTEDSS CYIPPLTTIKQDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVKRKTT LENLYFQGGGSWSHPQFEK
LacI-Link4	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVV SEVSGRLAADREVLALGSGSGSGSDIPATYEFTDGKHYITNEPIPPKG SGSGSGSEAARAAALAAVRAANTAACKAAVHNLLAQRVSGLIINYPLD DQDAIAVEAACTNVPALFLDVSDQTPINSIIFSHEDGTRLGVEHLVALG HQQIALLAGPLSSVSARLRLAGWHKYLTRNQIQPIAEREGDWSAMSG FQQTMQMLNEGIVPTAMLVANDQMALGAMRAITESGLRVGADISVVG YDDTEDSSCYIPPLTTIKQDFRLLGQTSVDRLLQLSQGQAVKGNQLLP VSLVKRKTTLLEENLYFQGGGSWSHPQFEK
LacI-Link5	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVV SEVSGRLAADREVLALEPPPPEPPPPDIPATYEFTDGKHYITNEPIPPK EPPPPEPPPPEAARAAALAAVRAANTAACKAAVHNLLAQRVSGLIINYP LDDQDAIAVEAACTNVPALFLDVSDQTPINSIIFSHEDGTRLGVEHLVAL GHQQIALLAGPLSSVSARLRLAGWHKYLTRNQIQPIAEREGDWSAMS GFQQTMQMLNEGIVPTAMLVANDQMALGAMRAITESGLRVGADISVV GYDDTEDSSCYIPPLTTIKQDFRLLGQTSVDRLLQLSQGQAVKGNQLL PVSLVKRKTTLLEENLYFQGGGSWSHPQFEK
LacI-A2-Link1-SpyTag	MKPVTLYDVAEYAGVSTATVSNVVNQASHVSAKTREKVEAAMAELNYI PNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVVS EVEFDRAGEVEMRKECEKLKKKDIPATYEFTDGKHYITNEPDERKKIE ERIAALEAAIREARTAACKAAVHNLLAQRVSGLIINYPLDDQDAIAVEAA CTNVPALFLDVSDQTPINSIIFSHEDGTRLGVEHLVALGHQQIALLAGPL SSVSARLRLAGWHKYLTRNQIQPIAEREGDWSAMSGFQQTMQMLNE GIVPTAMLVANDQMALGAMRAITESGLRVGADISVVGYDDTEDSSCSI PPLTTIKQDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVKRKTTLG SGSGSGSAHIVMVDAYKPTLEHHHHHH
LacI-A2-Link2-SpyTag	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVV SEVEYDLAAVEENIRIRDELLRPDIPATYEFTDGKHYITNEPIPPKDPEE LLLLQRRLEEALAADRRRRTAACKAAVHNLLAQRVSGLIINYPLDDQDA IAVEAACTNVPALFLDVSDQTPINSIIFSHEDGTRLGVEHLVALGHQQIA LLAGPLSSVSARLRLAGWHKYLTRNQIQPIAEREGDWSAMSGFQQTM QMLNEGIVPTAMLVANDQMALGAMRAITESGLRVGADISVVGYDDTE DSSCYIPPLTTIKQDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVK RKTTLGSGSGSGSAHIVMVDAYKPTLEHHHHHH

Construct	AS Sequence
LacI-Gentio	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVV SMVERSGVEACKAAVHNLLAQRVSGLIINYPLDDQDAIAVEAACTNVP ALFLDVSDQTPINSIIFSHEDGTRLGVEHLVALGHQQIALLAGPLSSVSA RLRLAGWHKYLTRNQIQPIAEREGDWSAMSGFQQTMQMLNEGIVPTA MLVANDQMALGAMRAITESGLRVGADISVVGYDDLEDGSCYIPPLTTIK HDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVKRKTTLALEENLYF QGGGSRKGEELFTGVVPILIELDGDVNGHKFFVRGEGEGDATIGKLSL KFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEG YVQERTIYFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHK LEYNFNSHKVYITADKQNNGIKANFTIRHNVEDGSVQLADHYQQNTPI GDGPVLLPHHHHHH
LLhT	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRAAQQLAGKQSKVVAIIVTRLDSLSENLAVQTMLPAFYEQGYDPIM MESQFSPQLVAEHLGVLKRRNIDGVVLFGFTGITEEMLAHWQSSLVLL ARDAKGFASVCYDDEGAIKILMQRLYDQGHRNISYLGVPHSDVTTGK RRHEAYLAFCKAHKLHPVAALPGLAMKQGYENVAKVITPETTALLCAT DTLALGASKYLQEQRIDTLQLASVGNTPLMKFLHPEIVTVDPGYAEAG RQAACQLIAQVTGRSEPQQIIIPATLSENLYFQGGGSWSHPQFEK
LLhF	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRVAQQLAGKSSRSIGLVIPDLENTSYTRIANYLERQARQRGYQLLI ACSEDQPDNEMRCIEHLLQRQVDAIIVSTSLPPEHPFYQRWANDPFPI VALDRALDREHFTSVVGADQDDAEMLAEELRKFPAETVLYLGALPELS VSFLREQGFRTAWKDDPREVHFLYANSYEREAAAQLFEKWLETHPMP QALFTTSFALLQGVMDVTLRRDGKLPSDLAIATFGDNELLDFLQCPVL AVAQRHRDVAERVLEIVLASLDEPRKPKPGLTRIKRNLYRRGVLSRSE NLYFQGGGSWSHPQFEK
LLhT-Lev1	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRAAQQLAGKQSKVVAIIVTRLDSLSENLAVQTMLPAFYEQGYDPIM MESQGEEGYGGFDLAAGREVLALTDEAARAAALAAVRAANTELAKNV QAILDEAKRRNIDGVVLFGFTGITEEMLAHWQSSLVLLARDAKGFASV CYDDEGAIKILMQRLYDQGHRNISYLGVPHSDVTTGKRRHEAYLAFCK AHKLHPVAALPGLAMKQGYENVAKVITPETTALLCATDTLALGASKYLQ EQRIDTLQLASVGNTPLMKFLHPEIVTVDPGYAEAGRQAACQLIAQVT GRSEPQQIIIPATLSLEHHHHHH
LLhT-Lev7	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRAAQQLAGKQSKVVAIIVTRLDSLSENLAVQTMLPAFYEQGYDPIM MEVSPEASIEERPQLVAEHLGVLKRRNIDGVVLFGFTGITEEMLAHWQ SSLVLLARDAKGFASVCYDDEGAIKILMQRLYDQGHRNISYLGVPHSD VTTGKRRHEAYLAFCKAHKLHPVAALPGLAMKQGYENVAKVITPETTA LLCATDTLALGASKYLQEQRIDTLQLASVGNTPLMKFLHPEIVTVDPGY AEAGRQAACQLIAQVTGRSEPQQIIIPATLSLEHHHHHH
LLhT-Lev8	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRAAQQLAGKQSKVVAIIVTRLDSLSENLAVQTMLPAFYEQGYDPIM MEAPVDTAKANIEEHLGVLKRRNIDGVVLFGFTGITEEMLAHWQSSLV LLARDAKGFASVCYDDEGAIKILMQRLYDQGHRNISYLGVPHSDVTTG KRRHEAYLAFCKAHKLHPVAALPGLAMKQGYENVAKVITPETTALLCA TDTLALGASKYLQEQRIDTLQLASVGNTPLMKFLHPEIVTVDPGYAEA GRQAACQLIAQVTGRSEPQQIIIPATLSLEHHHHHH

Construct	AS Sequence
LLhT-Link6	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRAAQQLAGKQSKVVAIIVTRLDSLSENLAVQTMLPAFYEQGYDPIM MEGSGSGSGSDSATHIKFSKRDEDGKELAGATMELRDSSGKTISTWI SDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKA TKGDAHILEGSGSGSGSFSPQLVAEHLGVLKRRNIDGVVLFGFTGITE EMLAHWQSSLVLLARDAKGFASVCYDDEGAIKILMQRLYDQGHRNISY LGVPHSDVTTGKRRHEAYLAFCKAHKLHPVAALPGLAMKQGYENVAK VITPETTALLCATDTLALGASKYLQEQRIDTLQLASVGNTPLMKFLHPEI VTVDPGYAEAGRQAACQLIAQVTGRSEPQQIIIPATLSLEHHHHHH
LLhT-Link7	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRAAQQLAGKQSKVVAIIVTRLDSLSENLAVQTMLPAFYEQGYDPIM MEVKWYPDKEYKATISCRVATAPDASGDSATHIKFSKRDEDGKELAGA TMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAIT FTVNEQGQVTVNGKATKGDAHILEKACKQASEGPQLVAEHLGVLKRR NIDGVVLFGFTGITEEMLAHWQSSLVLLARDAKGFASVCYDDEGAIKIL MQRLYDQGHRNISYLGVPHSDVTTGKRRHEAYLAFCKAHKLHPVAAL PGLAMKQGYENVAKVITPETTALLCATDTLALGASKYLQEQRIDTLQLA SVGNTPLMKFLHPEIVTVDPGYAEAGRQAACQLIAQVTGRSEPQQIIIP ATLSLEHHHHHH
LLhT-271A	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRAAQQLAGKQSKVVAIIVTRLDSLSENLAVQTMLPAFYEQGYDPIM MESQFSPQLVAEHLGVLKRRNIDGVVLFGFTGITEEMLAHWQSSLVLL ARDAKGFASVCYDDEGAIKILMQRLYDQGHRNISYLGVPHSDVTTGK RRHEAYLAFCKAHKLHPVAALPGLAMKQGYENVAKVITPETTALLCAT DTLALGASKYLQEQRIDTLQLASVGNTPLMKALHPEIVTVDPGYAEAG RQAACQLIAQVTGRSEPQQIIIPATLSENLYFQGGGSWSHPQFEK
LLhT-272A	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRAAQQLAGKQSKVVAIIVTRLDSLSENLAVQTMLPAFYEQGYDPIM MESQFSPQLVAEHLGVLKRRNIDGVVLFGFTGITEEMLAHWQSSLVLL ARDAKGFASVCYDDEGAIKILMQRLYDQGHRNISYLGVPHSDVTTGK RRHEAYLAFCKAHKLHPVAALPGLAMKQGYENVAKVITPETTALLCAT DTLALGASKYLQEQRIDTLQLASVGNTPLMKFAHPEIVTVDPGYAEAG RQAACQLIAQVTGRSEPQQIIIPATLSENLYFQGGGSWSHPQFEK
LLhT-272S	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRAAQQLAGKQSKVVAIIVTRLDSLSENLAVQTMLPAFYEQGYDPIM MESQFSPQLVAEHLGVLKRRNIDGVVLFGFTGITEEMLAHWQSSLVLL ARDAKGFASVCYDDEGAIKILMQRLYDQGHRNISYLGVPHSDVTTGK RRHEAYLAFCKAHKLHPVAALPGLAMKQGYENVAKVITPETTALLCAT DTLALGASKYLQEQRIDTLQLASVGNTPLMKFSHPEIVTVDPGYAEAG RQAACQLIAQVTGRSEPQQIIIPATLSENLYFQGGGSWSHPQFEK
LLhT-268A	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRAAQQLAGKQSKVVAIIVTRLDSLSENLAVQTMLPAFYEQGYDPIM MESQFSPQLVAEHLGVLKRRNIDGVVLFGFTGITEEMLAHWQSSLVLL ARDAKGFASVCYDDEGAIKILMQRLYDQGHRNISYLGVPHSDVTTGK RRHEAYLAFCKAHKLHPVAALPGLAMKQGYENVAKVITPETTALLCAT DTLALGASKYLQEQRIDTLQLASVGNTPAMKFLHPEIVTVDPGYAEAG RQAACQLIAQVTGRSEPQQIIIPATLSENLYFQGGGSWSHPQFEK

Construct	AS Sequence
LLhT-244A	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRAAQQLAGKQSKVVAIIVTRLDSLSENLAVQTMLPAFYEQGYDPIM MESQFSPQLVAEHLGVLKRRNIDGVVLFGFTGITEEMLAHWQSSLVLL ARDAKGFASVCYDDEGAIKILMQRLYDQGHRNISYLGVPHSDVTTGK RRHEAYLAFCKAHKLHPVAALPGLAMKQGYENVAKVITPETTALLCAT DTLAAGASKYLQEQRIDTLQLASVGNTPLMKFLHPEIVTVDPGYAEAG RQAACQLIAQVTGRSEPQQIIIPATLSENLYFQGGGSWSHPQFEK
LLhT-4Mut	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRAAQQLAGKQSKVVAIIVTRLDSLSENLAVQTMLPAFYEQGYDPIM MESQFSPQLVAEHLGVLKRRNIDGVVLFGFTGITEEMLAHWQSSLVLL ARDAKGFASVCYDDEGAIKILMQRLYDQGHRNISYLGVPHSDVTTGK RRHEAYLAFCKAHKLHPVAALPGLAMKQGYENVAKVITPETTALLCAT DTLAAGASKYLQEQRIDTLQLASVGNTPAMKAAHPEIVTVDPGYAEAG RQAACQLIAQVTGRSEPQQIIIPATLSENLYFQGGGSWSHPQFEK
LLhT-6Mut	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRAAQQLAGKQSKVVAIIVTRLDSLSENLAVQTMLPAFYEQGYDPIM MESQFSPQLVAEHLGVLKRRAADGVVLFGFTGITEEMLAHWQSSLVL LARDAKGFASVCYDDEGAIKILMQRLYDQGHRNISYLGVPHSDVTTGK RRHEAYLAFCKAHKLHPVAALPGLAAAQGYENVAKVITPETTALLCATD TLAAGASKYLQEQRIDTLQLASVGNTPAMKAAHPEIVTVDPGYAEAGR QAACQLIAQVTGRSEPQQIIIPATLSENLYFQGGGSWSHPQFEK-
LLhT-6Mut-Lev7-SpyCatcher-DogTag	MKPVTLYDVAEYAGVSTATVSNVVNQASHVSAKTREKVEAAMAELNYI PNRAAQQLAGKQSKVVAIIVTRLDSLSENLAVQTMLPAFYEQGYDPIM MEVSPEDSATHIKFSKRDEDGKELAGATMELRDSSGKTISTWISDGQV KDFYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKATKGDA HIASIEERPQLVAEHLGVLKRRNIDGVVLFGFTGITEEMLAHWQSSLVL LARDAKGFASVCYDDEGAIKILMQRLYDQGHRNISYLGVPHSDVTTGK RRHEAYLAFCKAHKLHPVAALPGLAAAQGYENVAKVITPETTALLCATD TLALGASAYLQEQRIDTLQLASVGNTPAMKAAHPEIVTVDPGYAEAGR QAACQLIAQVTGRSEPQQIIIPATLSGSGSGSGSDIPATYEFTDGKHYIT NEPIPPKHHHHHH
LLhT-6Mut-DogCatcher	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRAAQQLAGKQSKVVAIIVTRLDSLSENLAVQTMLPAFYEQGYDPIM MESQFSPQLVAEHLGVLKRRAADGVVLFGFTGITEEMLAHWQSSLVL LARDAKGFASVCYDDEGAIKILMQRLYDQGHRNISYLGVPHSDVTTGK RRHEAYLAFCKAHKLHPVAALPGLAAAQGYENVAKVITPETTALLCATD TLAAGASKYLQEQRIDTLQLASVGNTPAMKAAHPEIVTVDPGYAEAGR QAACQLIAQVTGRSEPQQIIIPATLSGSGSGSGSKLGEIEFIKVDKTDKK PLRGAVFSLQKQHPDYPDIYGAIDQNGTYQDVRTGEDGKLTFTNLSD GKYRLIENSEPPGYKPVQNKPIVSFRIVDGEVRDVTSIVPQLEHHHHH H
LacI-A2-Link1-SpyCatcher-DogTag	MKPVTLYDVAEYAGVSTATVSNVVNQASHVSAKTREKVEAAMAELNYI PNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVVS EVEFDRAGEVEMRKECEKLKKKGSDSATHIKFSKRDEDGKELAGATM ELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFT VNEQGQVTVNGKATKGDAHILEGSDERKKIEERIAALEAAIREARTAAA KAAVHNLLAQRVSGLIINYPLDDQDAIAVEAAATNVPALFLDVSDQTPIN SIIFSHEDGTRLGVEHLVALGHQQIALLAGPLSSVSARLRLAGWHKYLT RNQIQPIAEREGDWSAMSGFQQTMQMLNEGIVPTAMLVANDQMALG AMRAITESGLRVGADISVVGYDDTEDSSCSIPPLTTIKQDFRLLGQTSV DRLLQLSQGQAVKGNQLLPVSLVKRKTTLGSGSGSGSDIPATYEFTDG KHYITNEPIPPKLEHHHHHH

Construct	AS Sequence
LacI*-DogCatcher	MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAELNY IPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVV SMVERSGVEACKAAVHNLLAQRVSGLIINYPLDDQDAIAVEAACTNVP ALFLDVSDQTPINSIIFSHEDGTRLGVEHLVALGHQQIALLAGPLSSVSA RLRLAGWHKYLTRNQIQPIAEREGDWSAMSGFQQTMQMLNEGIVPTA MLVANDQMALGAMRAITESGLRVGADISVVGYDDEDSSCSIPPLTTIK QDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVKRKTTLALEGSGS GSGSKLGEIEFIKVDKTDKKPLRGAVFSLQKQHPDYPDIYGAIDQNGT YQDVRTGEDGKLTFTNLSDGKYRLIENSEPPGYKPVQNKPIVSFRIVD GEVRDVTSIVPQLEHHHHHH
CM-PotF	MKLGEIEFIKVDKTDKKPLRGAVFSLQKQHPDYPDIYGAIDQNGTYQD VRTGEDGKLTFTNLSDGKYRLIENSEPPGYKPVQNKPIVSFRIVDGEV RDVTSIVPQGSGSGSGSGSAHIVMVDAYKPTLEAEQKTLHIYNWSDYI APDTVANFEKETGIKVVYDVFDSNEVLEGKLMAGSTGFDLVVPSASFL ERQLTAGVFQPLDKSKLPEWKNLDPELLKLVAKHDPDNKFAMPYMWA TTGIGYNVDKVKAVLGENAPVDSWDLILKPENLEKLKSCGVSFLDAPE EVFATVLNYLGKDPNSTKADDYTGPATDLLLKLRPNIRYFHSSQYINDL ANGDICVAIGWAGDVWQASNRAKEAKNGVNVSFSIPKEGAMAFFDVF AMPADAKNKDEAYQFLNYLLRPDVVAHISDHVFYANANKAATPLVSAE VRENPGIYPPADVRAKLFTLKVQDPKIDRVRTRAWTKVKSGKLEGSGS GSGSAHIVMVDAYKPTLEHHHHHH

7.2 Additional purification data: chromatograms and SDS-PAGE analyses

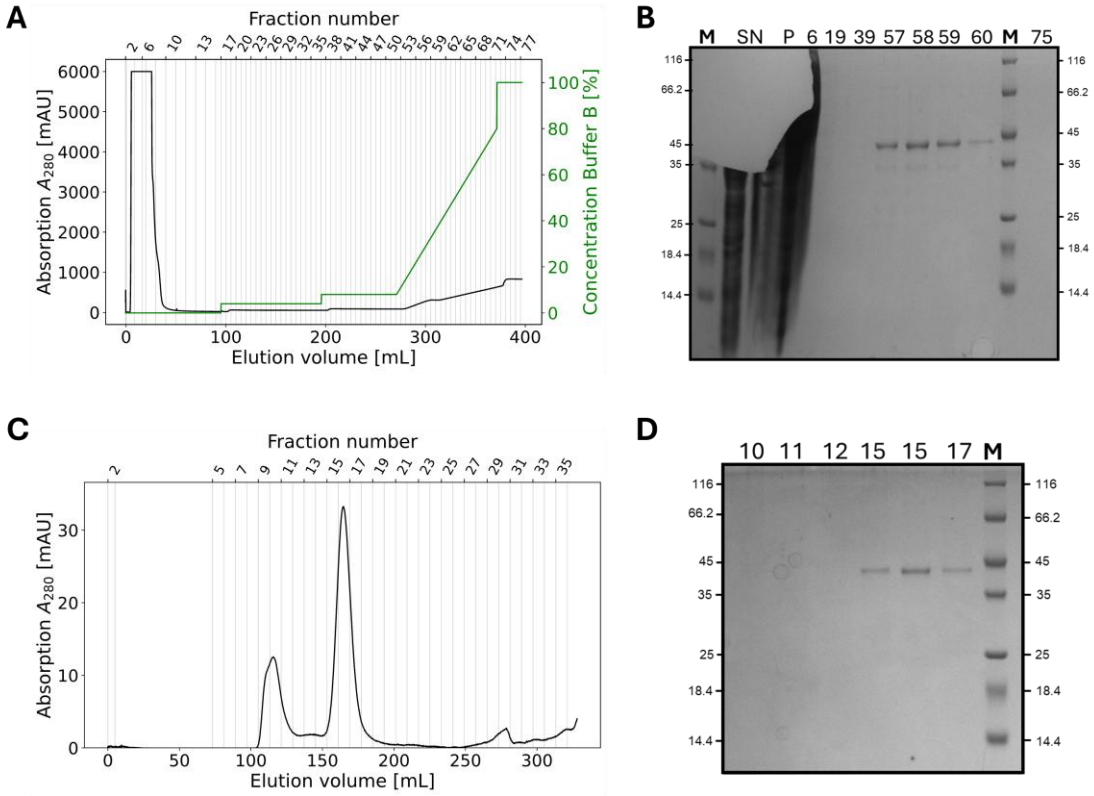


Figure 88. Purification of LacI-A1 by Ni^{2+}-IMAC and SEC. A) Ni^{2+}-IMAC chromatogram of the soluble fraction after expression of LacI-A1 (MW: 40 kDa). B) SDS-PAGE analysis of selected Ni^{2+}-IMAC fractions from panel A. C) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. D) SDS-PAGE analysis of selected SEC fractions from panel C. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

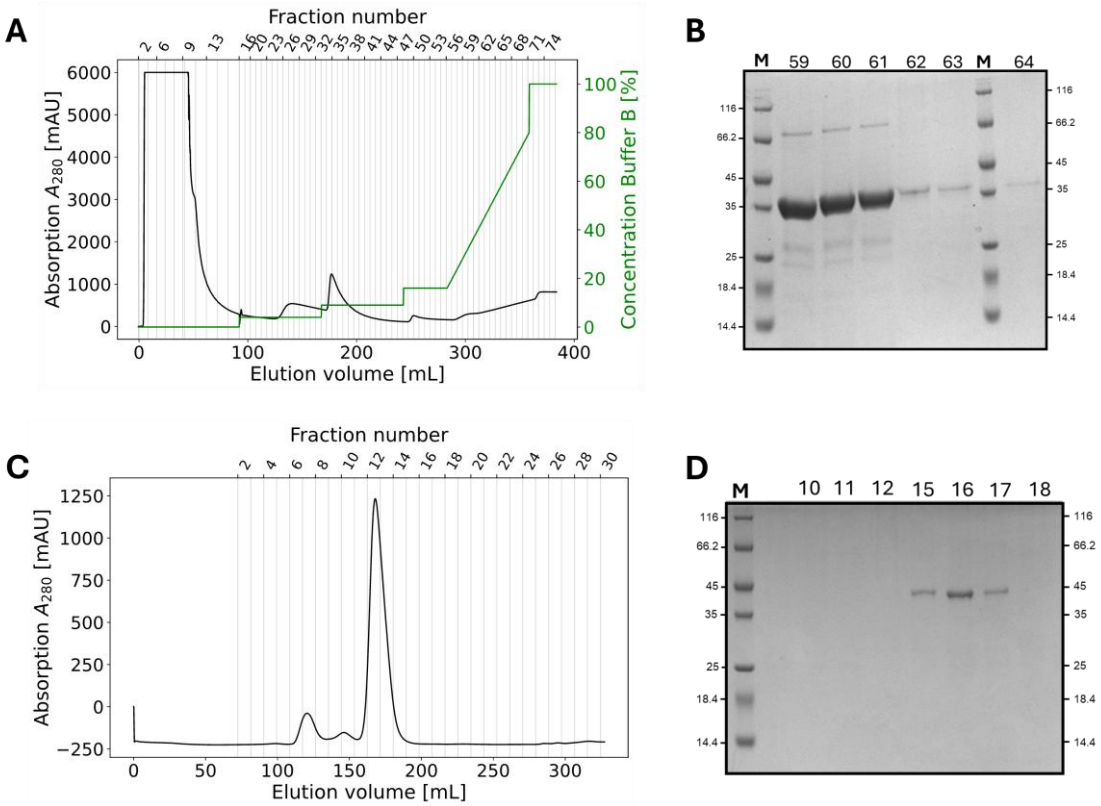


Figure 89. Purification of LacI-B1 by Ni^{2+}-IMAC and SEC. A) Ni^{2+}-IMAC chromatogram of the soluble fraction after expression of LacI-B1 (MW: 36 kDa). B) SDS-PAGE analysis of selected Ni^{2+}-IMAC fractions from panel A. C) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. D) SDS-PAGE analysis of selected SEC fractions from panel C. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

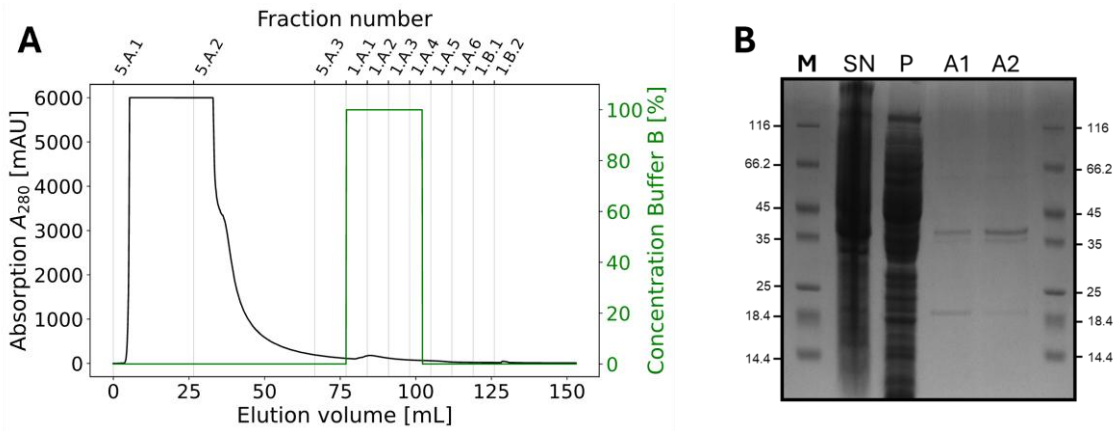


Figure 90. Purification of LacI-A2 by Strep-Tactin AC. A) Strep-Tactin AC chromatogram of the soluble fraction after expression of LacI-A2 (MW: 37 kDa). B) SDS-PAGE analysis of selected Strep-Tactin AC fractions from panel A, showing enrichment of LacI-A2 during elution with 100% buffer B. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

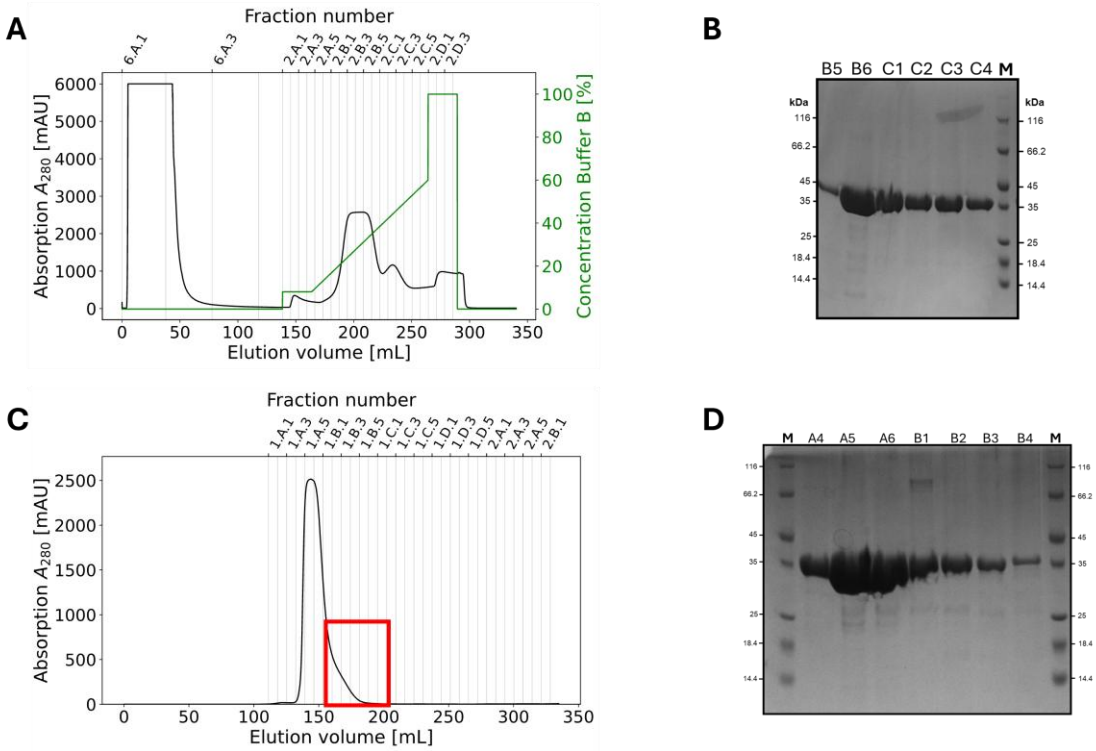


Figure 91. Purification of LacI-B2 by Ni^{2+}-IMAC and SEC. A) Ni^{2+}-IMAC chromatogram of the soluble fraction after expression of LacI-B2 (MW: 36 kDa). B) SDS-PAGE analysis of selected Ni^{2+}-IMAC fractions from panel A. C) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. D) SDS-PAGE analysis of selected SEC fractions from panel C. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

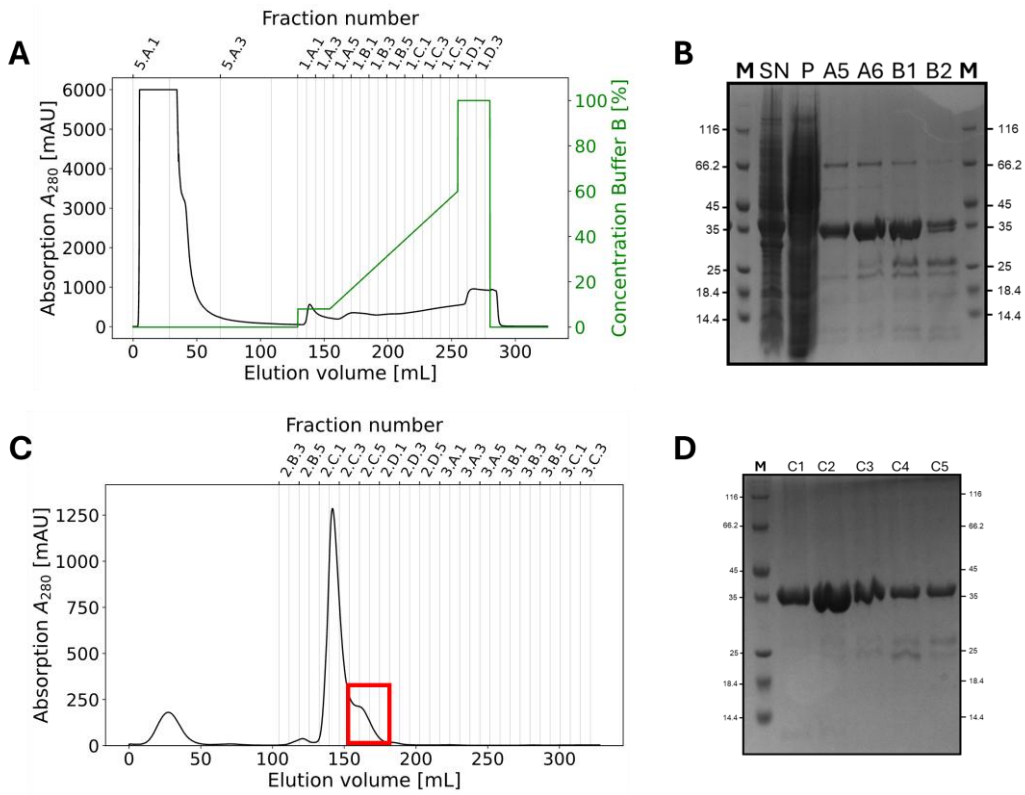


Figure 92. Purification of LacI-B3 by Ni^{2+}-IMAC and SEC. A) Ni^{2+}-IMAC chromatogram of the soluble fraction after expression of LacI-B3 (MW: 36 kDa). B) SDS-PAGE analysis of selected Ni^{2+}-IMAC fractions from panel A. C) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. D) SDS-PAGE analysis of selected SEC fractions from panel C. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

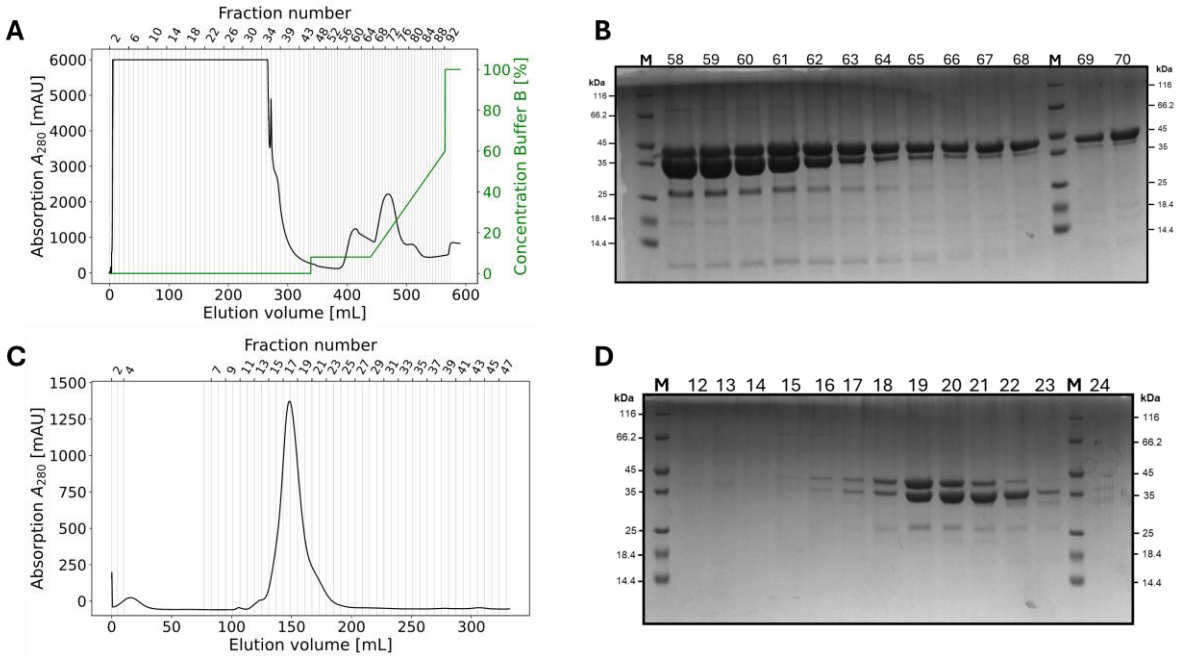


Figure 93. Purification analysis of co-expressed LacI-A1 and LacI-B1 by Ni^{2+}-IMAC and SEC. A) Ni^{2+}-IMAC chromatogram of the soluble fraction after co-expression of LacI-A1 (MW: 40 kDa) and LacI-B1 (MW: 36 kDa). B) SDS-PAGE analysis of selected Ni^{2+}-IMAC fractions from panel A. C) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. D) SDS-PAGE analysis of selected SEC fractions from panel C. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

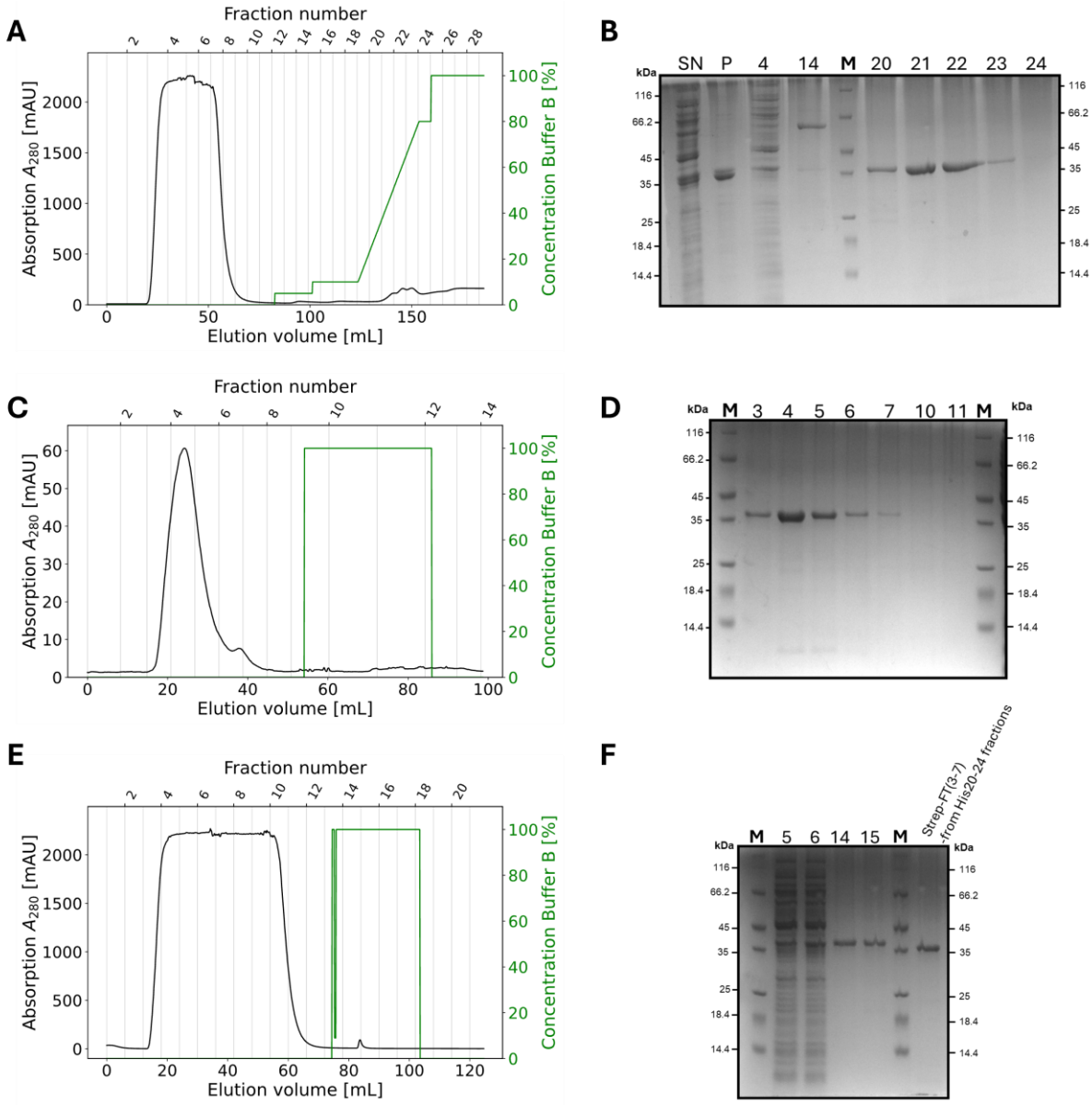


Figure 94. Purification analysis of co-expressed LacI-A2 and LacI-B2 by Ni^{2+}-IMAC and Strep-Tactin AC. A) Ni^{2+}-IMAC chromatogram of the soluble fraction after co-expression of LacI-A2 (MW: 37 kDa) and LacI-B2 (MW: 36 kDa). B) SDS-PAGE analysis of selected Ni^{2+}-IMAC fractions from panel A, showing enrichment of the His-tagged target protein during elution with increasing buffer B concentration. C) Strep-Tactin AC chromatogram of the Ni^{2+}-IMAC elution fractions. D) SDS-PAGE analysis of selected fractions from panel C, showing that only LacI-A2 was detected in the elution fractions after Ni^{2+}-IMAC. E) Strep-Tactin AC chromatogram of the Ni^{2+}-IMAC flow-through. F) SDS-PAGE analysis of selected fractions from panel E, showing that only LacI-B2 was detected in the Ni^{2+}-IMAC flow-through. The separation of LacI-A2 and LacI-B2 across the purification steps indicates that no stable co-purifying heterodimer was detected under the tested conditions. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

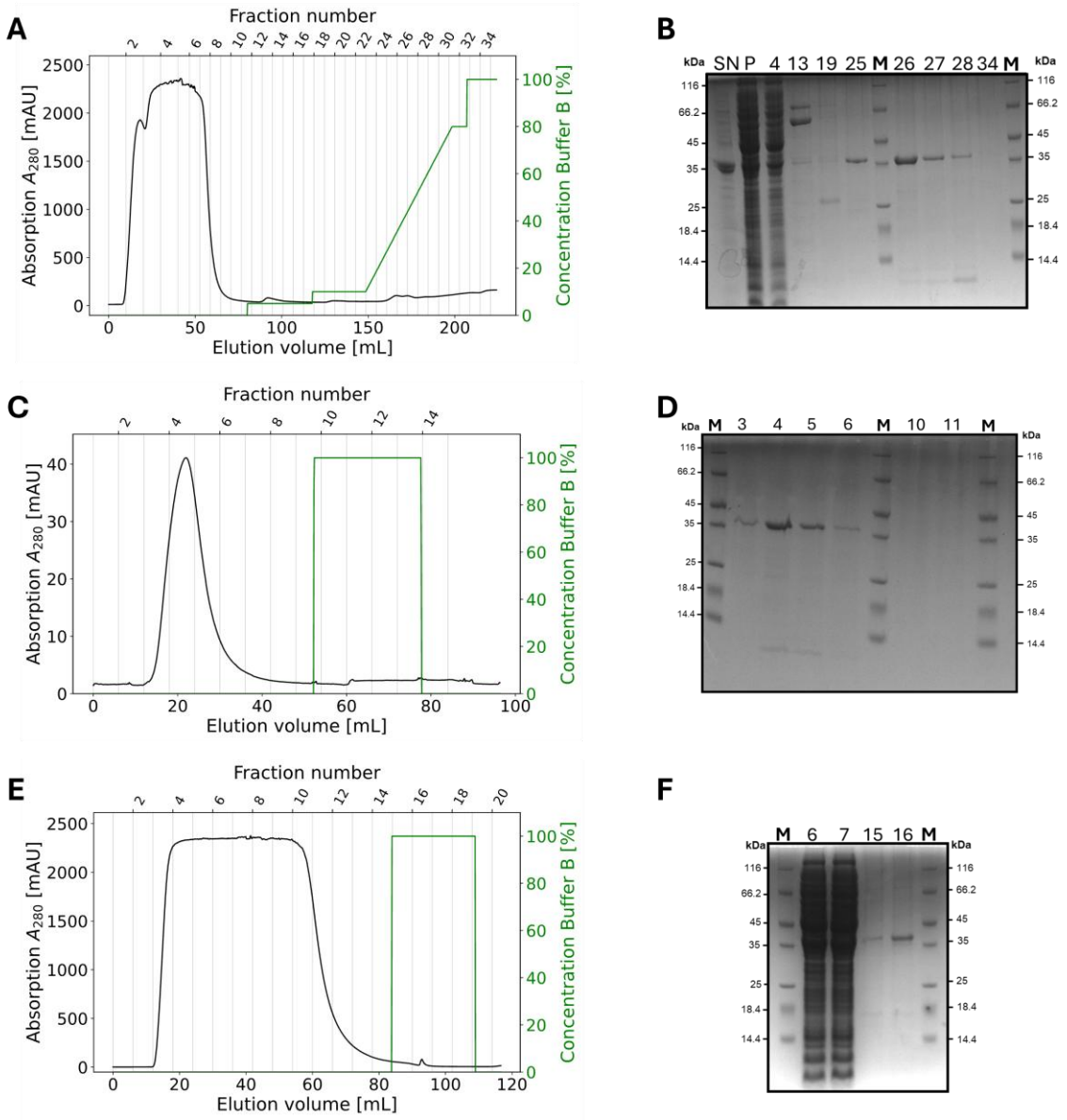


Figure 95. Purification analysis of co-expressed LacI-A2 and LacI-B3 by Ni^{2+}-IMAC and Strep-Tactin AC. A) Ni^{2+}-IMAC chromatogram of the soluble fraction after co-expression of LacI-A2 (MW: 37 kDa) and LacI-B3 (MW: 36 kDa). B) SDS-PAGE analysis of selected Ni^{2+}-IMAC fractions from panel A, showing enrichment of the His-tagged target protein during elution with increasing buffer B concentration. C) Strep-Tactin AC chromatogram of the Ni^{2+}-IMAC elution fractions. D) SDS-PAGE analysis of selected fractions from panel C, showing that only LacI-A2 was detected in the elution fractions after Ni^{2+}-IMAC. E) Strep-Tactin AC chromatogram of the Ni^{2+}-IMAC flow-through. F) SDS-PAGE analysis of selected fractions from panel E, showing that only LacI-B3 was detected in the Ni^{2+}-IMAC flow-through. The separation of LacI-A2 and LacI-B3 across the purification steps indicates that no stable co-purifying heterodimer was detected under the tested conditions. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

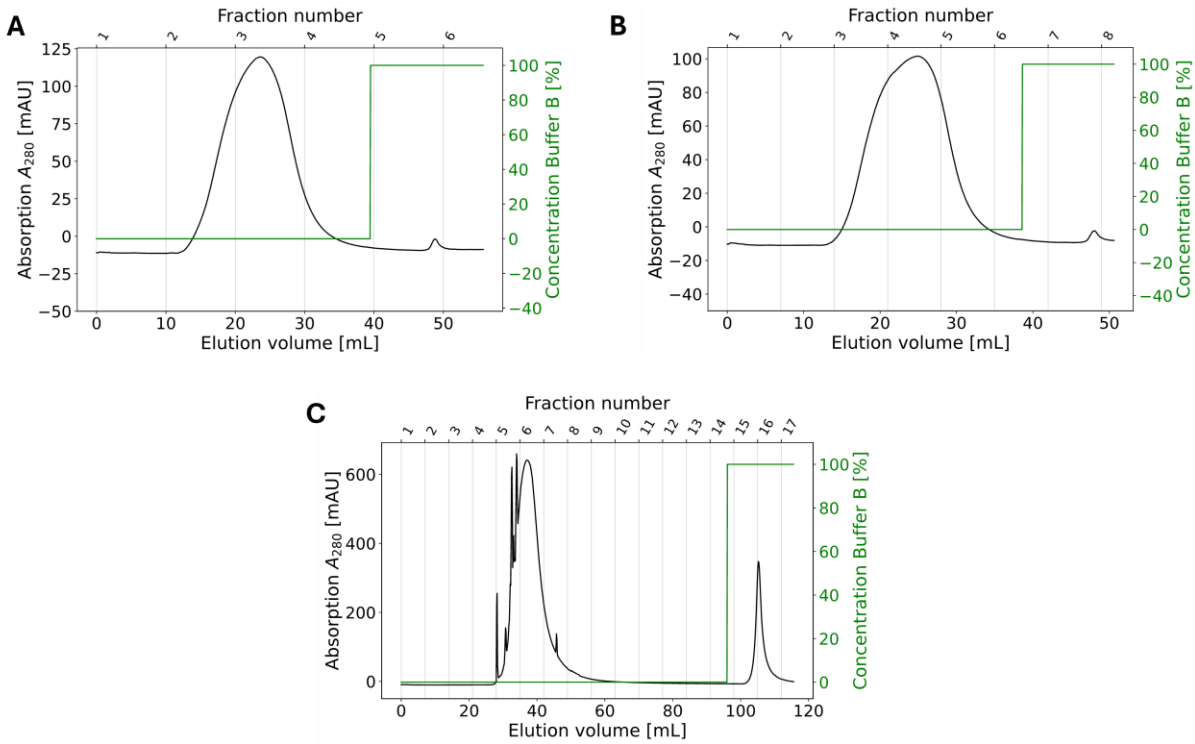


Figure 96. Chromatograms of splitGFP fusion constructs used as controls for the design of a BiFC assay. The corresponding SDS-PAGE analyses are shown in Figure 38. A) Strep-Tactin AC of A1-splitGFP. B) Strep-Tactin AC of RBP-splitGFP. C) Strep-Tactin AC of WT-splitGFP. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. The buffer B concentration is shown on the right y-axis in green.

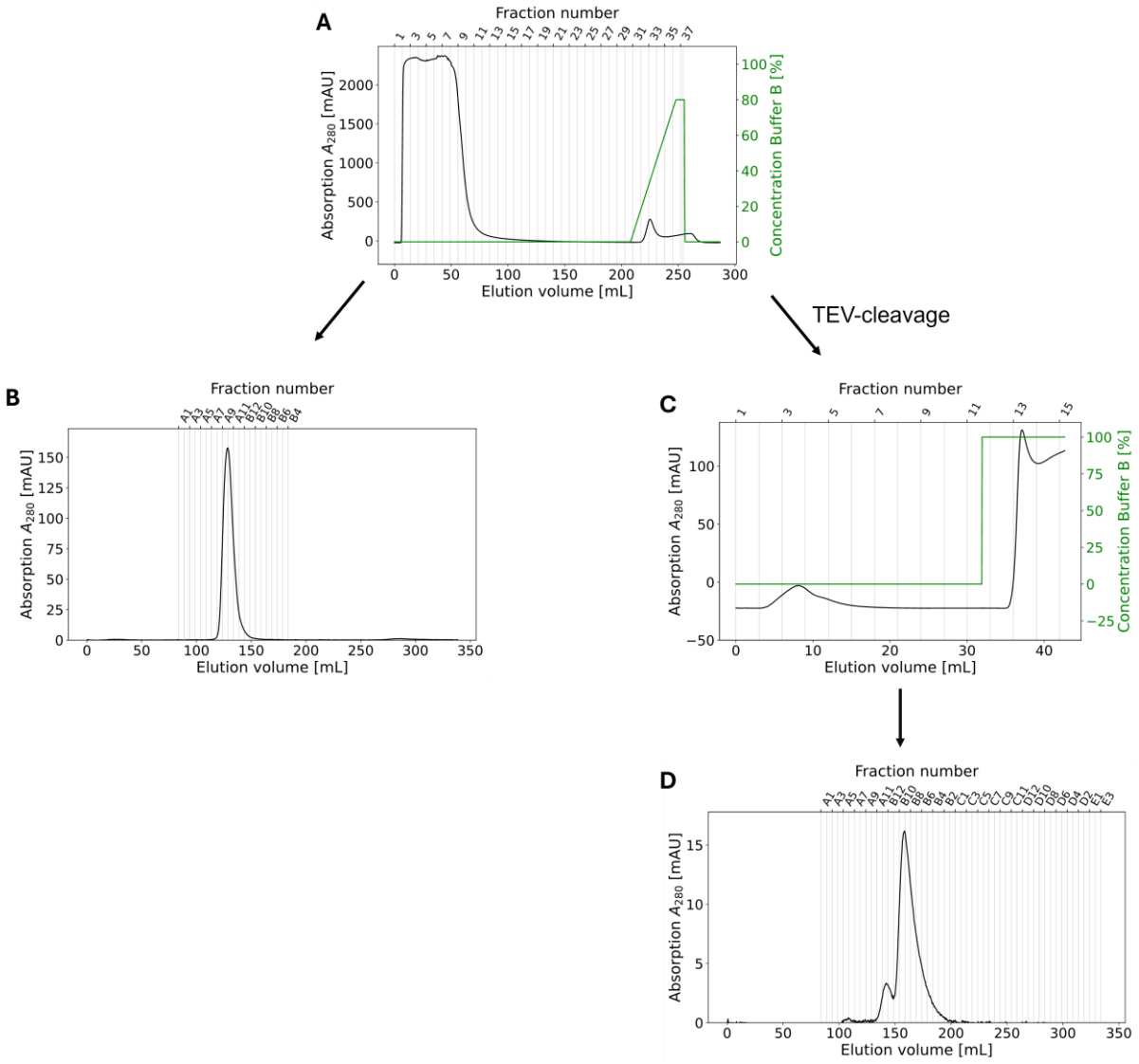


Figure 97. Purification and proteolytic processing of LacI-splitGFP. A) Ni^{2+}-IMAC chromatogram of LacI-splitGFP after recombinant expression. B) SEC profile of the purified uncleaved LacI-splitGFP sample, showing a single major peak. SEC was performed using a HiLoad 26/600 Superdex 75 pg column. C) Reverse Ni^{2+}-IMAC chromatogram after TEV cleavage, used to separate the cleaved sample from His_6-tagged components. D) SEC profile of the cleaved LacI-splitGFP sample after reverse Ni^{2+}-IMAC, performed using a HiLoad 26/600 Superdex 75 pg column. In contrast to the uncleaved sample shown in panel B, the cleaved construct no longer elutes as a single apparently monodisperse species but separates into two distinct peaks, consistent with different oligomeric or conformational states.Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

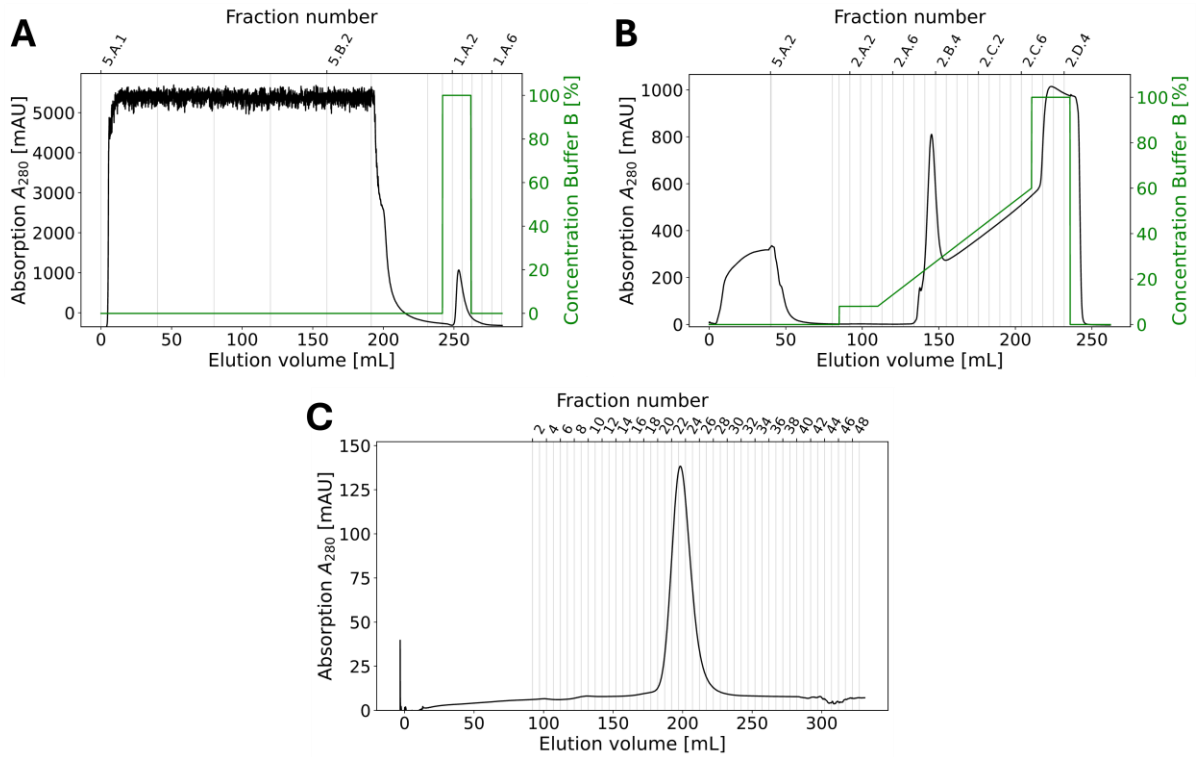


Figure 98. Chromatograms of the three-step purification of Het-SpyTagxCatcher. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green. A) Chromatogram of the Strep-Tactin affinity chromatography. B) Chromatogram of Ni2+-IMAC. C) SEC chromatogram of the Strep-Tactin elution fraction, performed using a HiLoad 26/600 Superdex 200 pg column.

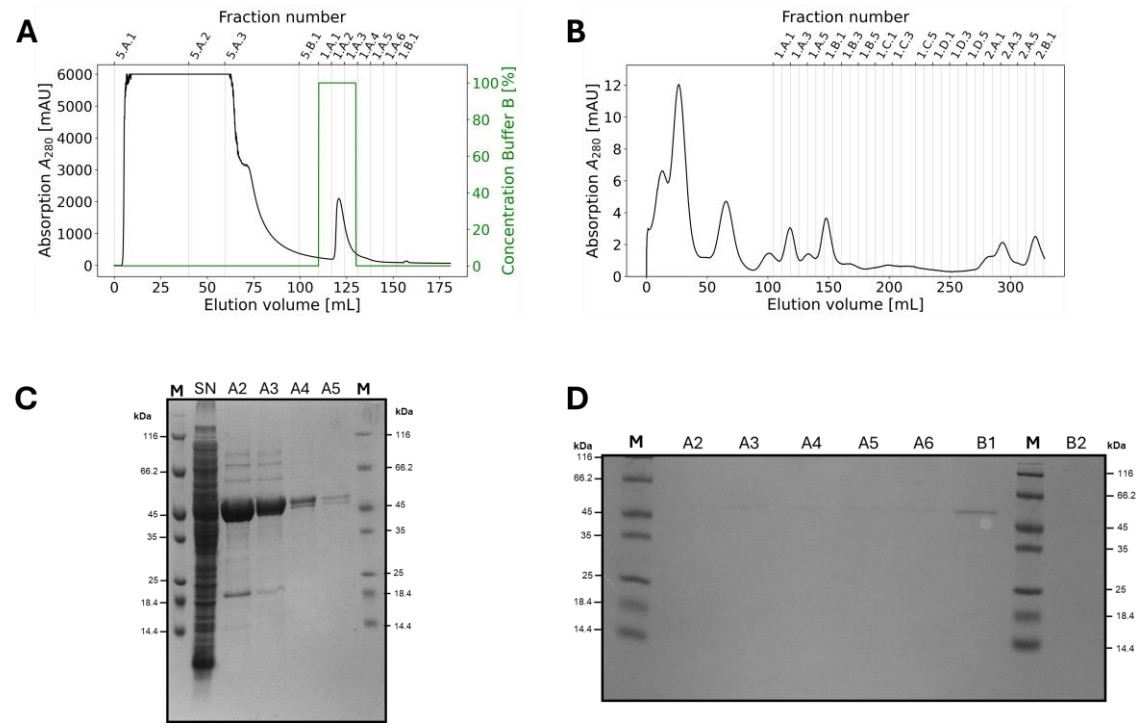


Figure 99. Purification of LacI*-SpyCatcher by Strep-Tactin AC and SEC. A) Strep-Tactin AC chromatogram of the soluble fraction after recombinant expression of LacI*-SpyCatcher (MW: 48 kDa). B) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. C, D) SDS-PAGE analyses of selected fractions from panels A and B, confirming enrichment of the target protein in the major elution peaks. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

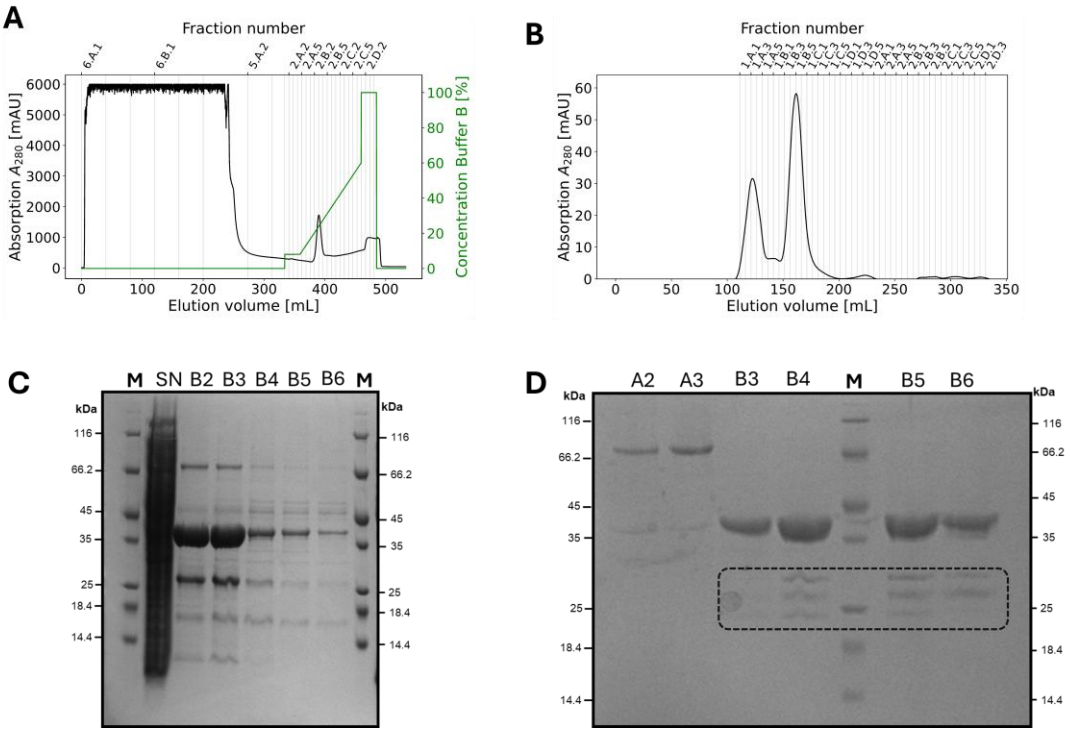


Figure 100. Purification of LacI-A2-SpyTag by Ni^{2+}-IMAC and SEC. A) Ni^{2+}-IMAC chromatogram of the soluble fraction after recombinant expression of LacI-A2-SpyTag (MW: 39 kDa), with the corresponding SDS-PAGE analysis shown below. B) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. C, D) SDS-PAGE analyses of selected Ni^{2+}-IMAC and SEC fractions, confirming enrichment of the target protein in the major elution peaks. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

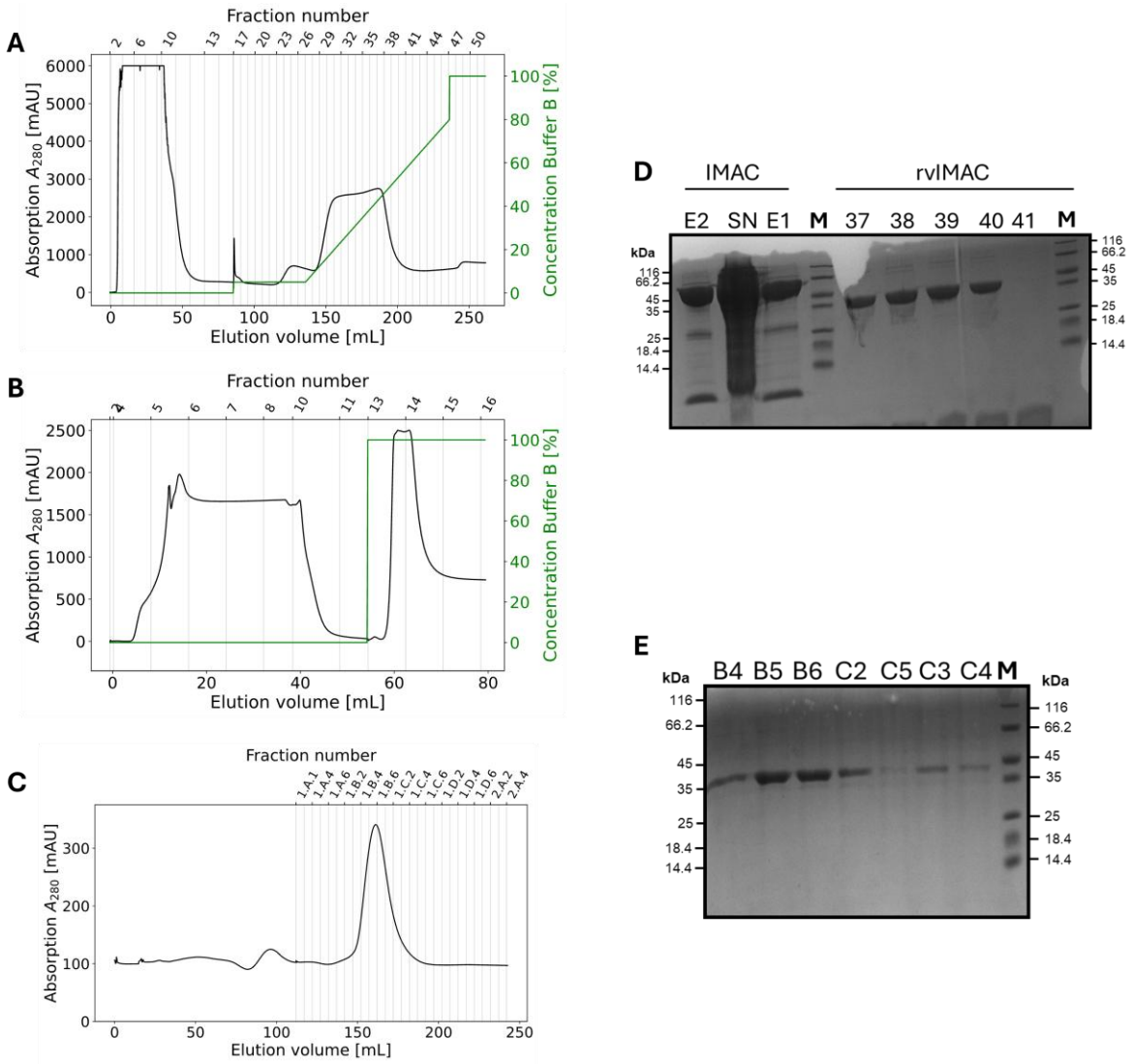


Figure 101. Purification of LacI-Lev1 by Ni^{2+}-IMAC, reverse Ni^{2+}-IMAC after TEV cleavage and SEC. A) Ni^{2+}-IMAC chromatogram of the soluble fraction after recombinant expression of His_6-tagged LacI-Lev1 (MW: 62 kDa), with the corresponding SDS-PAGE analysis shown below. B) Reverse Ni^{2+}-IMAC chromatogram after TEV cleavage, used to separate cleaved LacI-Lev1 (MW: 39 kDa) from His_6-tagged components. C) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. D, E) SDS-PAGE analyses of selected fractions from panels B and C, confirming the presence of cleaved LacI-Lev1 in the reverse Ni^{2+}-IMAC flow-through and in the major SEC elution peak. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

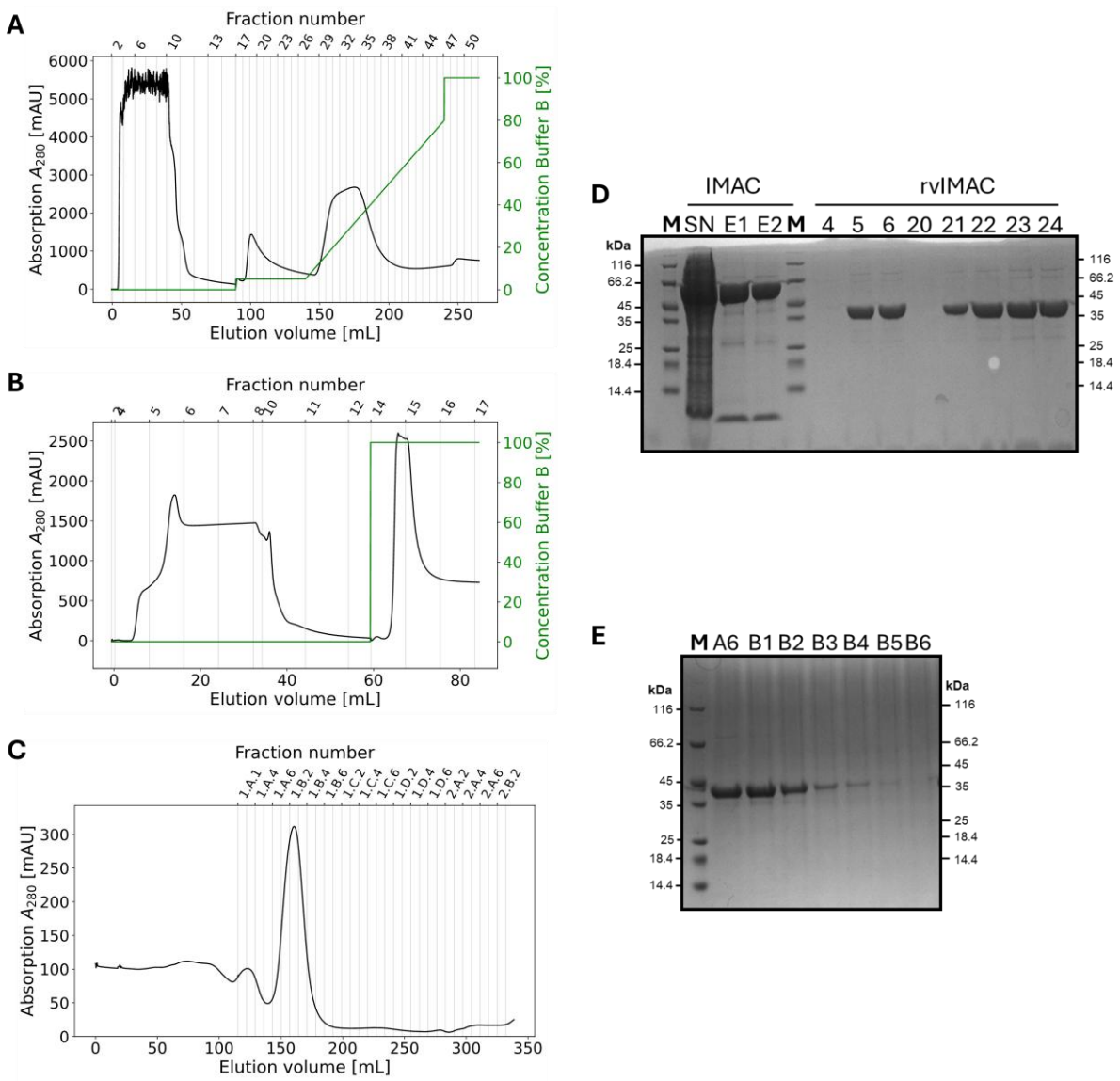


Figure 102. Purification of LacI-Lev2 by Ni^{2+}-IMAC, reverse Ni^{2+}-IMAC after TEV cleavage and SEC. A) Ni^{2+}-IMAC chromatogram of the soluble fraction after recombinant expression of His_6-tagged LacI-Lev1 (MW: 62 kDa), with the corresponding SDS-PAGE analysis shown below. B) Reverse Ni^{2+}-IMAC chromatogram after TEV cleavage, used to separate cleaved LacI-Lev2 (MW: 39 kDa) from His_6-tagged components. C) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. D, E) SDS-PAGE analyses of selected fractions from panels B and C, confirming the presence of cleaved LacI-Lev1 in the reverse Ni^{2+}-IMAC flow-through and in the major SEC elution peak. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

.

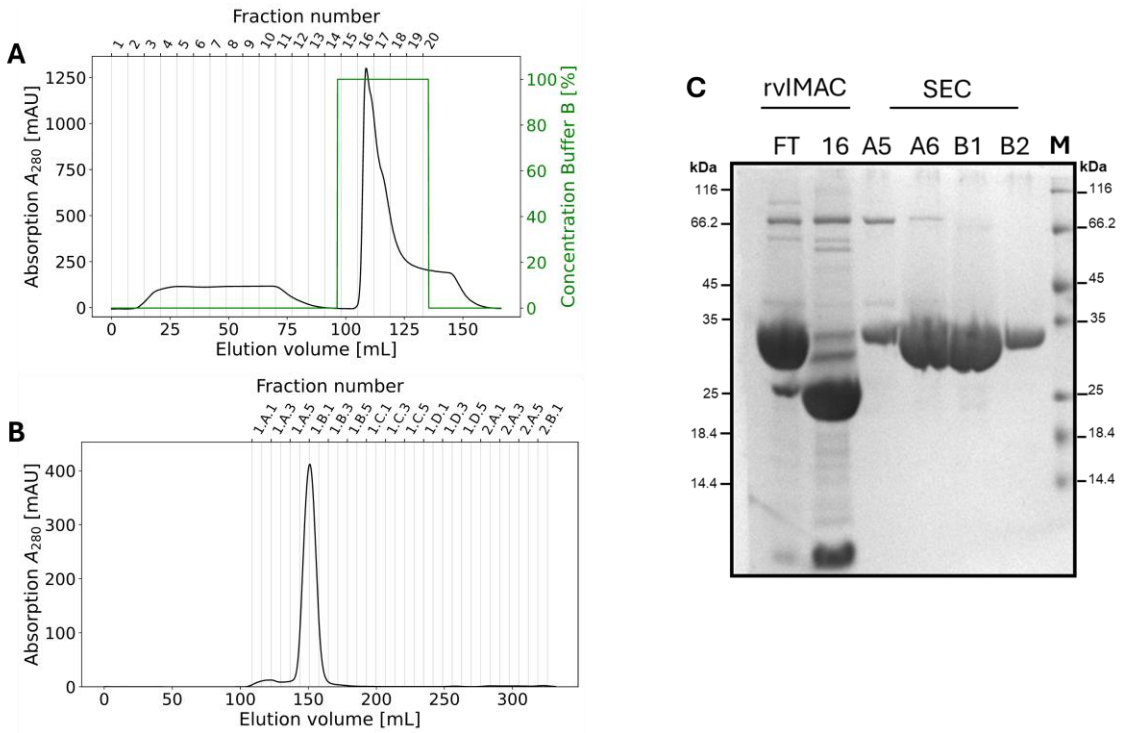


Figure 103. Purification of PBP1 by reverse Ni^{2+}-IMAC after TEV cleavage and SEC. A) Reverse Ni^{2+}-IMAC chromatogram after TEV cleavage, used to separate cleaved PBP1 (MW: 32 kDa) from His_6-tagged cleaved part. B) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. C) SDS-PAGE analysis of selected fractions from panels A and B, confirming the presence of PBP1 in the reverse Ni^{2+}-IMAC flow-through and in the major SEC elution peak. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

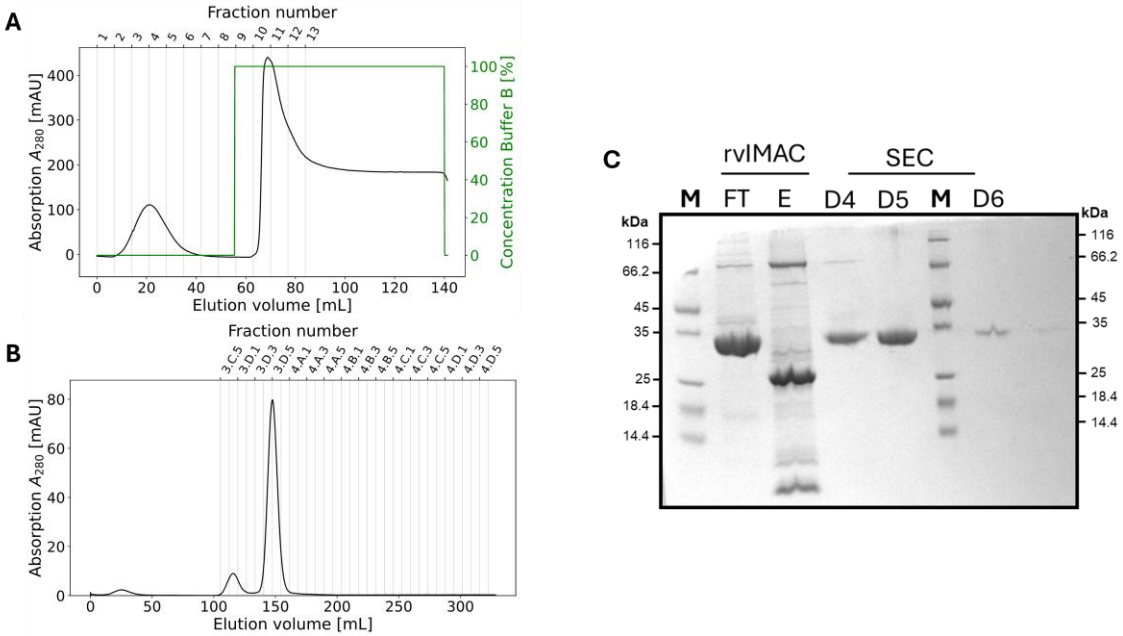


Figure 104. Purification of PBP2 by reverse Ni^{2+}-IMAC after TEV cleavage and SEC. A) Reverse Ni^{2+}-IMAC chromatogram after TEV cleavage, used to separate cleaved PBP2 (MW: 32 kDa) from His_6-tagged cleaved part. B) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. C) SDS-PAGE analysis of selected fractions from panels A and B, confirming the presence of PBP2 in the reverse Ni^{2+}-IMAC flow-through and in the major SEC elution peak. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

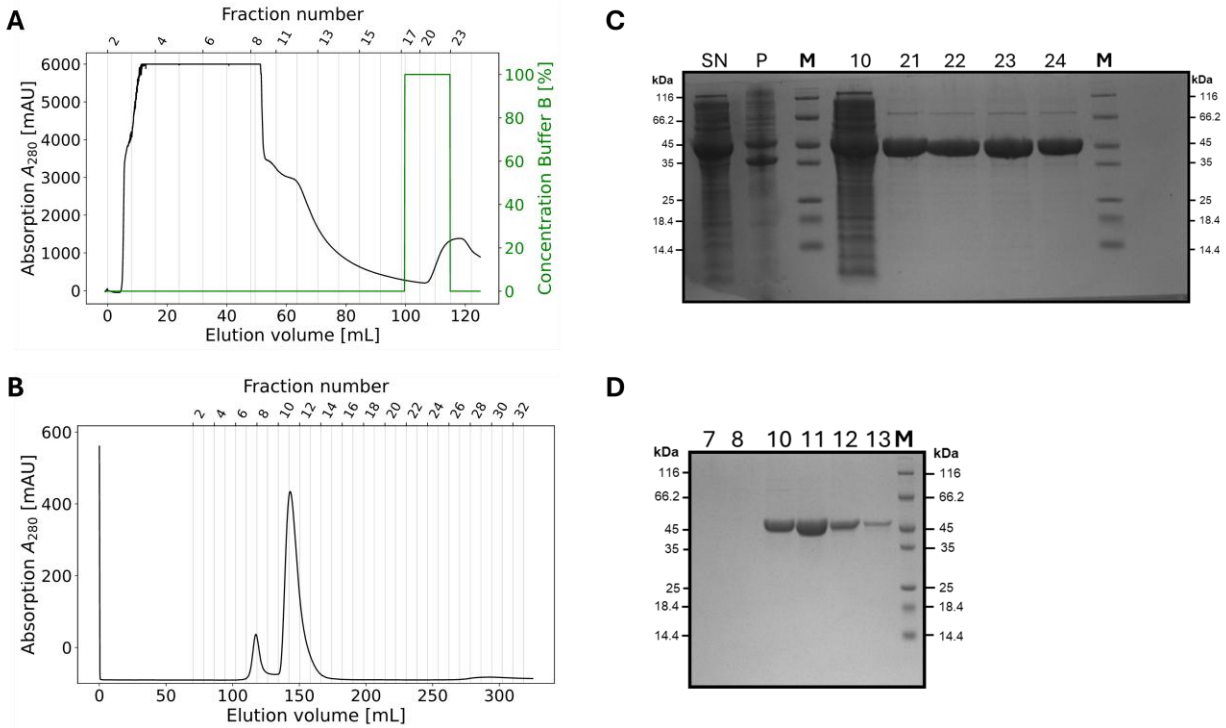


Figure 105. Purification of LacI-Link1 by Strep-Tactin AC and SEC. A) Strep-Tactin AC chromatogram of the soluble fraction after recombinant expression of LacI-Link1 (MW: 44 kDa). B) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. C, D) SDS-PAGE analyses of selected fractions from panels A and B, confirming enrichment of LacI-Link1 in the major elution peaks. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

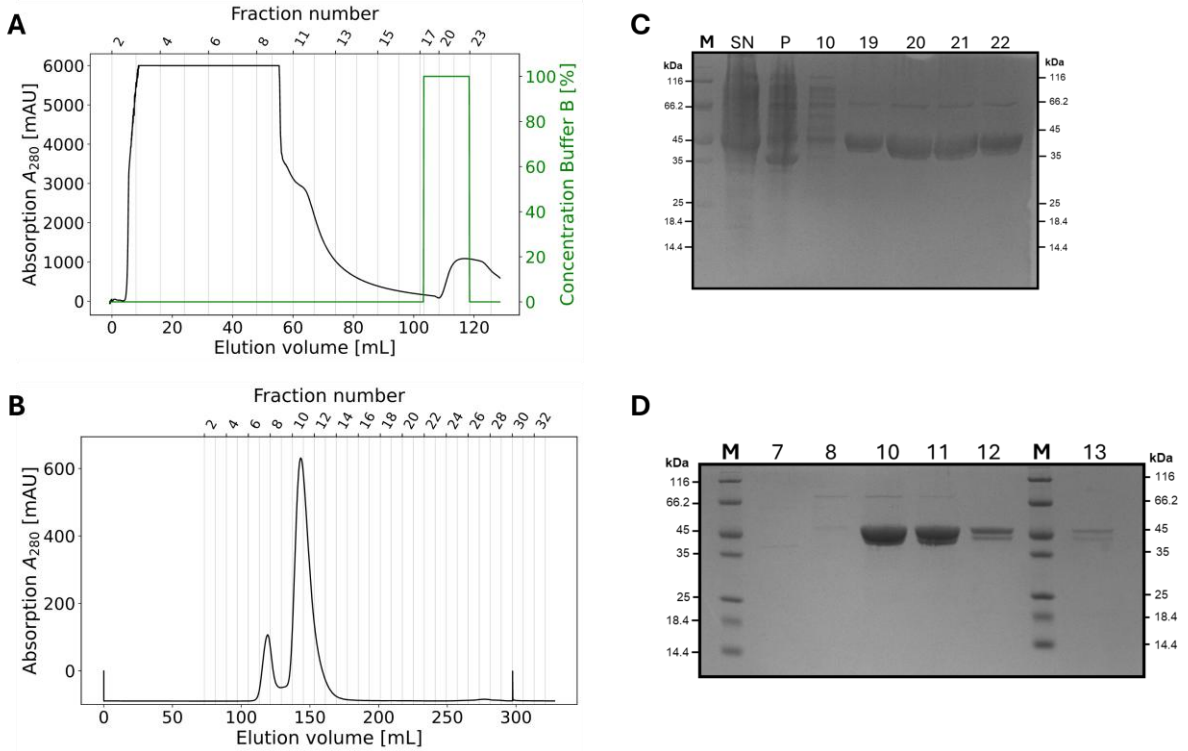


Figure 106. Purification of LacI-Link2 by Strep-Tactin AC and SEC. A) Strep-Tactin AC chromatogram of the soluble fraction after recombinant expression of LacI-Link2 (MW: 45 kDa). B) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. C, D) SDS-PAGE analyses of selected fractions from panels A and B, confirming enrichment of LacI-Link2 in the major elution peaks. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

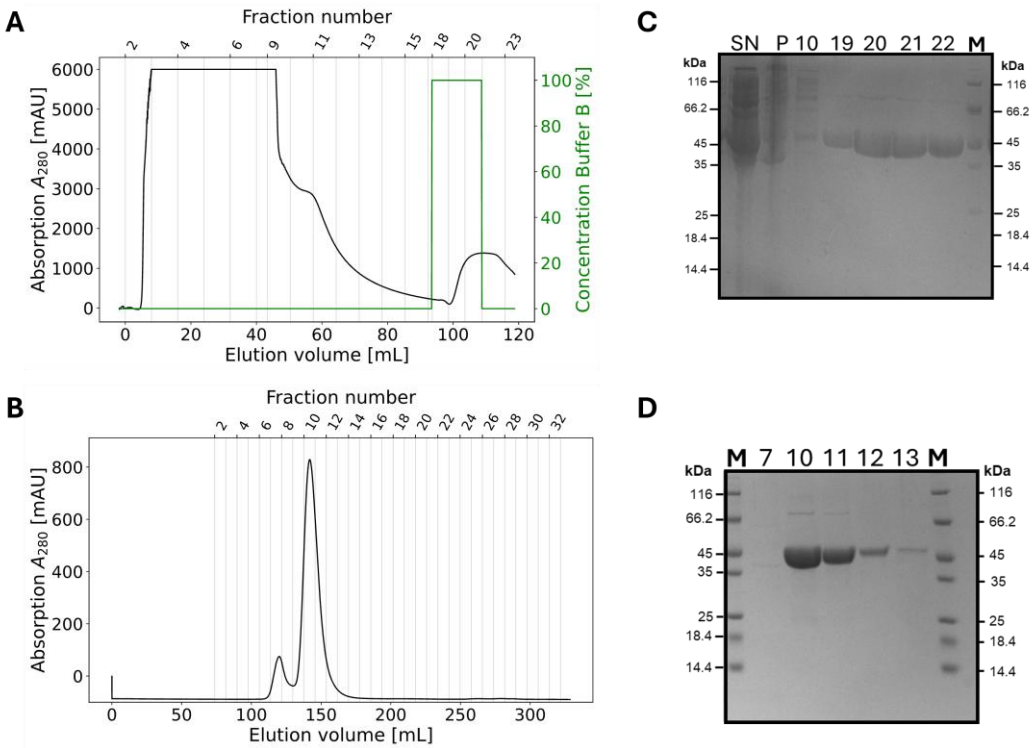


Figure 107. Purification of LacI-Link3 by Strep-Tactin AC and SEC. A) Strep-Tactin AC chromatogram of the soluble fraction after recombinant expression of LacI-Link3 (MW: 44 kDa). B) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. C, D) SDS-PAGE analyses of selected fractions from panels A and B, confirming enrichment of LacI-Link3 in the major elution peaks. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

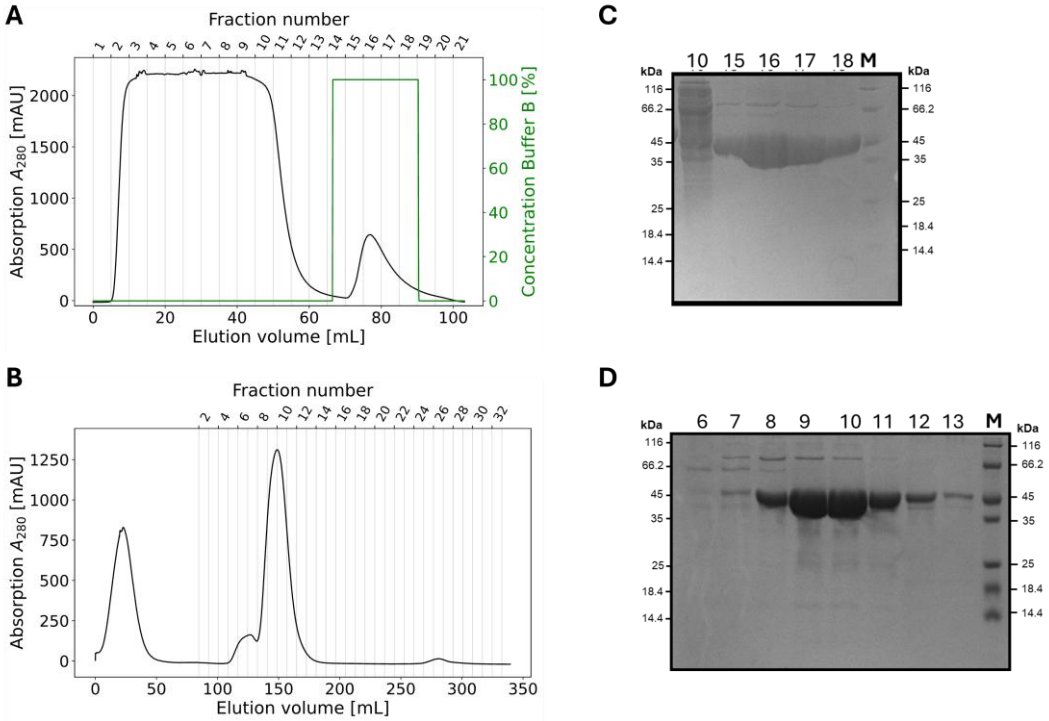


Figure 108. Purification of LacI-Link4 by Strep-Tactin AC and SEC. A) Strep-Tactin AC chromatogram of the soluble fraction after recombinant expression of LacI-Link4 (MW: 44 kDa). B) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. C, D) SDS-PAGE analyses of selected fractions from panels A and B, confirming enrichment of LacI-Link4 in the major elution peaks. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

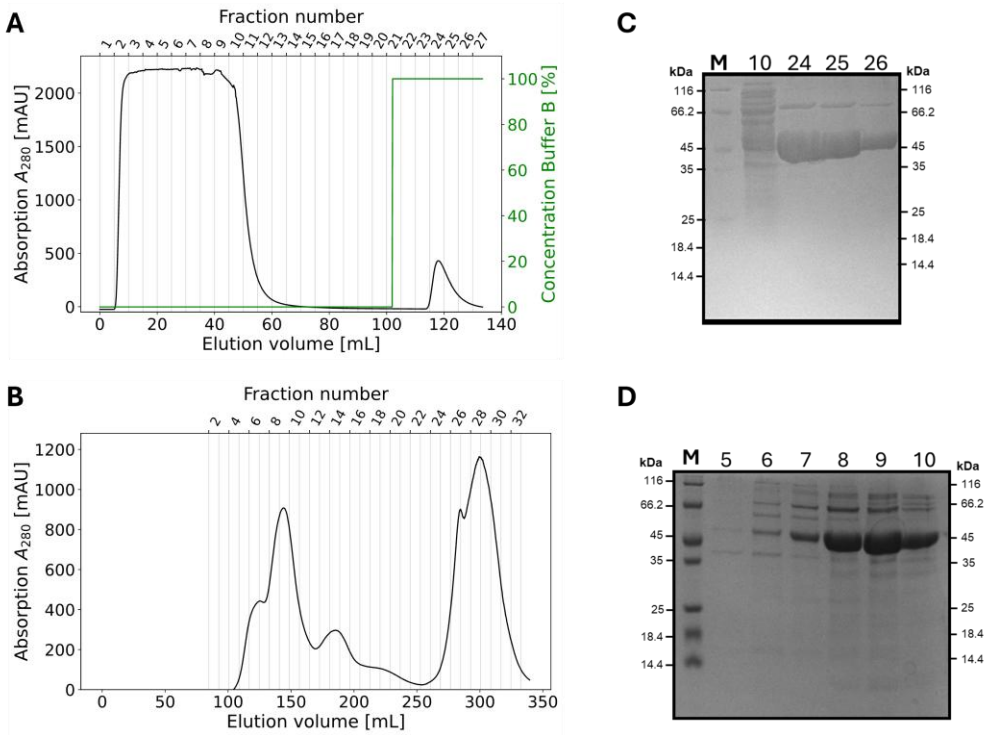


Figure 109. Purification of LacI-Link5 by Strep-Tactin AC and SEC. A) Strep-Tactin AC chromatogram of the soluble fraction after recombinant expression of LacI-Link5 (MW: 44 kDa). B) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. C, D) SDS-PAGE analyses of selected fractions from panels A and B, confirming enrichment of LacI-Link5 in the major elution peaks. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

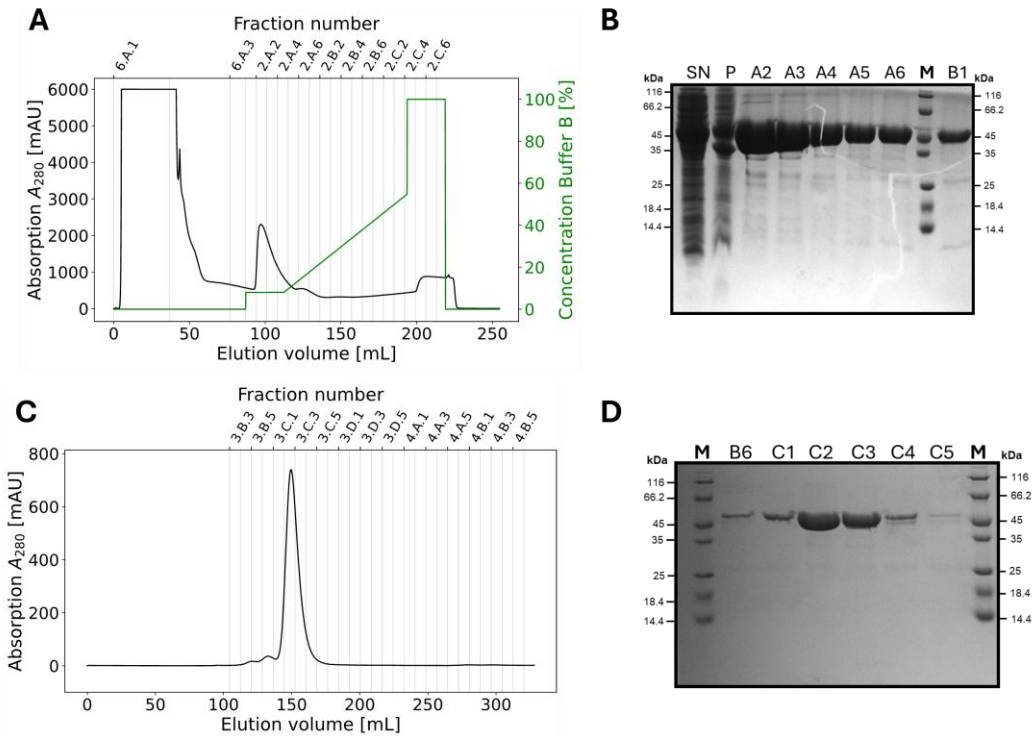


Figure 110. Purification of LacI-A2-Link1-SpyTag by Ni^{2+}-IMAC and SEC. A) Ni^{2+}-IMAC chromatogram of the soluble fraction after recombinant expression of LacI-A2-Link1-SpyTag (MW: 44 kDa). B) SDS-PAGE analysis of selected Ni^{2+}-IMAC fractions from panel A. C) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. D) SDS-PAGE analysis of selected SEC fractions from panel C. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

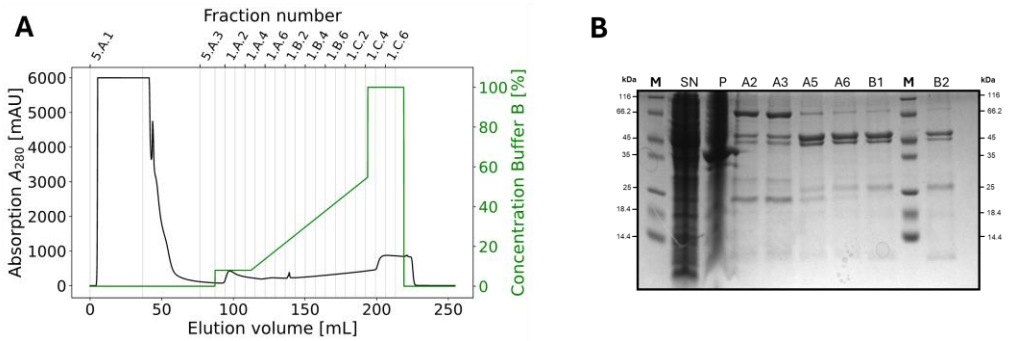


Figure 111. Purification of LacI-A2-Link2-SpyTag by Ni^{2+}-IMAC and SEC. A) Ni^{2+}-IMAC chromatogram of the soluble fraction after recombinant expression of LacI-A2-Link2-SpyTag (MW: 45 kDa). B) SDS-PAGE analysis of selected Ni^{2+}-IMAC fractions from panel A. C) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. D) SDS-PAGE analysis of selected SEC fractions from panel C. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

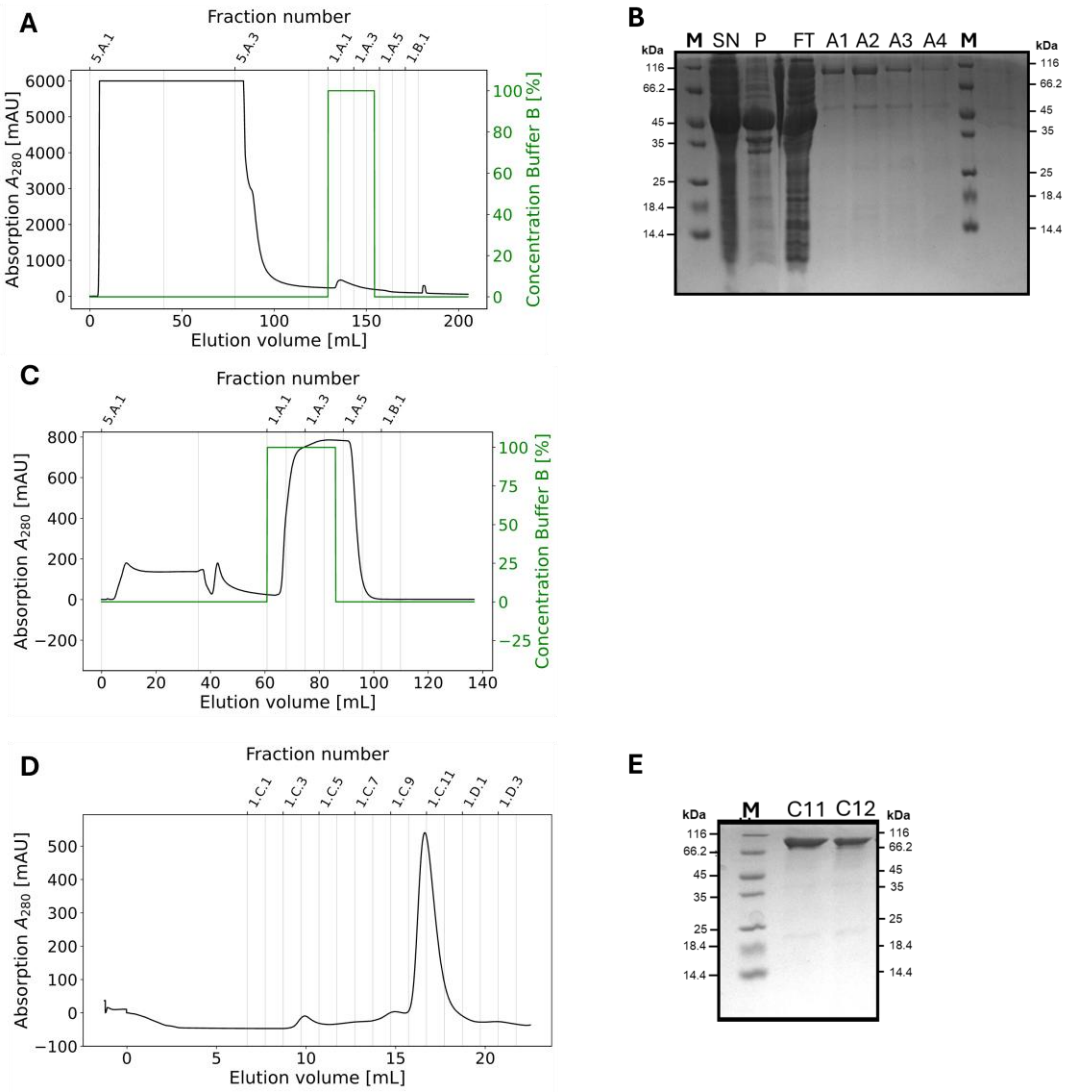


Figure 112. Purification of Het-Link1-SpyTag×Catcher by sequential Strep-Tactin AC, Ni^{2+}-IMAC, and analytical SEC. A) Strep-Tactin AC chromatogram for enrichment of Het-Link1-SpyTag×Catcher after *in vitro* ligation (MW: 94 kDa). B) SDS-PAGE analysis of selected Strep-Tactin AC fractions from panel A. C) Subsequent Ni^{2+}-IMAC chromatogram of the Strep-enriched elution fractions for further purification of the heterodimeric construct. D) Analytical SEC profile of the affinity-purified Het-Link1-SpyTag×Catcher sample, performed using a Superdex 200 Increase 10/300 GL column. E) SDS-PAGE analysis of selected SEC fractions from panel D. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

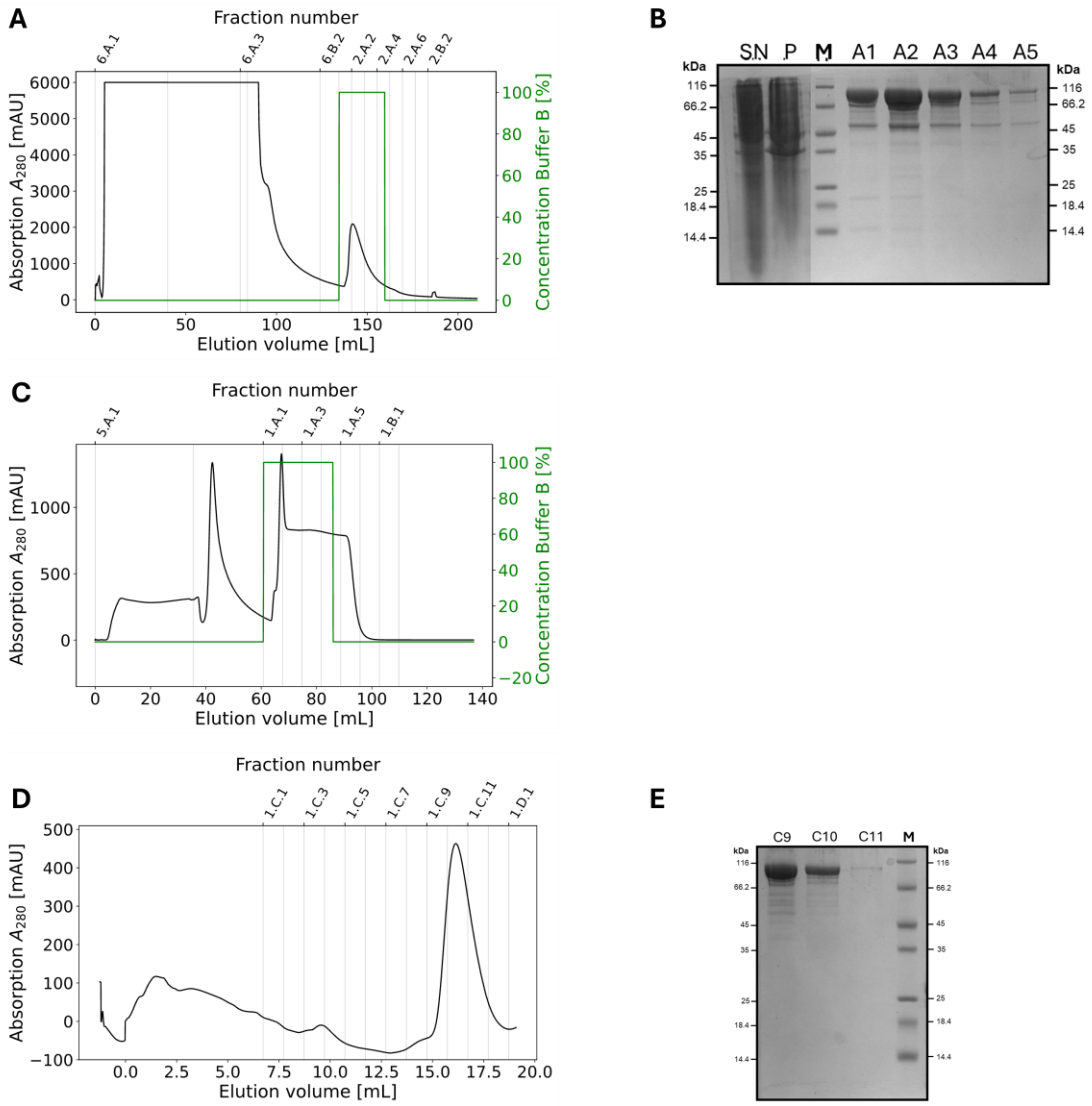


Figure 113. Purification of Het-Link2-SpyTag×Catcher by sequential Strep-Tactin AC, Ni^{2+}-IMAC, and analytical SEC. A) Strep-Tactin AC chromatogram for enrichment of Het-Link2-SpyTag×Catcher after *in vitro* ligation (MW: 93 kDa). B) SDS-PAGE analysis of selected Strep-Tactin AC fractions from panel A. C) Subsequent Ni^{2+}-IMAC chromatogram of the Strep-enriched elution fractions for further purification of the heterodimeric construct. D) Analytical SEC profile of the affinity-purified Het-Link2-SpyTag×Catcher sample, performed using a Superdex 200 Increase 10/300 GL column. E) SDS-PAGE analysis of selected SEC fractions from panel D. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

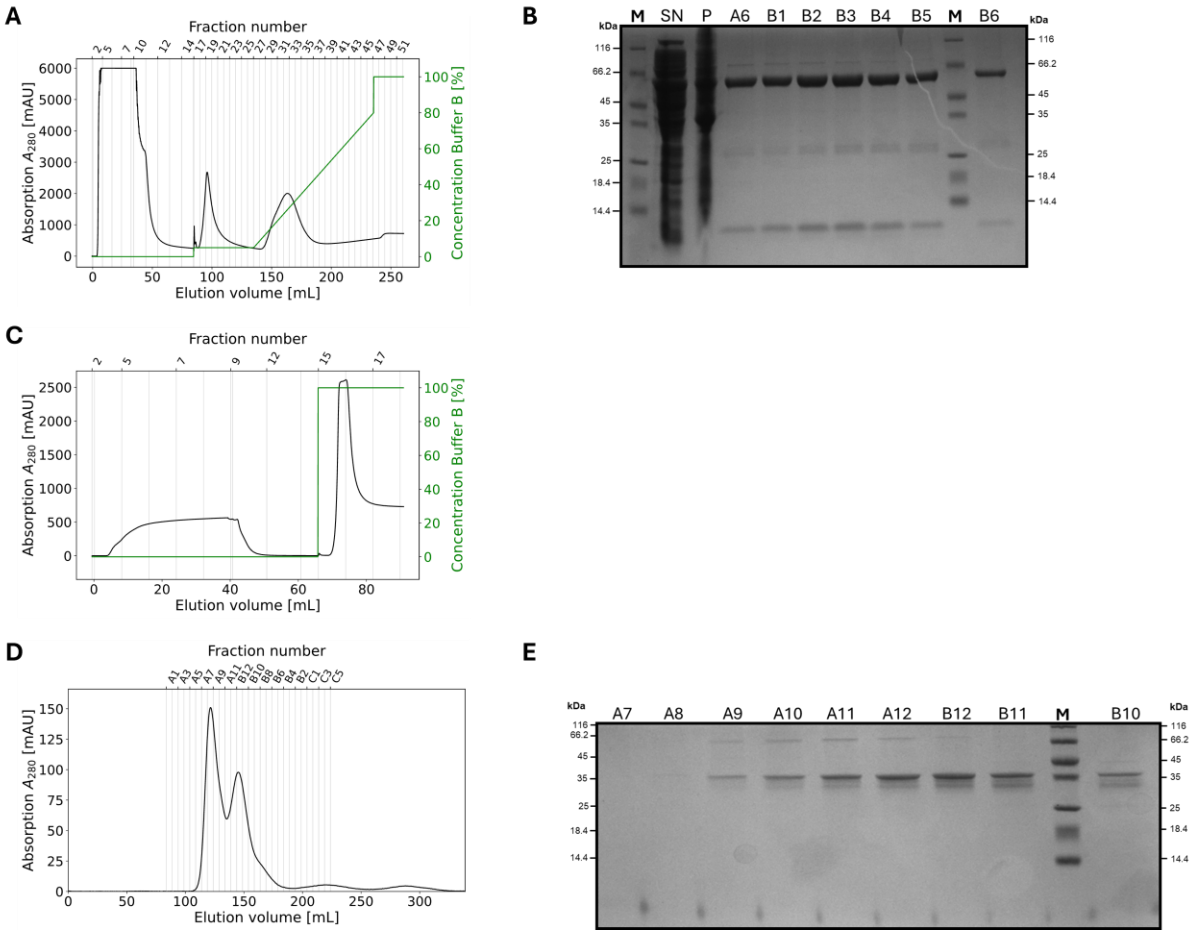


Figure 114. Purification of LacI-Gentio by Ni^{2+}-IMAC, reverse Ni^{2+}-IMAC after TEV cleavage, and SEC. A) Ni^{2+}-IMAC chromatogram of the soluble fraction after recombinant expression of His_6-tagged LacI-Gentio. B) SDS-PAGE analysis of selected Ni^{2+}-IMAC fractions, showing enrichment of the protein of interest (MW: 53 kDa). C) Reverse Ni^{2+}-IMAC chromatogram after TEV cleavage, used to separate cleaved LacI-Gentio (MW: 36 kDa) from His_6-tagged components. D) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. E) SDS-PAGE analysis of selected fractions from panels C and D, confirming the presence of LacI-Gentio in the reverse Ni^{2+}-IMAC flow-through and in the major SEC elution peak. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

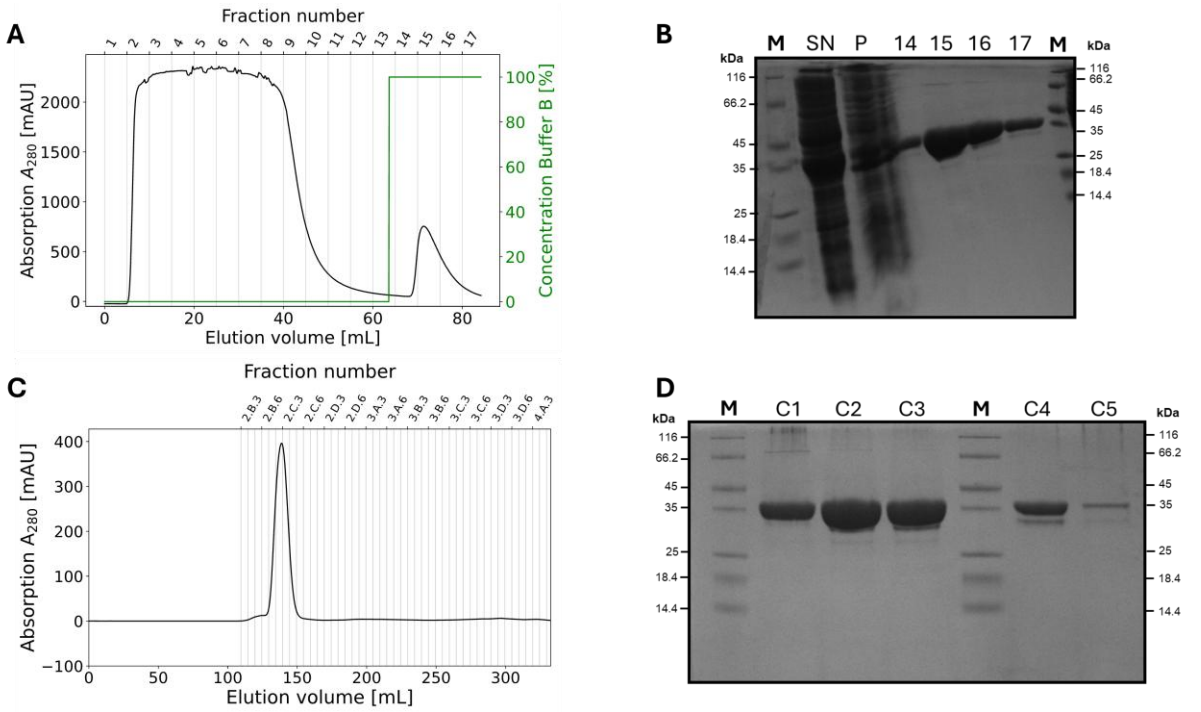


Figure 115. Purification of LLhT by Strep-Tactin AC and SEC. A) Strep-Tactin AC chromatogram of the soluble fraction after recombinant expression of LLhT (MW: 39 kDa). B) SDS-PAGE analysis of selected Strep-Tactin AC fractions from panel A. C) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. D) SDS-PAGE analysis of selected SEC fractions from panel C. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

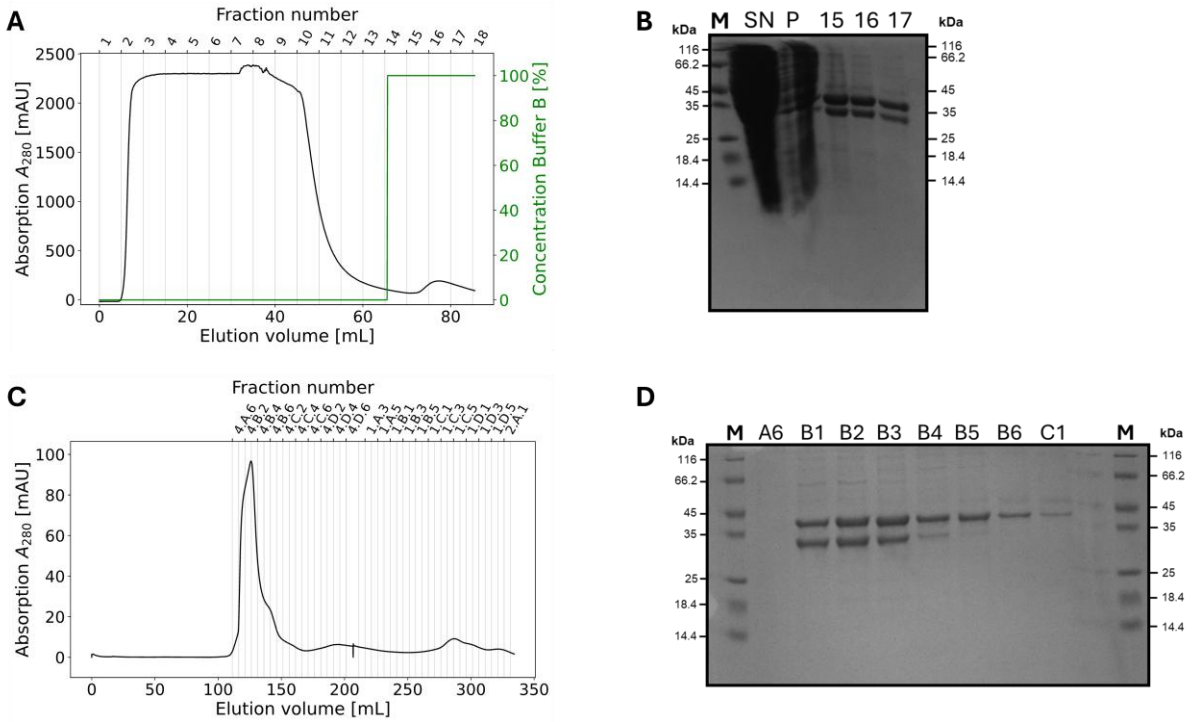


Figure 116. Purification of LLhF by Strep-Tactin AC and SEC. A) Strep-Tactin AC chromatogram of the soluble fraction after recombinant expression of LLhF (MW: 40 kDa). B) SDS-PAGE analysis of selected Strep-Tactin AC fractions from panel A. C) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. D) SDS-PAGE analysis of selected SEC fractions from panel C. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

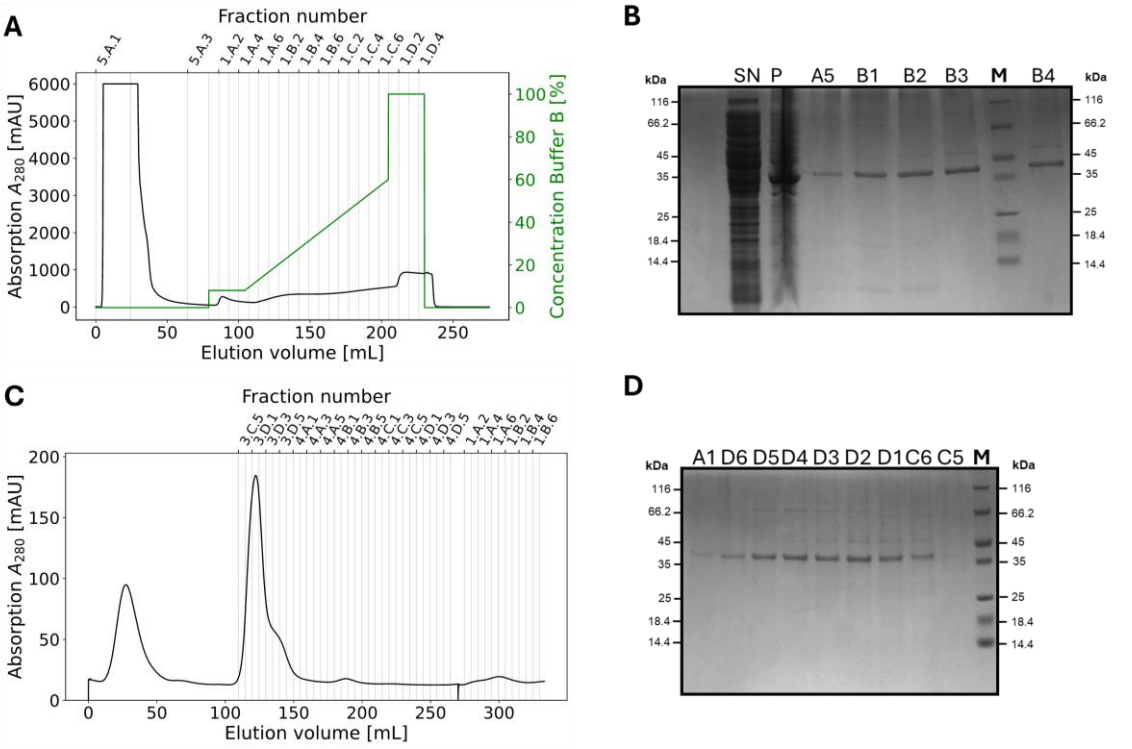


Figure 117. Purification of LLhT-Lev1 by Ni^{2+}-IMAC and SEC. A) Ni^{2+}-IMAC chromatogram of the soluble fraction after recombinant expression of LLhT-Lev1 (MW: 39 kDa). B) SDS-PAGE analysis of selected Ni^{2+}-IMAC fractions from panel A. C) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. D) SDS-PAGE analysis of selected SEC fractions from panel C. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

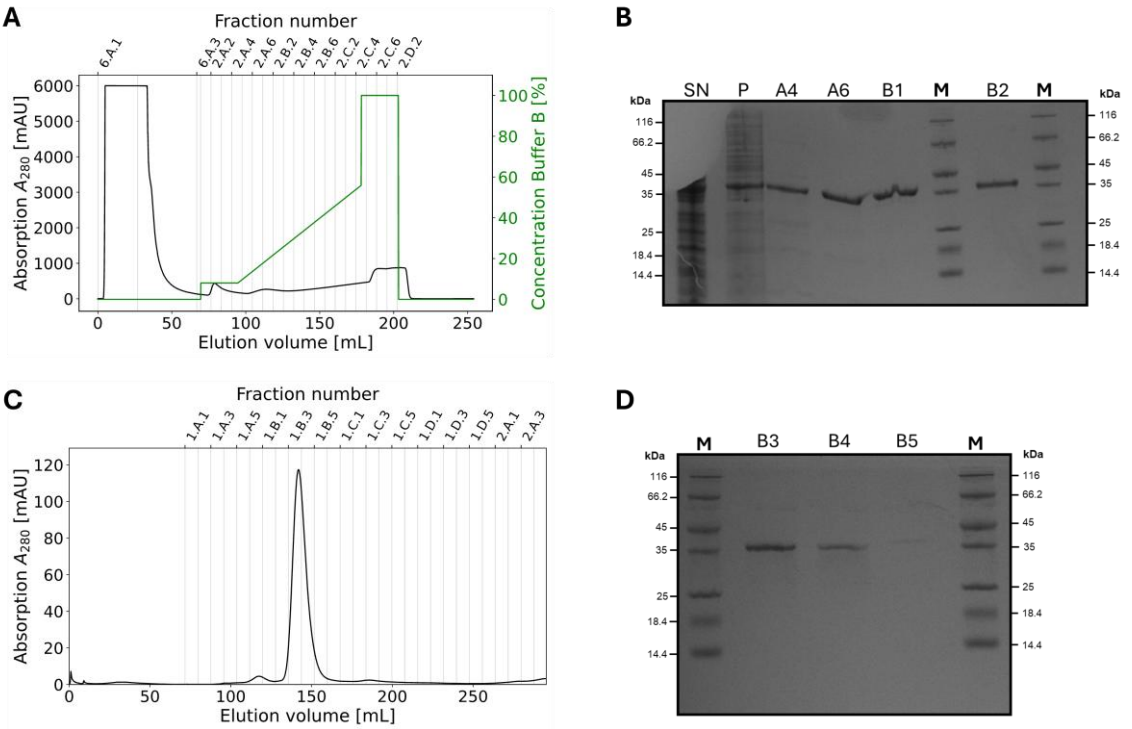


Figure 118. Purification of LLhT-Lev7 by Ni^{2+}-IMAC and SEC. A) Ni^{2+}-IMAC chromatogram of the soluble fraction after recombinant expression of LLhT-Lev7 (MW: 37 kDa). B) SDS-PAGE analysis of selected Ni^{2+}-IMAC fractions from panel A. C) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. D) SDS-PAGE analysis of selected SEC fractions from panel C. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

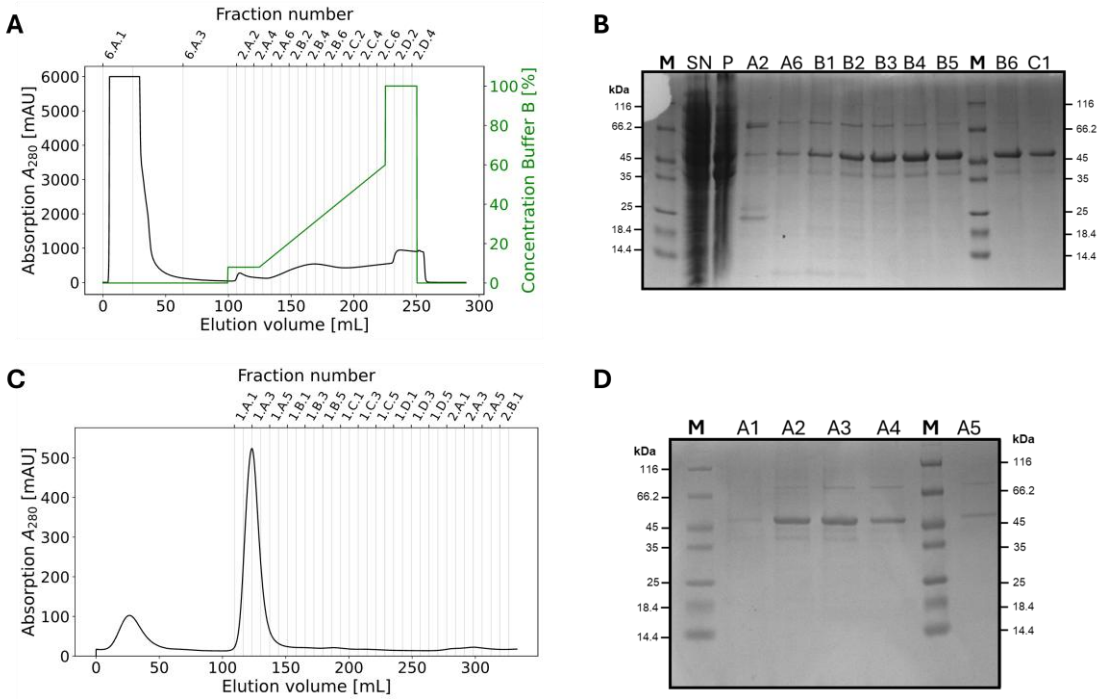


Figure 119. Purification of LLhT-Link6 by Ni^{2+}-IMAC and SEC. A) Ni^{2+}-IMAC chromatogram of the soluble fraction after recombinant expression of LLhT-Link6 (MW: 46 kDa). B) SDS-PAGE analysis of selected Ni^{2+}-IMAC fractions from panel A. C) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. D) SDS-PAGE analysis of selected SEC fractions from panel C. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

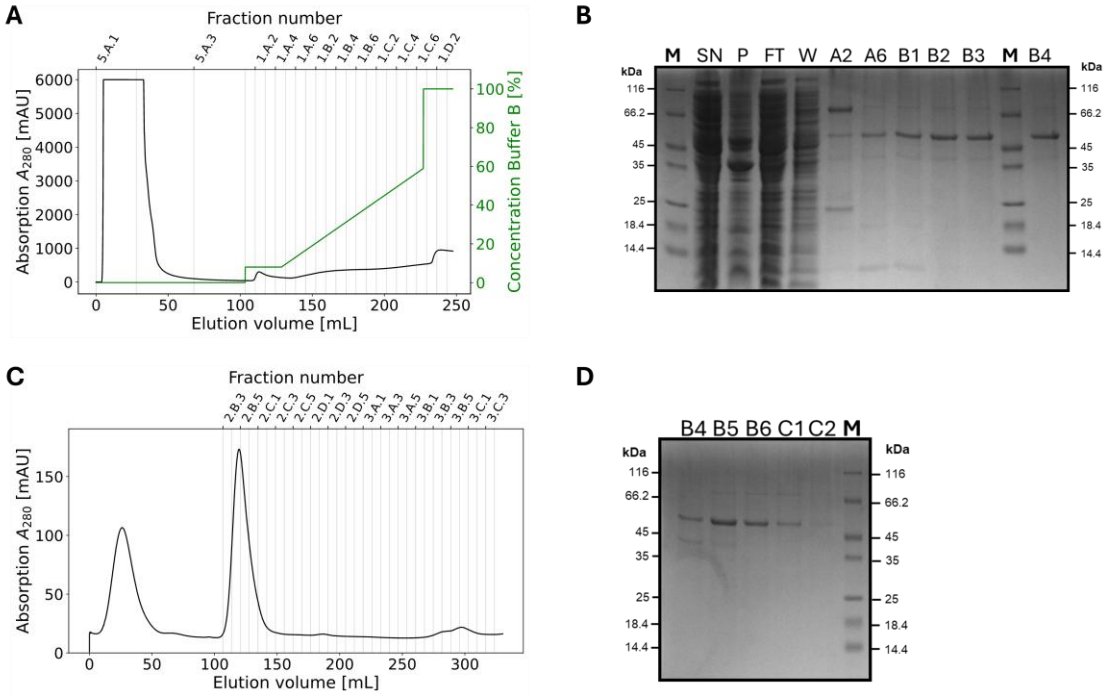


Figure 120. Purification of LLhT-Link7 by Ni^{2+}-IMAC and SEC. A) Ni^{2+}-IMAC chromatogram of the soluble fraction after recombinant expression of LLhT-Link7 (MW: 48 kDa). B) SDS-PAGE analysis of selected Ni^{2+}-IMAC fractions from panel A. C) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. D) SDS-PAGE analysis of selected SEC fractions from panel C. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

.

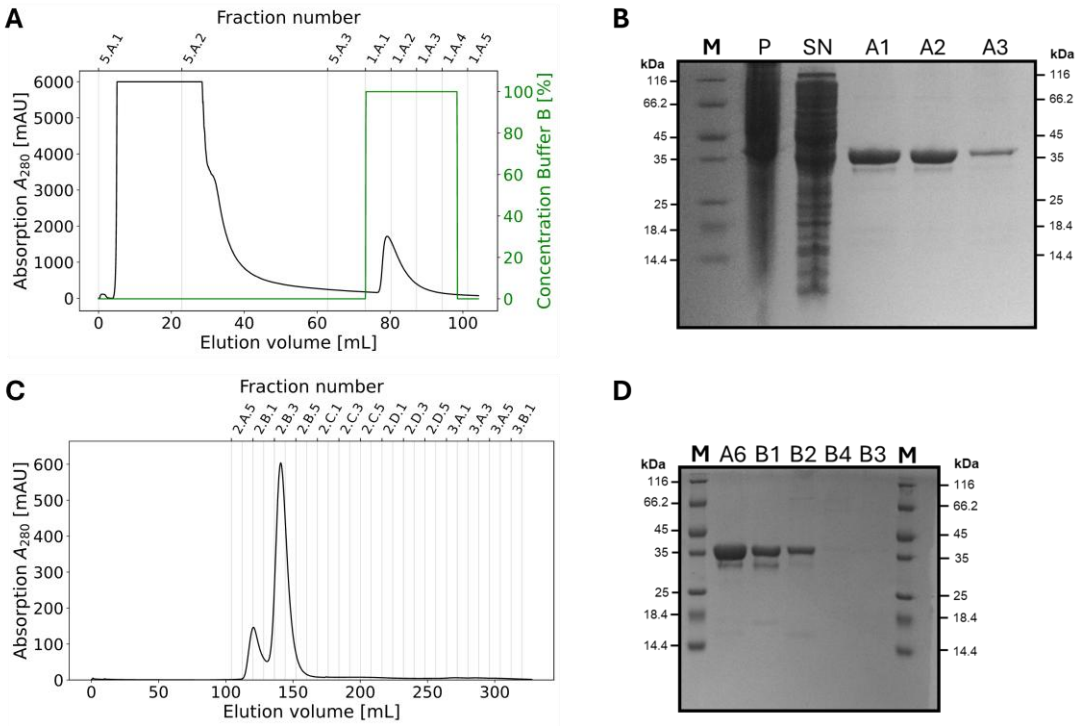


Figure 121. Purification of LLhT-L244A by Strep-Tactin AC and SEC. A) Strep-Tactin AC chromatogram of the soluble fraction after recombinant expression of LLhT-L244A (MW: 36 kDa). B) SDS-PAGE analysis of selected Strep-Tactin AC fractions from panel A. C) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. D) SDS-PAGE analysis of selected SEC fractions from panel C. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

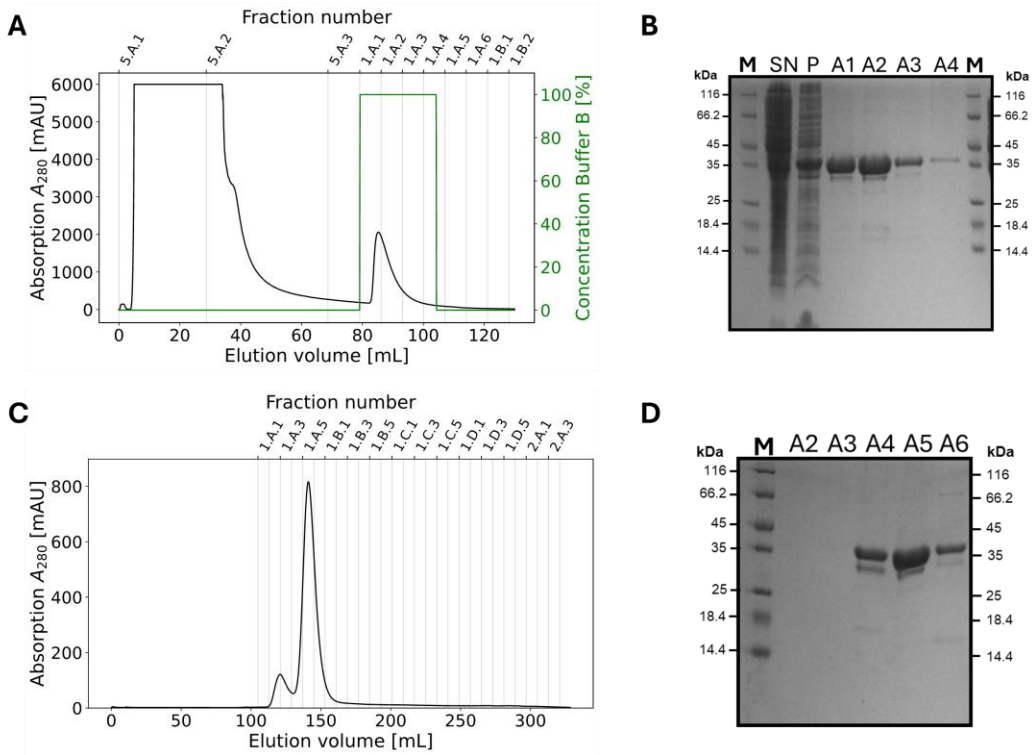


Figure 122. Purification of LLhT-L268A by Strep-Tactin AC and SEC. A) Strep-Tactin AC chromatogram of the soluble fraction after recombinant expression of LLhT-L268A (MW: 36 kDa). B) SDS-PAGE analysis of selected Strep-Tactin AC fractions from panel A. C) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. D) SDS-PAGE analysis of selected SEC fractions from panel C. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

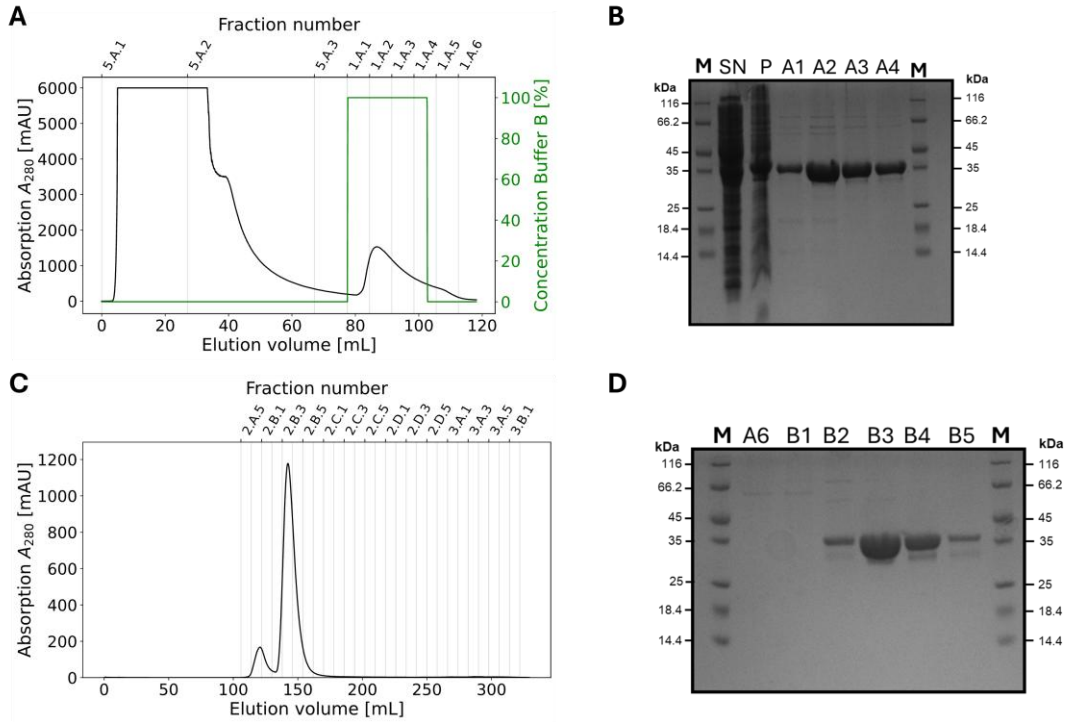


Figure 123. Purification of LLhT-F271A by Strep-Tactin AC and SEC. A) Strep-Tactin AC chromatogram of the soluble fraction after recombinant expression of LLhT-F271A (MW: 36 kDa). B) SDS-PAGE analysis of selected Strep-Tactin AC fractions from panel A. C) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. D) SDS-PAGE analysis of selected SEC fractions from panel C. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

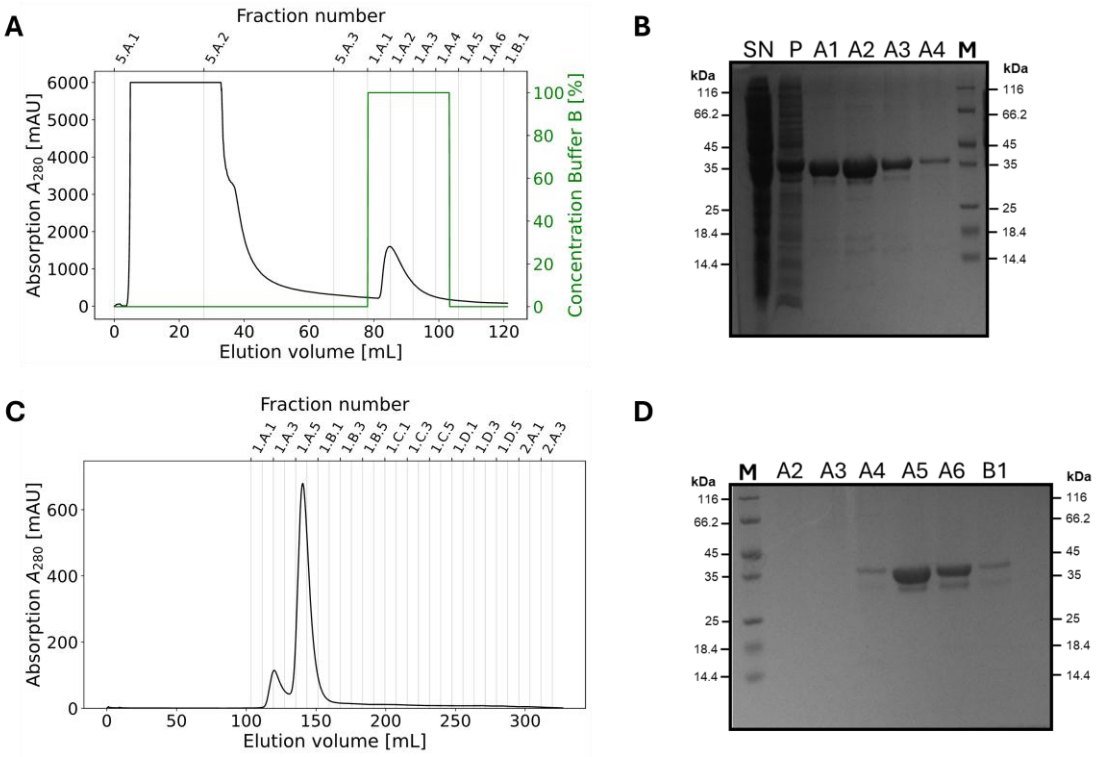


Figure 124. Purification of LLhT-L272A by Strep-Tactin AC and SEC. A) Strep-Tactin AC chromatogram of the soluble fraction after recombinant expression of LLhT-L272A (MW: 36 kDa). B) SDS-PAGE analysis of selected Strep-Tactin AC fractions from panel A. C) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. D) SDS-PAGE analysis of selected SEC fractions from panel C. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

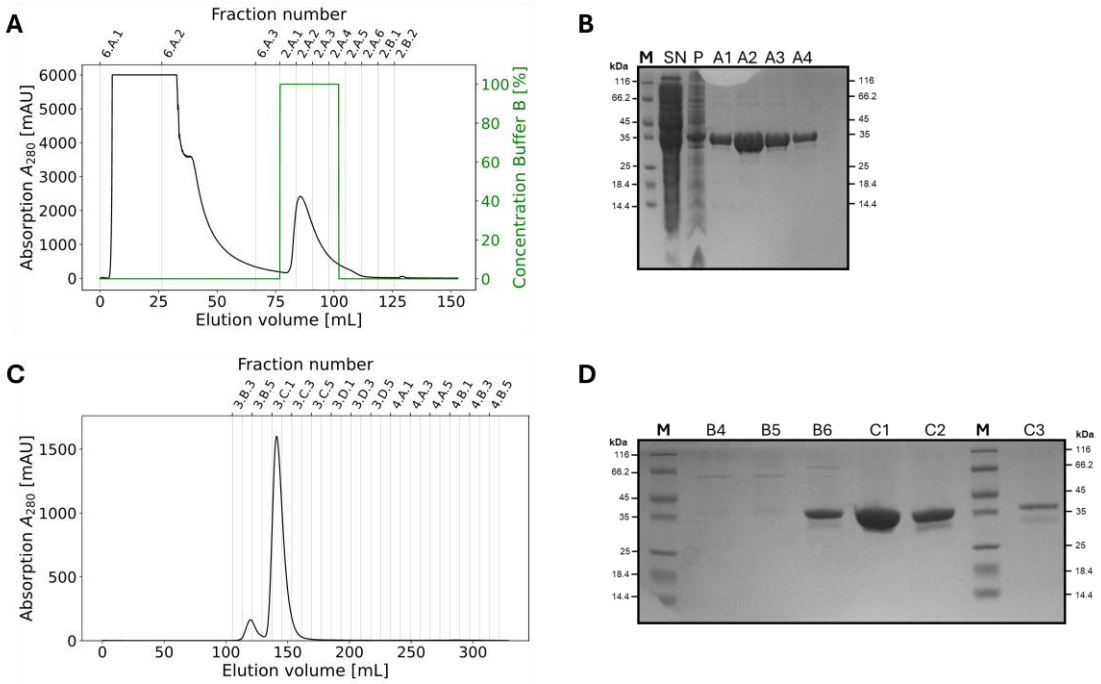


Figure 125. Purification of LLhT-L272S by Strep-Tactin AC and SEC. A) Strep-Tactin AC chromatogram of the soluble fraction after recombinant expression of LLhT-L272S (MW: 36 kDa). B) SDS-PAGE analysis of selected Strep-Tactin AC fractions from panel A. C) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. D) SDS-PAGE analysis of selected SEC fractions from panel C. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

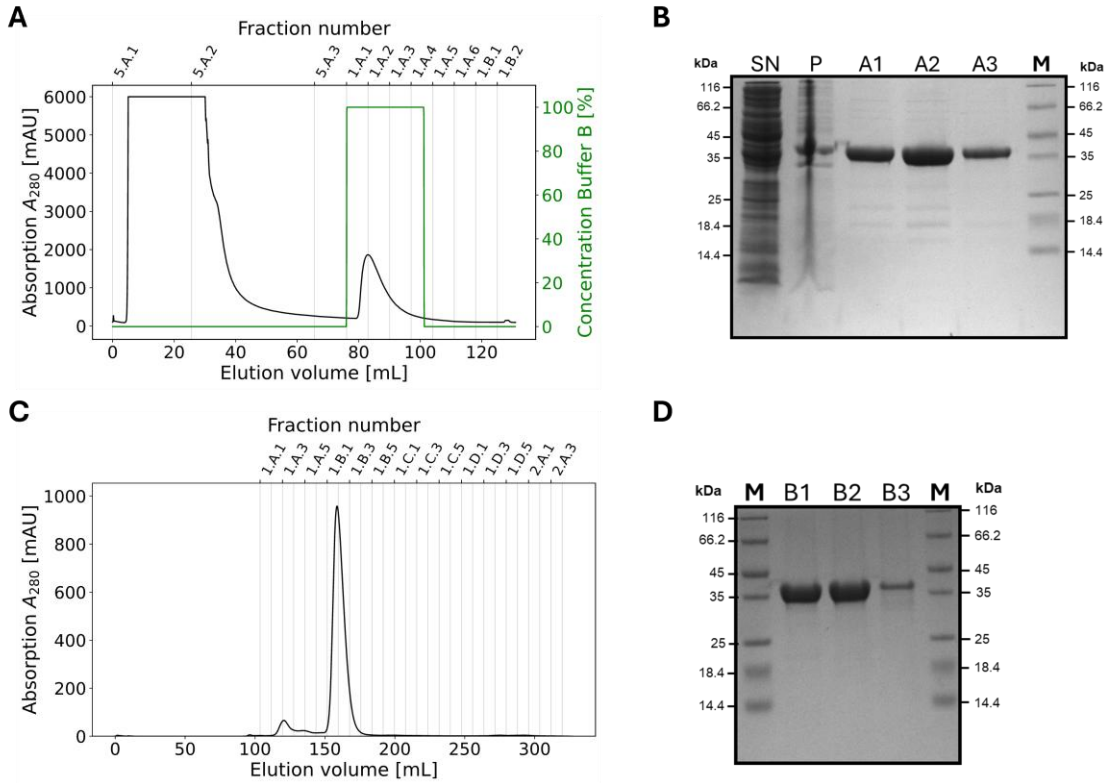


Figure 126. Purification of LLhT-4Mut by Strep-Tactin AC and SEC. A) Strep-Tactin AC chromatogram of the soluble fraction after recombinant expression of LLhT-4Mut (MW: 36 kDa). B) SDS-PAGE analysis of selected Strep-Tactin AC fractions from panel A. C) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. D) SDS-PAGE analysis of selected SEC fractions from panel C. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

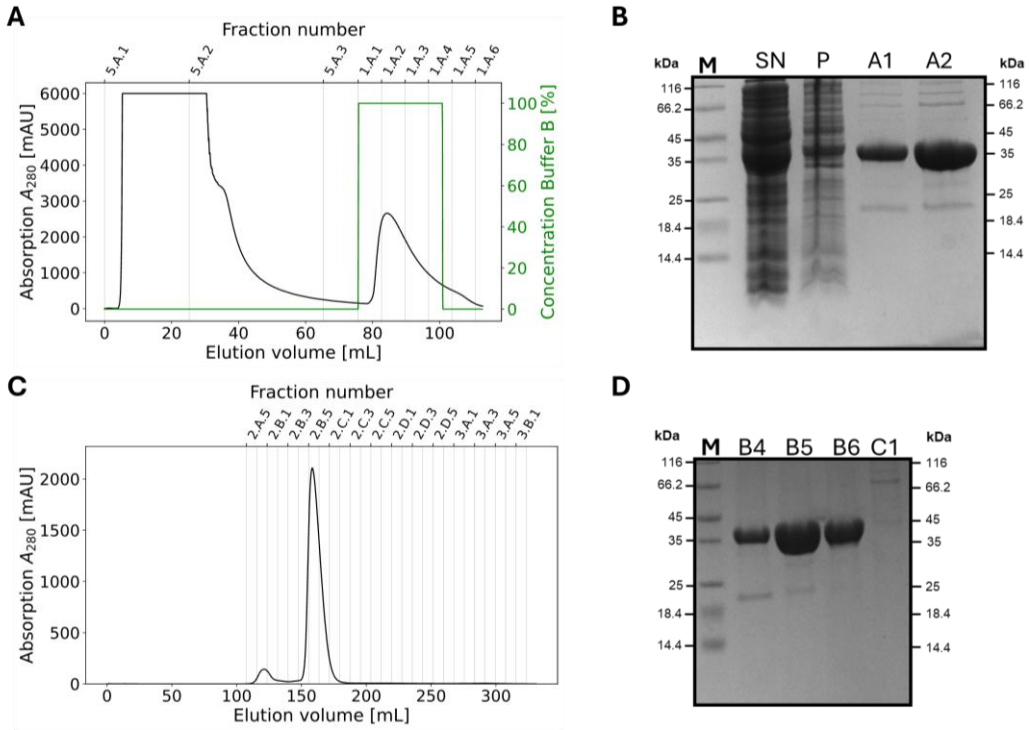


Figure 127. Purification of LLhT-6Mut by Strep-Tactin AC and SEC. A) Strep-Tactin AC chromatogram of the soluble fraction after recombinant expression of LLhT-6Mut (MW: 36 kDa). B) SDS-PAGE analysis of selected Strep-Tactin AC fractions from panel A. C) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. D) SDS-PAGE analysis of selected SEC fractions from panel C. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

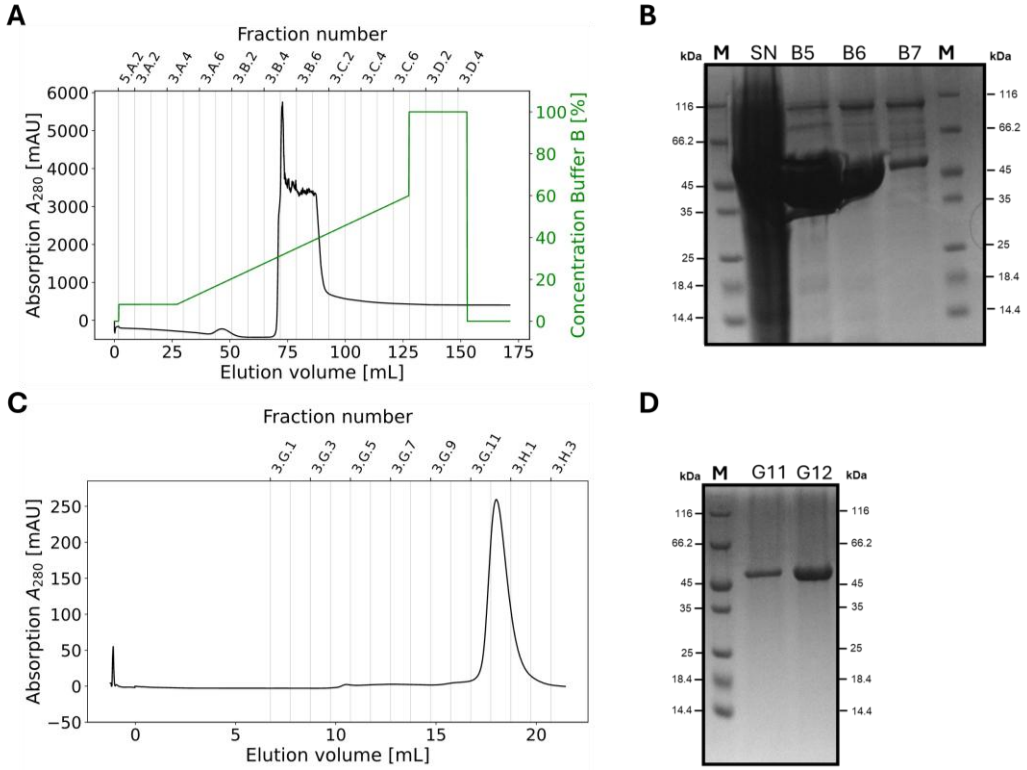


Figure 128. Purification of LLhT-6Mut-DogCatcher by Ni^{2+}-IMAC and SEC. A) Ni^{2+}-IMAC chromatogram of the soluble fraction after recombinant expression of LLhT-6Mut-DogCatcher (MW: 46 kDa). B) SDS-PAGE analysis of selected Ni^{2+}-IMAC fractions from panel A. C) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. D) SDS-PAGE analysis of selected SEC fractions from panel C. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

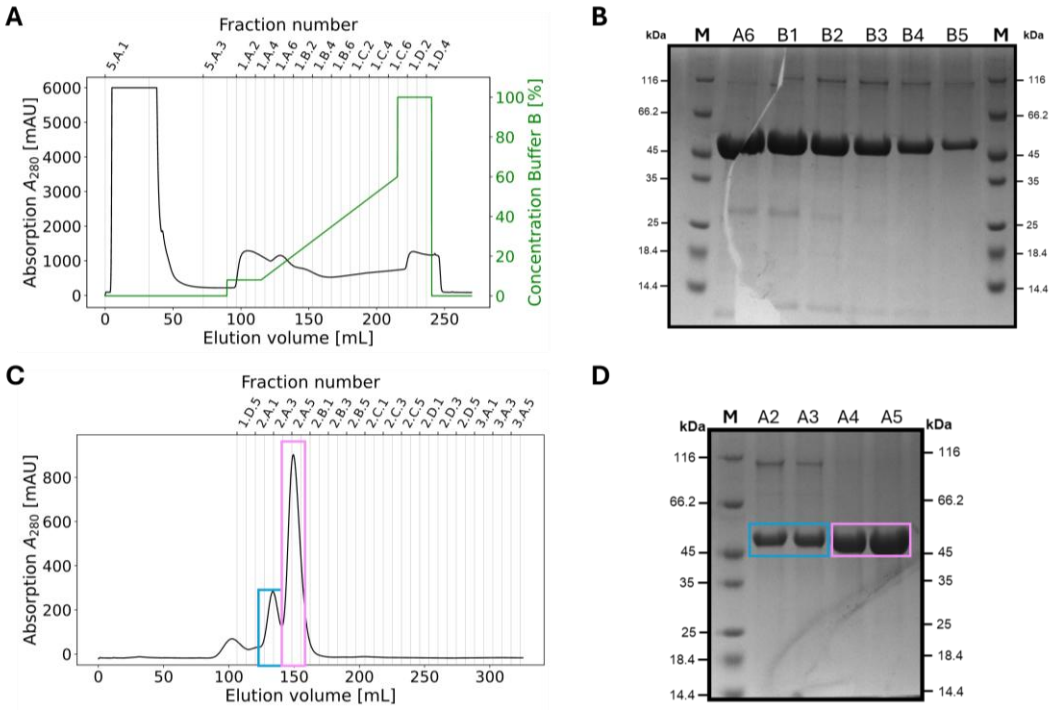


Figure 129. Purification of LacI*-DogCatcher by Ni^{2+}-IMAC and SEC. A) Ni^{2+}-IMAC chromatogram of soluble LacI*-DogCatcher (MW: 49 kDa). B) SDS-PAGE of selected Ni^{2+}-IMAC fractions. C) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. D) SDS-PAGE of selected SEC fractions. The first and second SEC peaks are highlighted in blue and purple, respectively. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

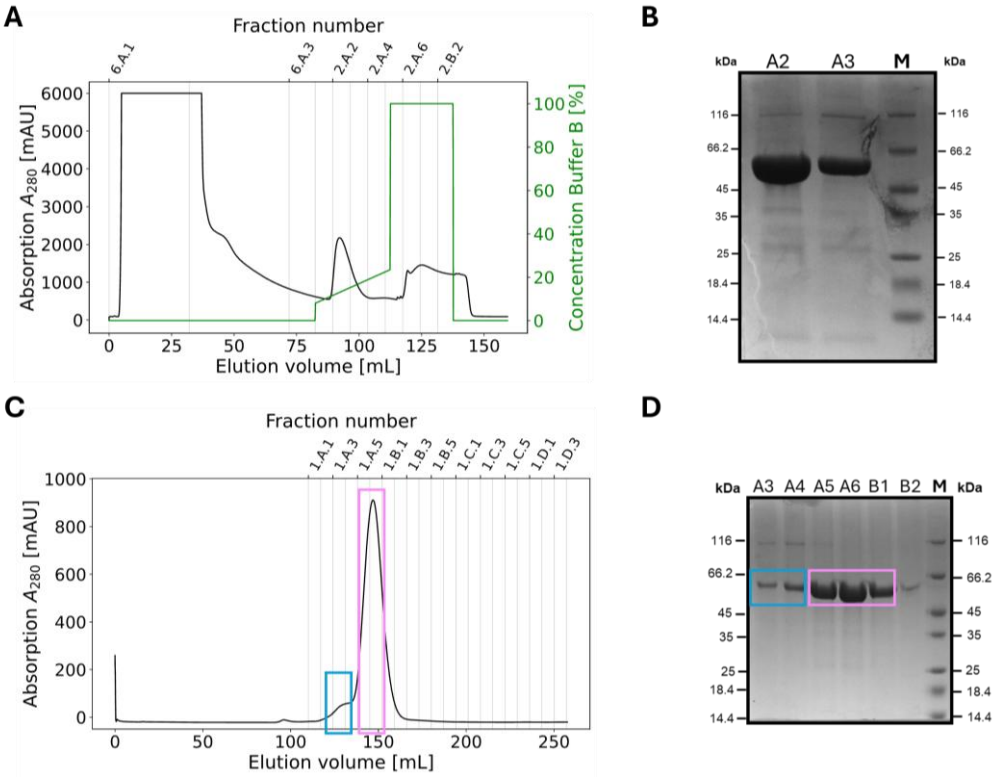


Figure 130. Purification of LacI-A2-Link1-SpyCatcher-DogTag by Ni^{2+}-IMAC and SEC. A) Ni^{2+}-IMAC chromatogram of soluble LacI-A2-Link1-SpyCatcher-DogTag (MW: 54 kDa). B) SDS-PAGE of selected Ni^{2+}-IMAC fractions. C) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. D) SDS-PAGE of selected SEC fractions. The first and second SEC peaks are highlighted in blue and purple, respectively. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

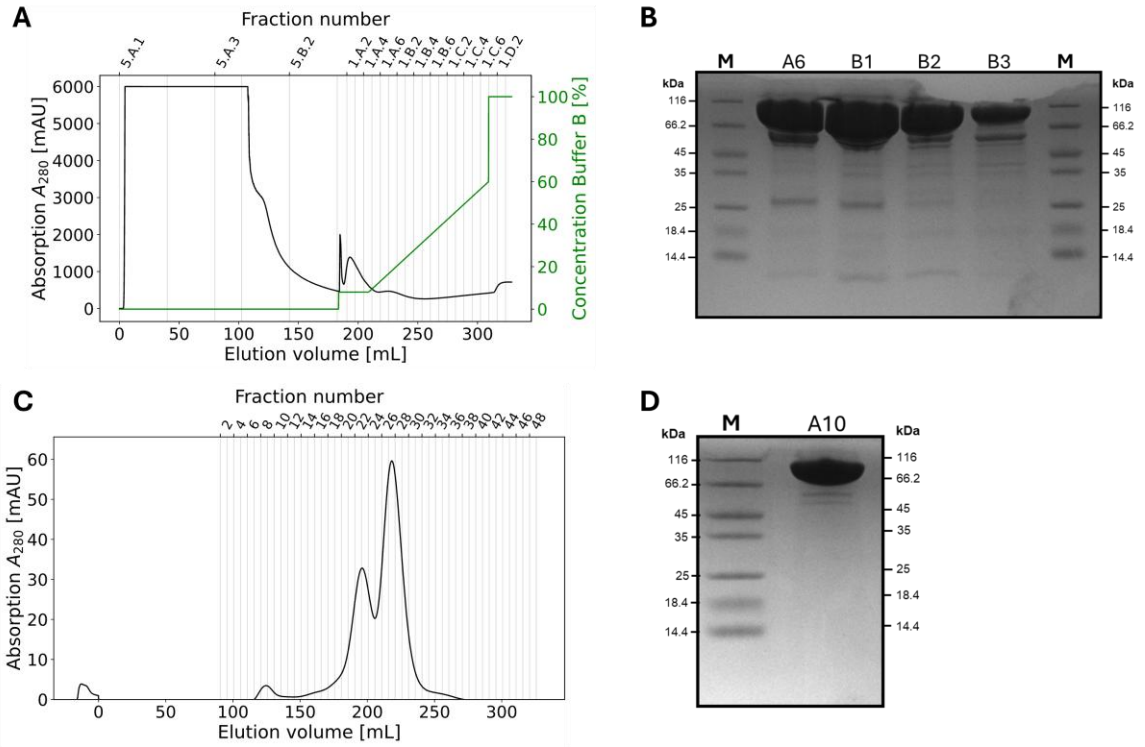


Figure 131. Purification of Het-Link1-DogTagxCatcher by Ni^{2+}-IMAC and SEC.A) Ni^{2+}-IMAC chromatogram of the soluble fraction after recombinant expression of Het-Link1-DogTag×Catcher (MW: 103 kDa). B) SDS-PAGE analysis of selected Ni^{2+}-IMAC fractions from panel A. C) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 200 pg column. D) SDS-PAGE analysis of selected SEC fractions from panel C. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

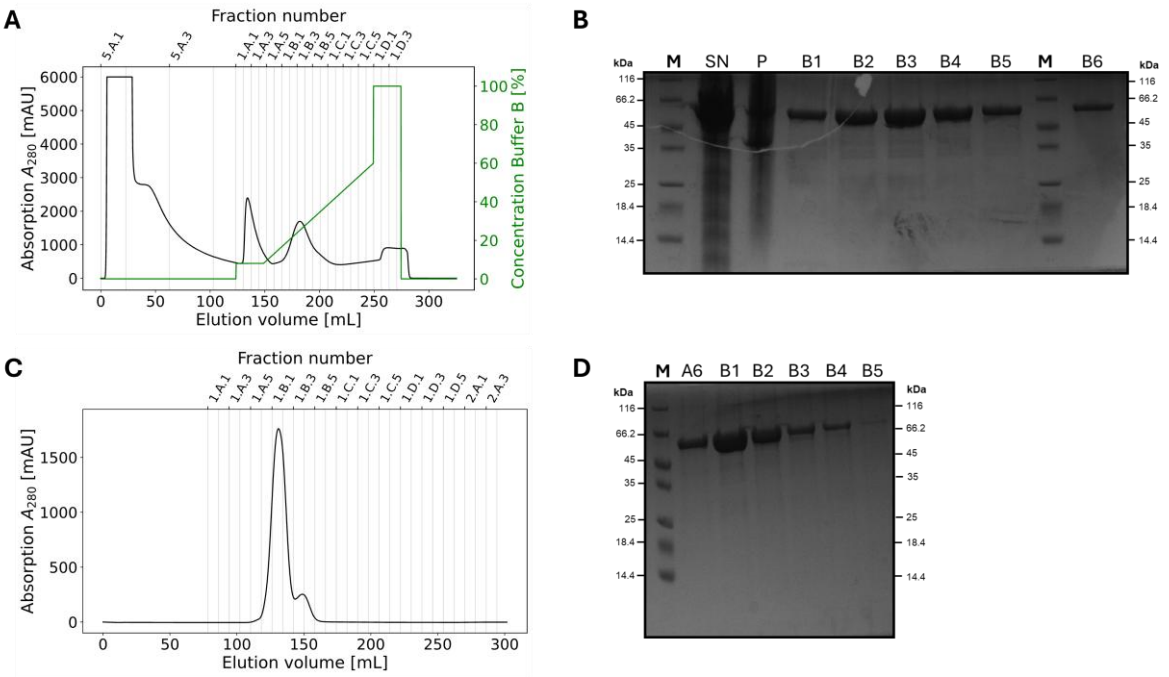


Figure 132. Purification of CM-PotF by Ni^{2+}-IMAC and SEC. A) Ni^{2+}-IMAC chromatogram of the soluble fraction after recombinant expression of CM-PotF (MW: 56 kDa). B) SDS-PAGE analysis of selected Ni^{2+}-IMAC fractions from panel A. C) SEC chromatogram of the subsequent purification step using a HiLoad 26/600 Superdex 75 pg column. D) SDS-PAGE analysis of selected SEC fractions from panel C. Chromatograms show A_{280} as a function of elution volume, with collected fractions indicated. For affinity chromatograms, the buffer B concentration is shown on the right y-axis in green.

7.3 Characterisation of IPTG binding with fluorescence titration

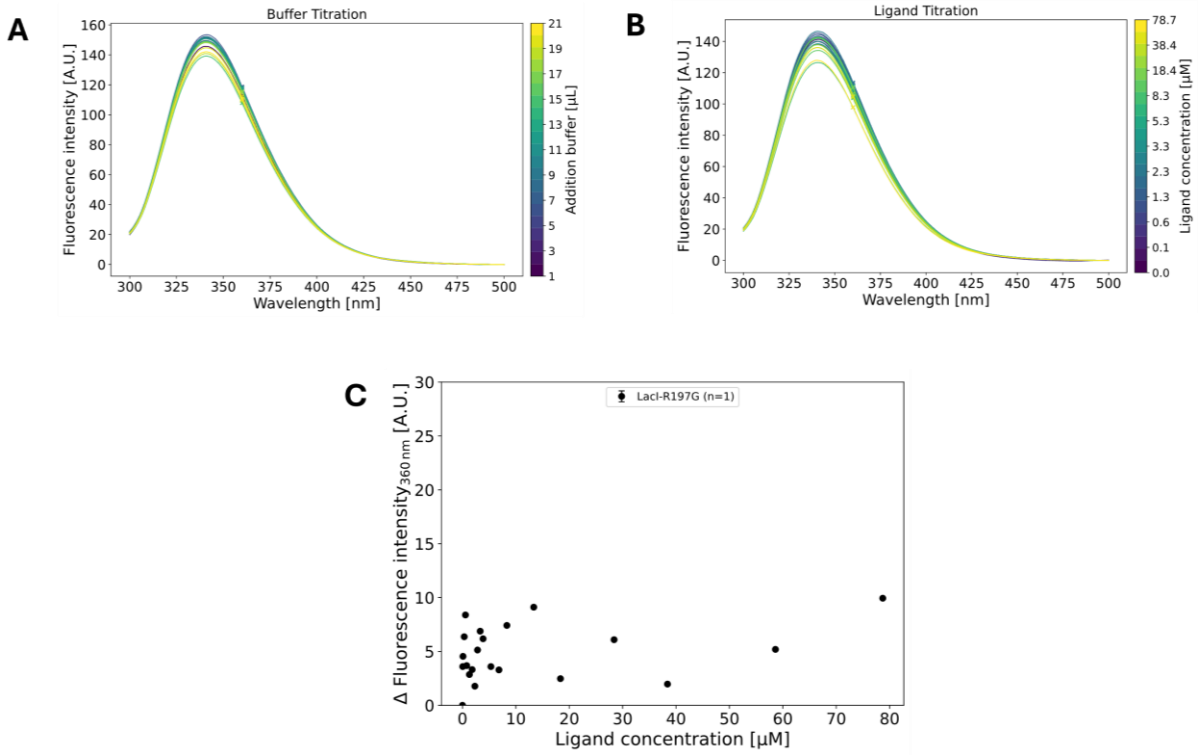


Figure 133.Fluorescence-based analysis of IPTG binding to ligand-deficient variant LacI-R197G. Fluorescence emission spectra were recorded at a constant protein concentration. The addition of either buffer volume or IPTG concentration is indicated by the colour scale. A) Buffer control measurement of LacI-R197G. B) Fluorescence emission spectra of LacI-R197G recorded upon addition of increasing IPTG concentrations. C) IPTG-dependent change in fluorescence intensity at 360 nm plotted against IPTG concentration for LacI-R197G. Data points represent mean values, and error bars indicate the standard deviation of the indicated number of replicates. Measurements were performed in Buffer C using a protein concentration of 0.2 mg·mL^{-1}.

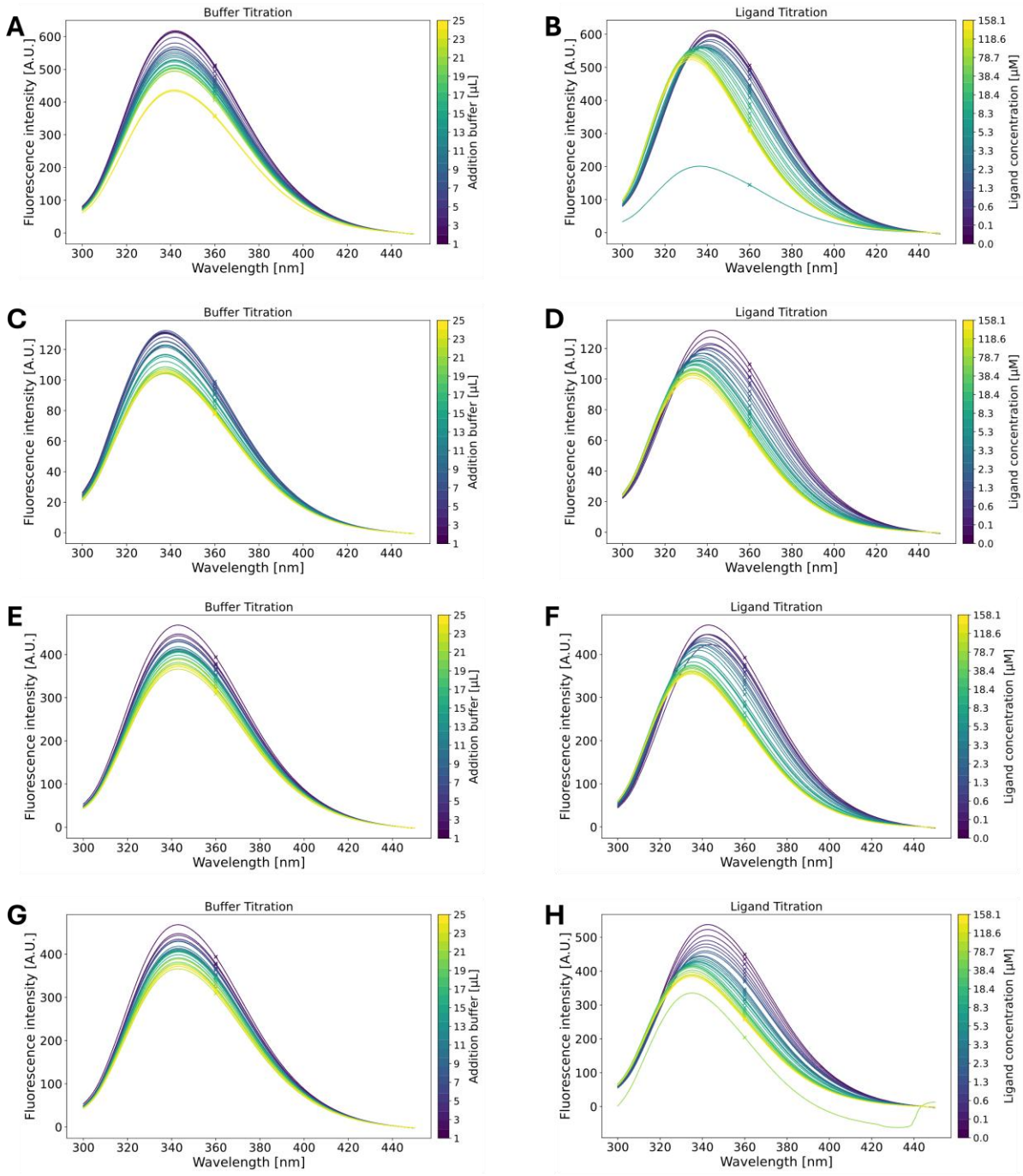


Figure 134. Fluorescence-based analysis of IPTG binding to LacI interface variants. Fluorescence emission spectra were recorded at constant protein concentration after sequential buffer addition as a control and upon titration with increasing IPTG concentrations. The added buffer volume or IPTG concentration is indicated by the colour scale. A, B) LacI-B1. C, D) LacI-A2. E, F) LacI-B2. G, H) LacI-B3. Measurements were performed in Buffer C using a protein concentration of 0.2 mg·mL^{-1}.

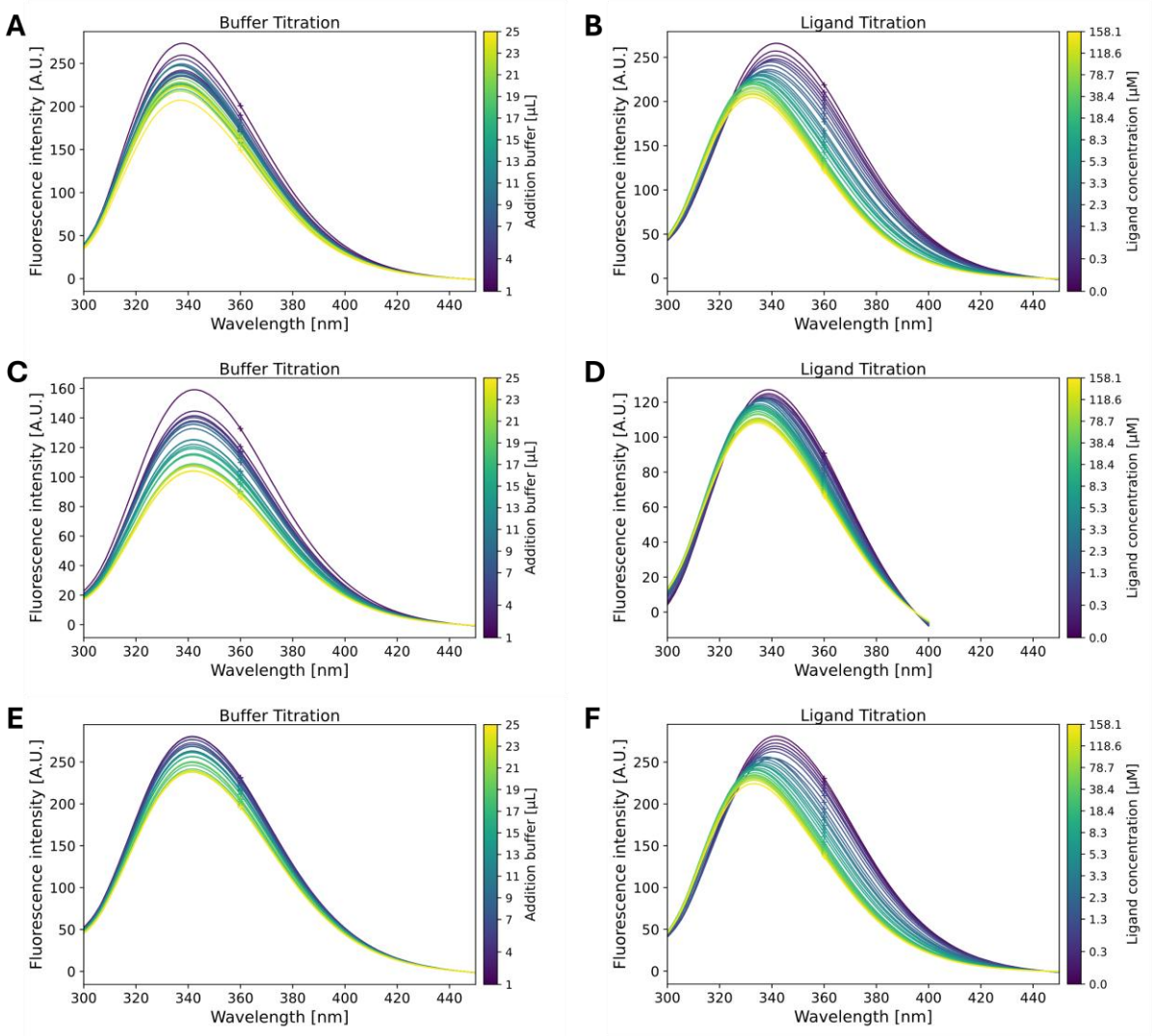


Figure 135. Fluorescence-based analysis of IPTG binding to variants of heterodimeric LacI. Fluorescence emission spectra were recorded at constant protein concentration after sequential buffer addition as a control and upon titration with increasing IPTG concentrations. The added buffer volume or IPTG concentration is indicated by the colour scale. A) B) LacI-A2-SpyTag, C) D) LacI*-SpyCatcher, E) F) Het-SpyTagxCatcher. Measurements were performed in Buffer C using a protein concentration of 0.2 mg·mL^{-1}.

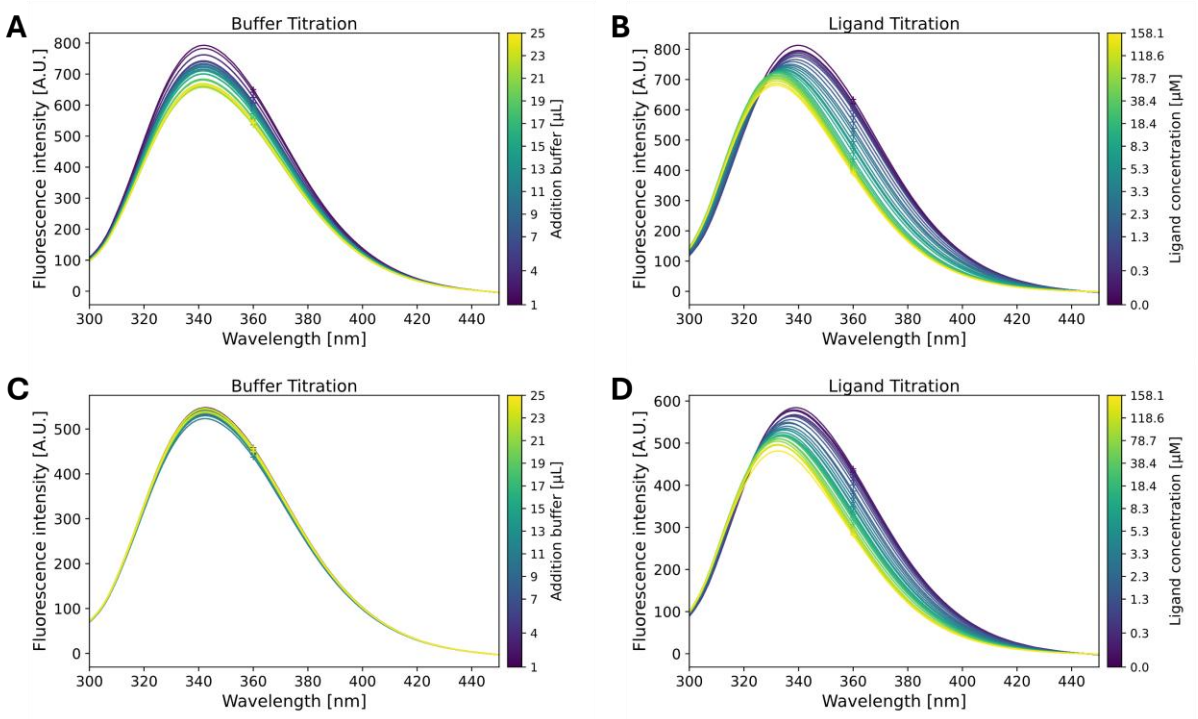


Figure 136. Fluorescence-based analysis of IPTG binding to LacI extension variants. Fluorescence emission spectra were recorded at constant protein concentration after sequential buffer addition as a control and upon titration with increasing IPTG concentrations. The added buffer volume or IPTG concentration is indicated by the colour scale. A) B) LacI-Lev1, C) D) LacI-Lev2. Measurements were performed in Buffer C using a protein concentration of 0.2 mg·mL^{-1}.

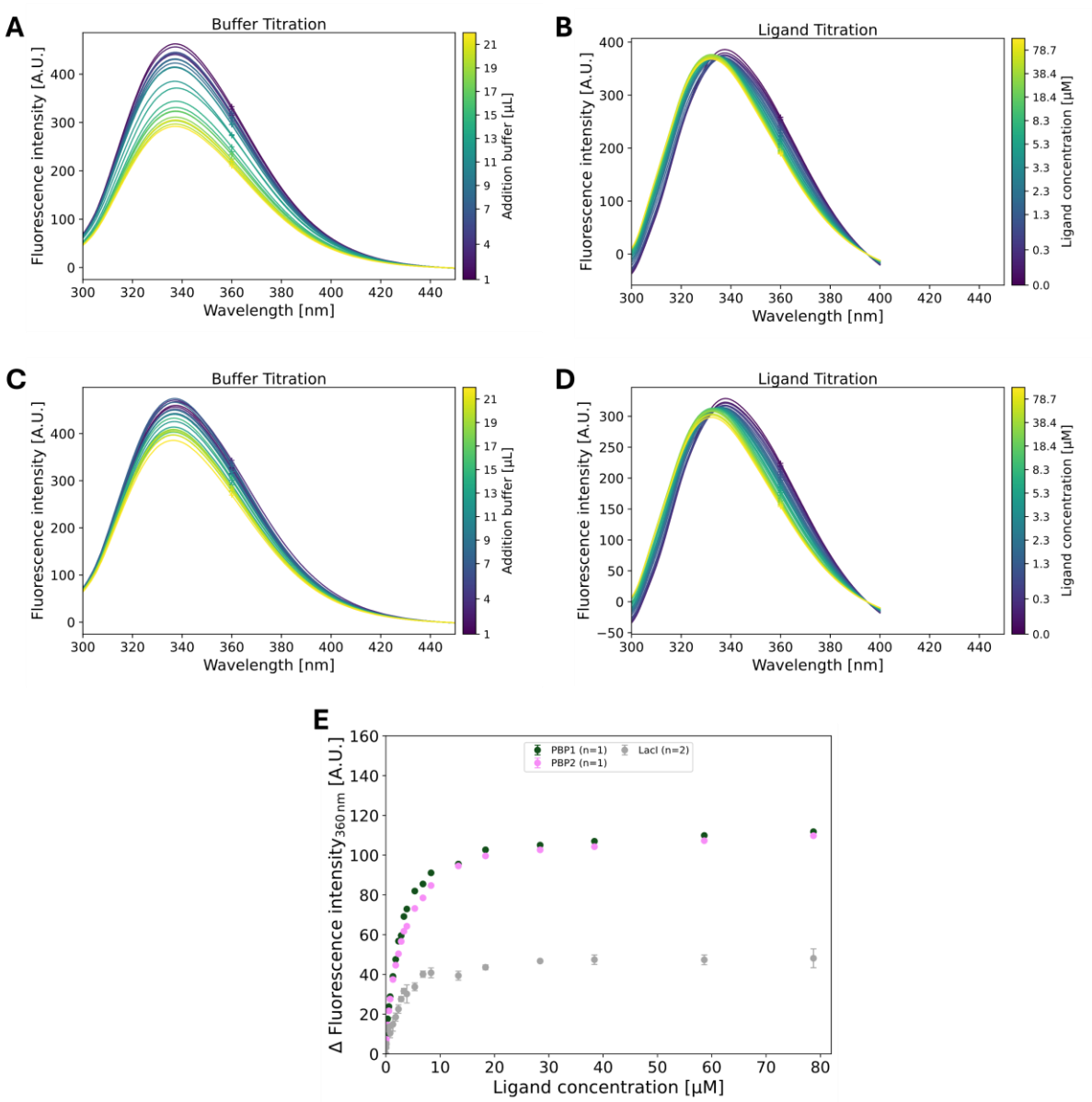
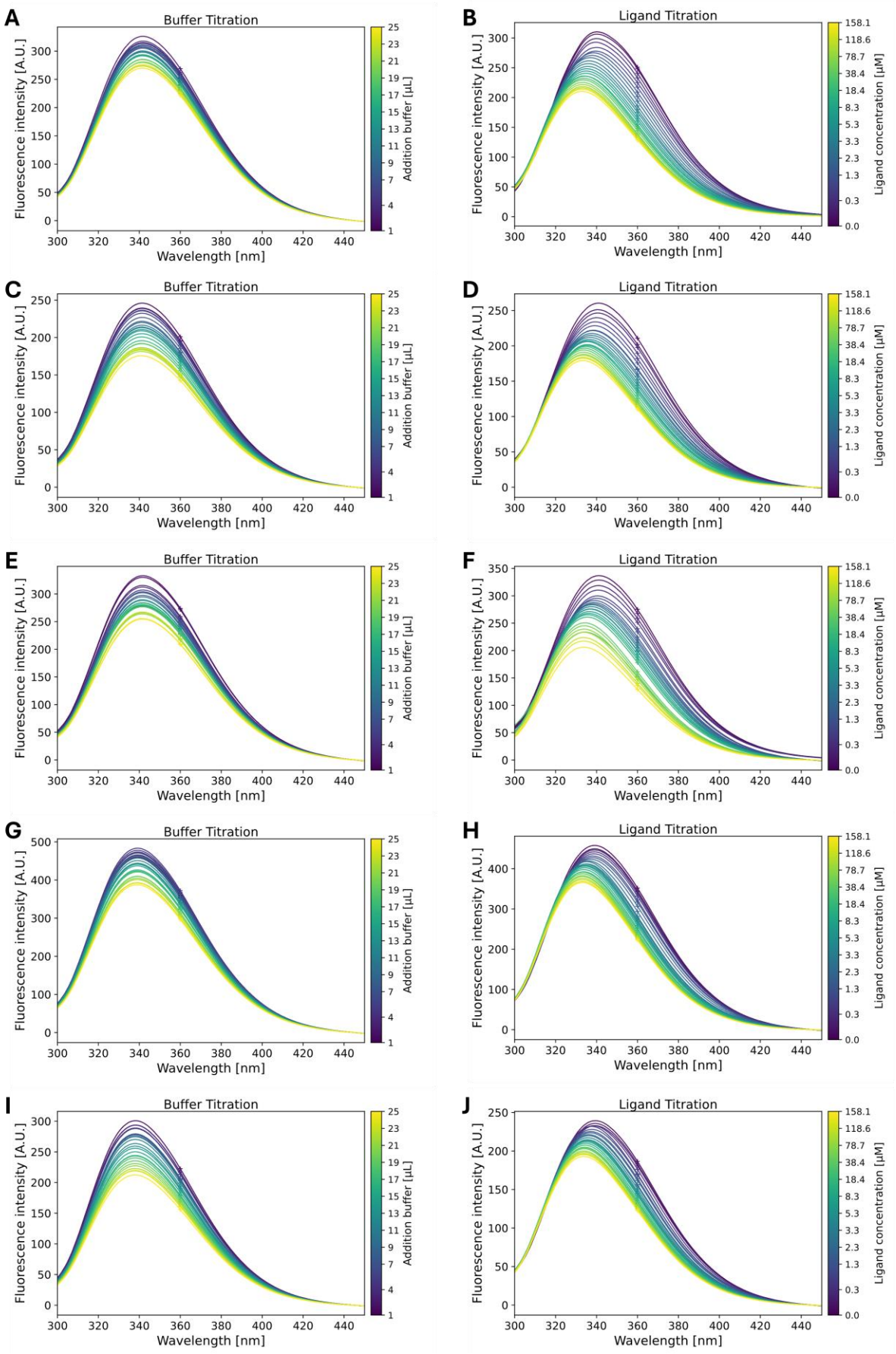


Figure 137. Fluorescence-based analysis of IPTG binding to truncated LacI extension variants lacking the DNA binding domain. Fluorescence emission spectra were recorded at constant protein concentration after sequential buffer addition as a control and upon titration with increasing IPTG concentrations. The added buffer volume or IPTG concentration is indicated by the colour scale. A) B) PBP1, C) D) PBP2, (E) IPTG-dependent change in fluorescence intensity at 360 nm plotted against IPTG concentration for PBP1 (green), PBP2 (pink), and LacI (grey). Data points represent mean values, and error bars indicate the standard deviation of the indicated number of replicates. Measurements were performed in Buffer C using a protein concentration of 0.2 mg·mL^{-1}.


A
Buffer Titration
B
Ligand Titration
C
Buffer Titration
D
Ligand Titration
E
Buffer Titration
F
Ligand Titration
G
Buffer Titration
H
Ligand Titration
I
Buffer Titration
J
Ligand Titration
Fluorescence intensity [A.U.]
Wavelength [nm]
Addition buffer [µL]
Ligand concentration [µM]

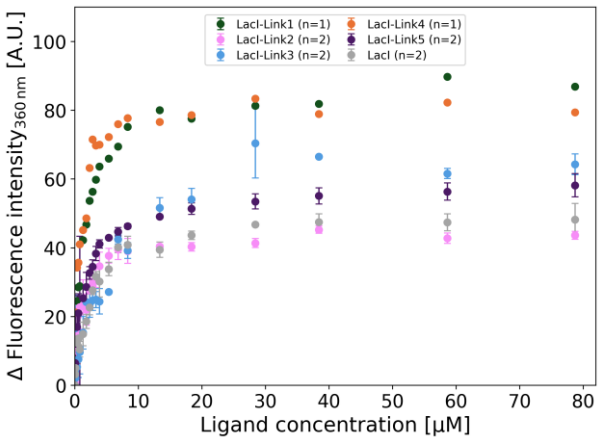


Figure 138. Fluorescence-based analysis of IPTG binding to LacI extension variants with DogTag. Fluorescence emission spectra were recorded at constant protein concentration after sequential buffer addition as a control and upon titration with increasing IPTG concentrations. The added buffer volume or IPTG concentration is indicated by the colour scale. A), B) LacI-Link1, C), D) LacI-Link2, E), F) LacI-Link3, G), H) LacI-Link4, I), J) LacI-Link5, K) IPTG-dependent change in fluorescence intensity at 360 nm plotted against IPTG concentration for LacI Linker variants (LacI-Link1 (green), LacI-Link2 (pink), LacI-Link3 (blue), LacI-Link4 (orange), LacI-Link5 (purple)) and LacI (grey). Data points represent the mean of the indicated number of replicates, and error bars indicate the standard deviation. Measurements were performed in Buffer C using a protein concentration of 0.2 mg·mL^{-1}.

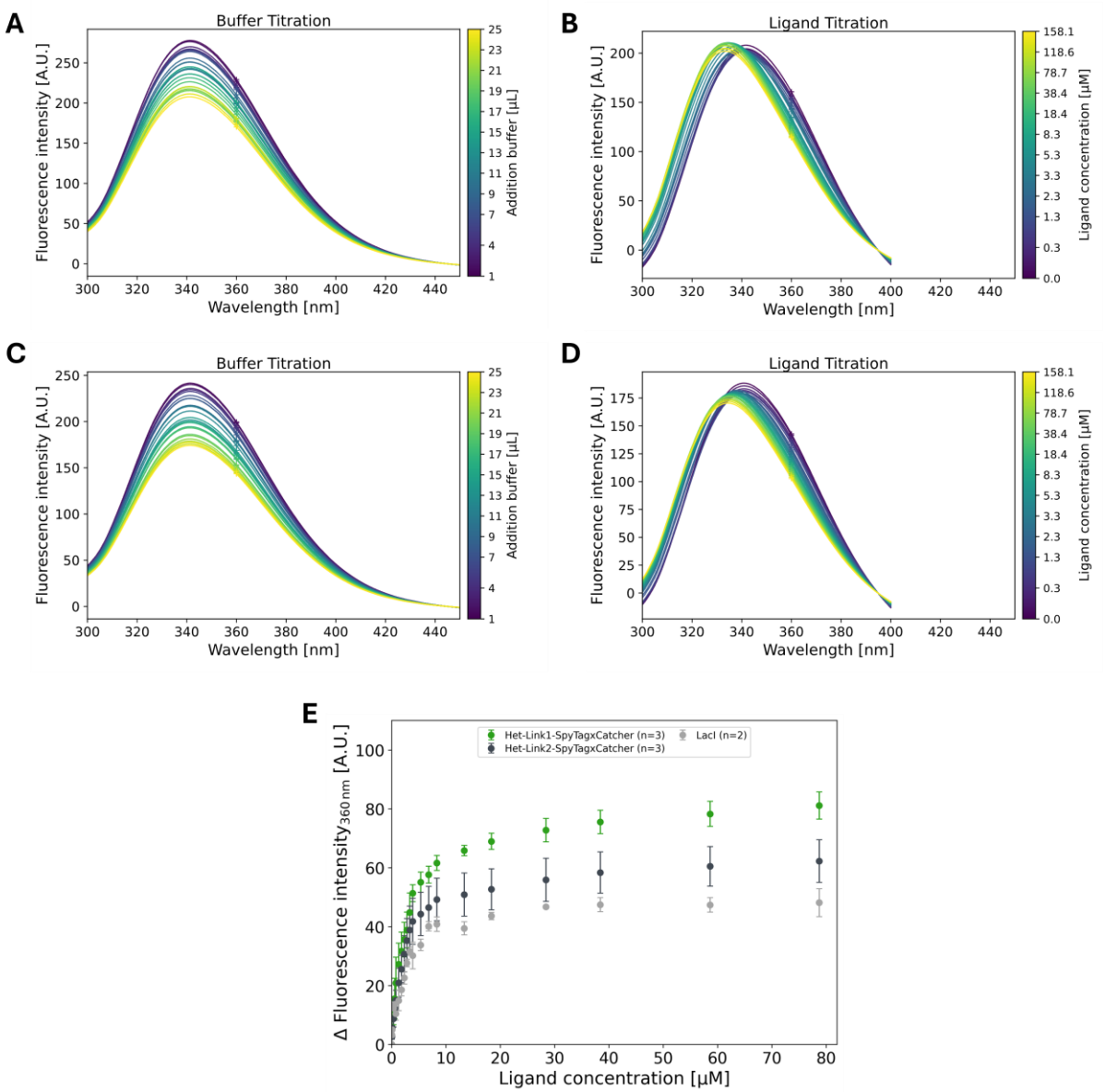


Figure 139. Fluorescence-based analysis of IPTG binding to heterodimeric Linker variants. Fluorescence emission spectra were recorded at constant protein concentration after sequential buffer addition as a control and upon titration with increasing IPTG concentrations. The added buffer volume or IPTG concentration is indicated by the colour scale. A), B) Het-Link1-SpyTagxCatcher, C), D) Het-Link2-SpyTagxCatcher. E) IPTG-dependent change in fluorescence intensity at 360 nm plotted against IPTG concentration for heterodimeric LacI Linker variants (Het-Link1-SpyTagxCatcher (green), Het-Link2-SpyTagxCatcher (dark grey)) and LacI (grey). Data points represent the mean of the indicated number of replicates, and error bars indicate the standard deviation. Measurements were performed in Buffer C using a protein concentration of 0.2 mg·mL^{-1}.

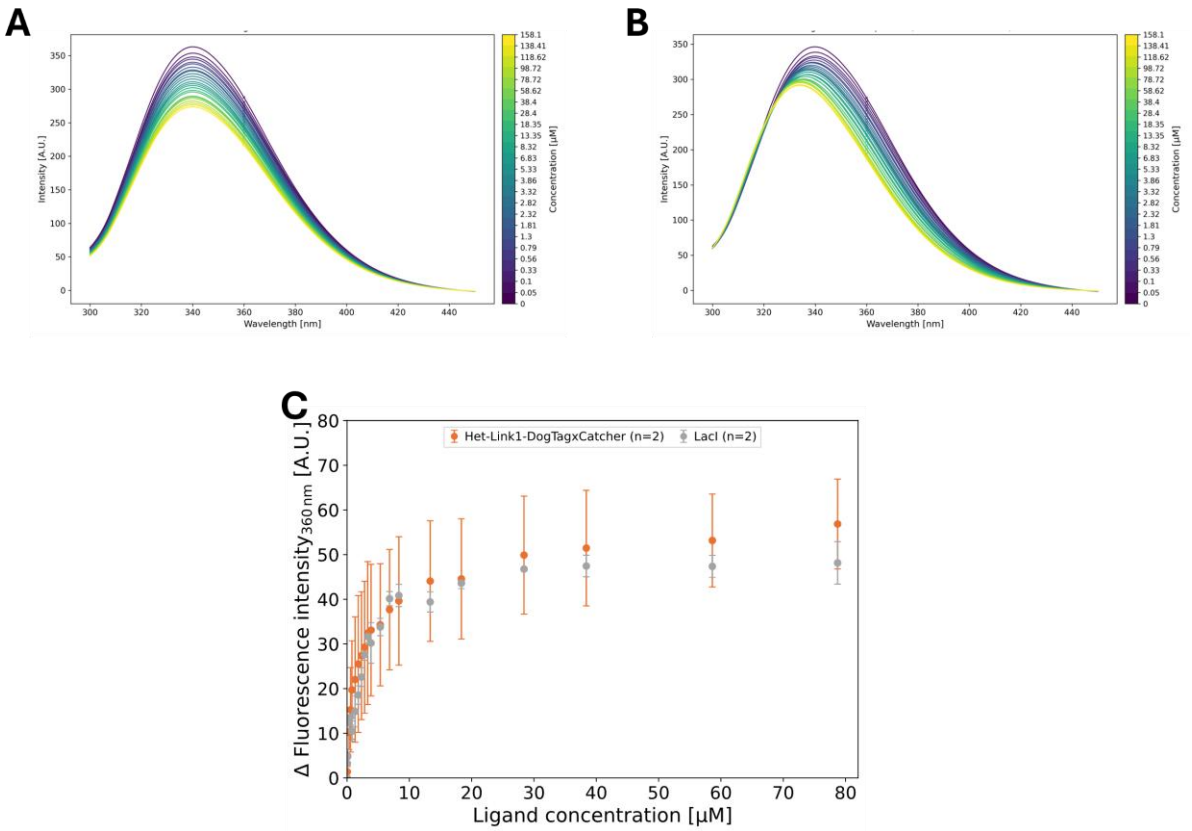


Figure 140. Fluorescence-based analysis of IPTG binding to heterodimeric Linker variants. Fluorescence emission spectra were recorded at constant protein concentration after sequential buffer addition as a control and upon titration with increasing IPTG concentrations. The added buffer volume or IPTG concentration is indicated by the colour scale. A), B) Het-Link1-DogTagxCatcher, C) IPTG-dependent change in fluorescence intensity at 360 nm for heterodimeric LacI Linker variants (Het-Link1-DogTagxCatcher (orange) and LacI (grey)). Data points represent the mean of the indicated number of replicates, and error bars indicate the standard deviation. Measurements were performed in Buffer C using a protein concentration of 0.2 mg·mL^{-1}.

7.4 Analysis of secondary structure content with circular dichroism spectroscopy

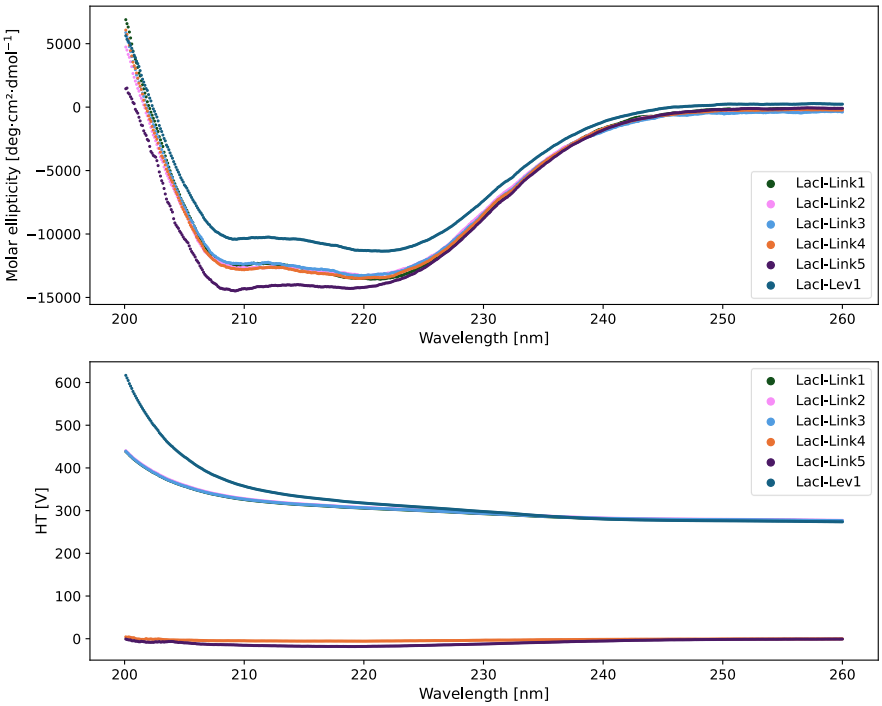


Figure 141. Circular dichroism analysis of LacI linker variants. The upper panel shows far-UV CD spectra of LacI-Link1 (green), LacI-Link2 (pink), LacI-Link3 (blue), LacI-Link4 (orange), LacI-Link5 (purple), and wild-type LacI (grey). The lower panel shows the corresponding high-tension (HT) voltage traces recorded during the measurements. HT voltage traces were not recorded for LacI-Link4 and LacI-Link5. Measurements were performed in Buffer C using a protein concentration of 0.2 mg·mL⁻¹.

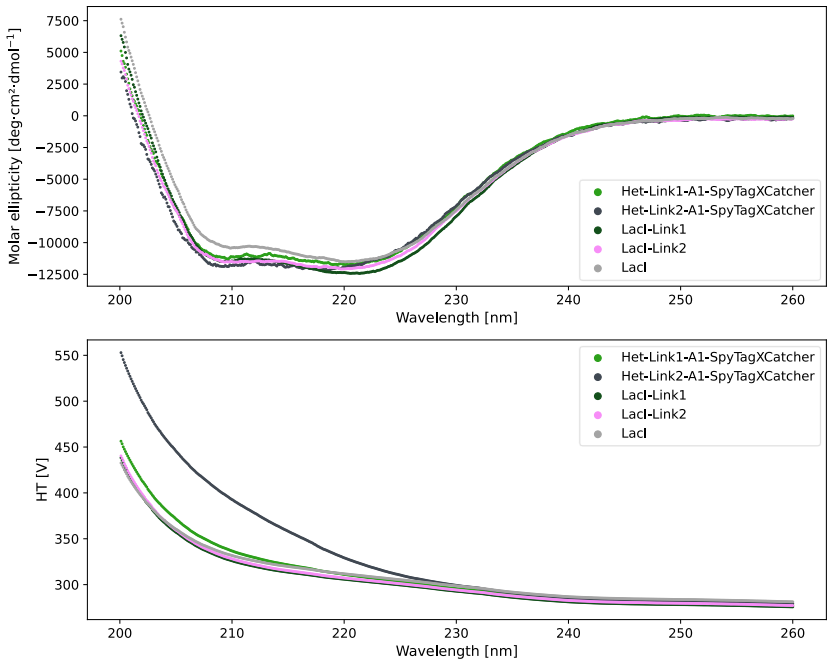


Figure 142. Circular dichroism analysis of heterodimeric LacI linker variants. Far-UV CD spectra of Het-Link1-SpyTag×Catcher (light green), Het-Link2-SpyTag×Catcher (dark grey), LacI-Link1 (dark green), LacI-Link2 (pink), and wild-type LacI (grey) are shown in the upper panel. The corresponding high-tension (HT) voltage traces are shown in the lower panel. Measurements were performed in Buffer C using a protein concentration of 0.2 mg·mL⁻¹.

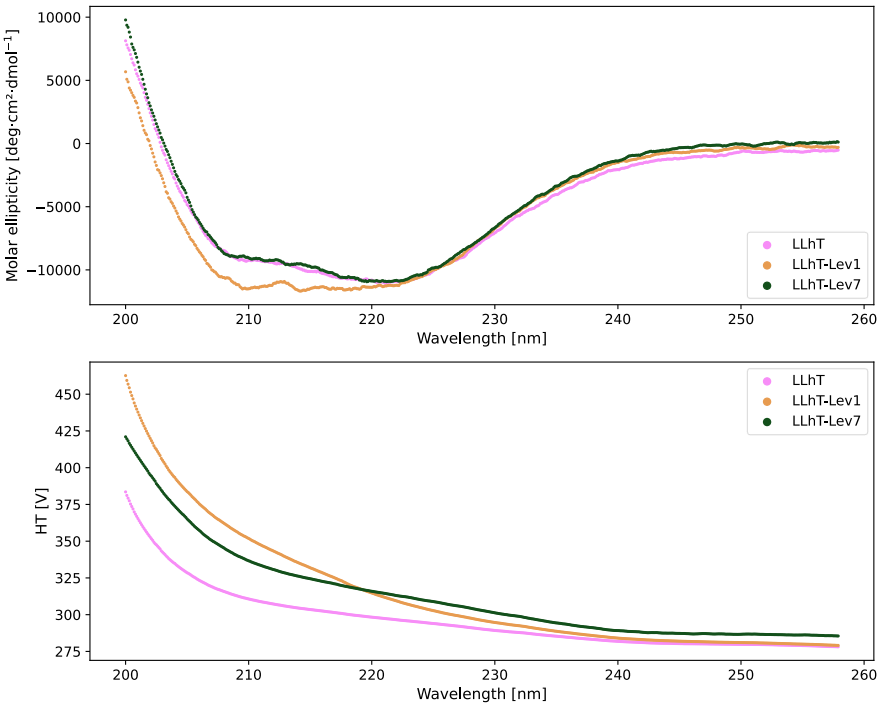


Figure 143. Circular dichroism analysis of LLhT structural extension variants. Far-UV CD spectra of LLhT-Lev1 (orange), LLhT-Lev7 (green), and LLhT (pink) are shown in the upper panel. The corresponding high-tension (HT) voltage traces are shown in the lower panel. Measurements were performed in Buffer C using a protein concentration of 0.2 mg·mL⁻¹.

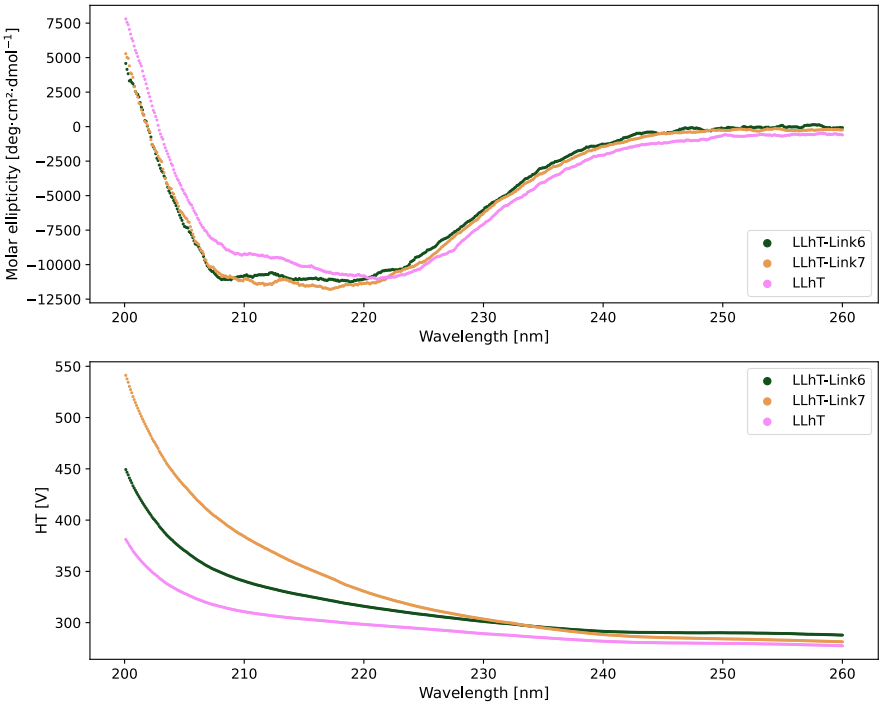


Figure 144. Circular dichroism analysis of LLhT linker variants. Far-UV CD spectra of LLhT-Link6 (green), LLhT-Link7 (orange), and LLhT (pink) are shown in the upper panel. The corresponding high-tension (HT) voltage traces are shown in the lower panel. Measurements were performed in Buffer C using a protein concentration of 0.2 mg·mL⁻¹.

7.5 Analysis of thermal stability evaluated with a change of molar ellipticity determined with circular dichroism spectroscopy

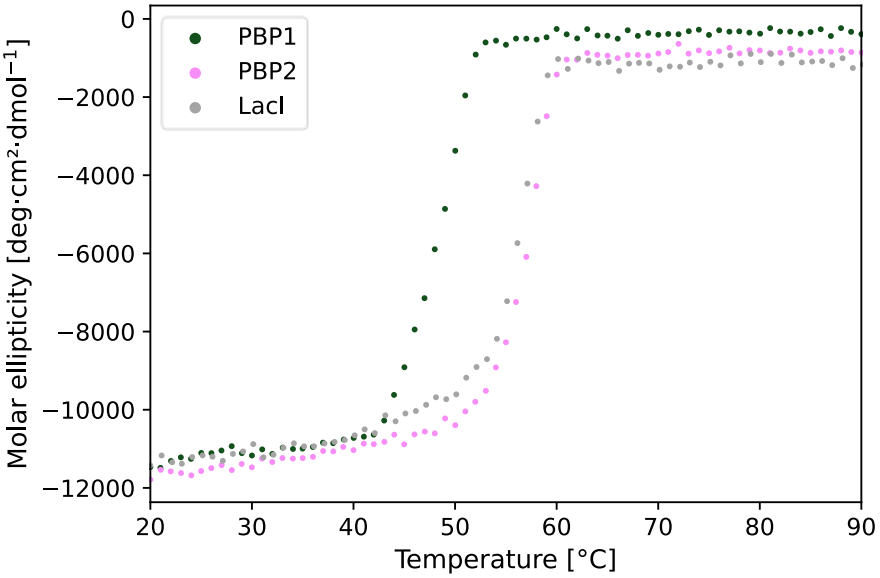


Figure 145. CD-based thermal unfolding analysis of truncated PBP variants and wild-type LacI. Thermal unfolding of PBP1 (green), PBP2 (pink), and wild-type LacI (grey) was monitored by the temperature-dependent change in molar ellipticity at 222 nm. Measurements were performed in Buffer C using a protein concentration of 0.2 mg·mL⁻¹.

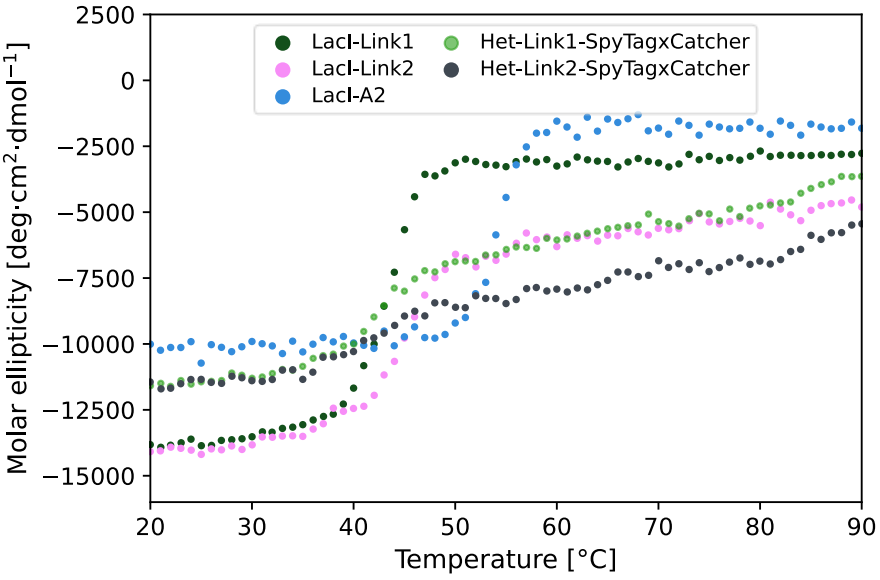


Figure 146. CD-based thermal unfolding analysis of heterodimeric LacI linker variants and their monomeric counterparts. Thermal unfolding of LacI-Link1 (dark green), LacI-Link2 (pink), LacI-A2 (blue), Het-Link1-SpyTag×Catcher (light green), Het-Link2-SpyTag×Catcher (dark grey), and wild-type LacI (grey) was monitored by the temperature-dependent change in molar ellipticity at 222 nm. The unfolding profiles illustrate the stabilising effect associated with SpyTag/SpyCatcher-mediated heterodimer formation. Measurements were performed in Buffer C using a protein concentration of 0.2 mg·mL⁻¹.

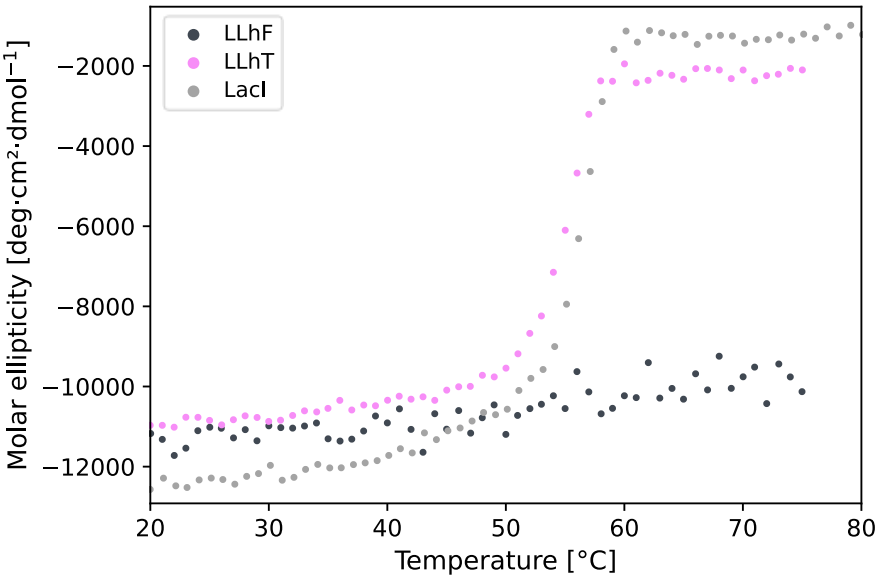


Figure 147. CD-based thermal unfolding analysis of chimeric LacI variants and wild-type LacI. Thermal unfolding of LLhF (dark grey), LLhT (pink), and wild-type LacI (grey) was monitored by the temperature-dependent change in molar ellipticity at 222 nm. Measurements were performed in Buffer C using a protein concentration of 0.2 mg·mL⁻¹.

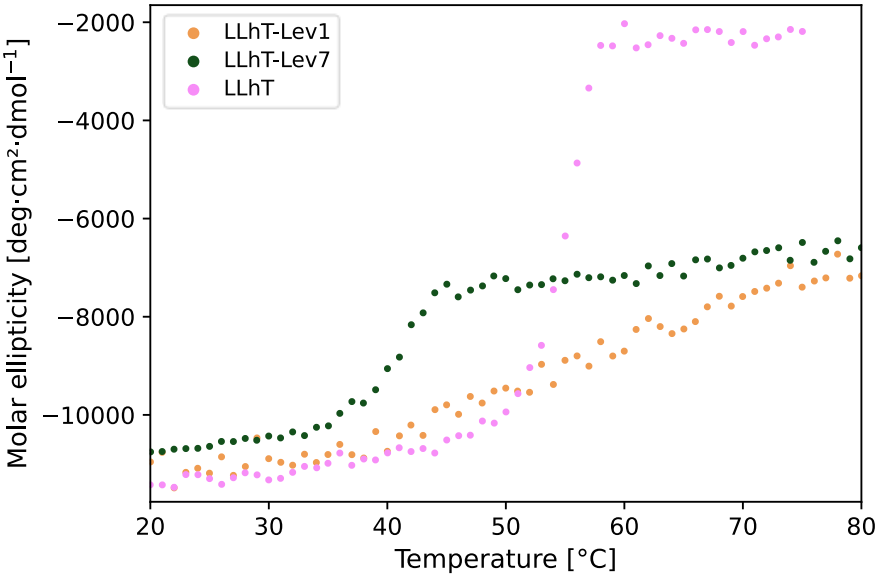


Figure 148. CD-based thermal unfolding analysis of LLhT extension variants and parental LLhT. Thermal unfolding of LLhT-Lev1 (orange), LLhT-Lev2 (green), and parental LLhT (pink) was monitored by the temperature-dependent change in molar ellipticity at 222 nm. Measurements were performed in Buffer C using a protein concentration of 0.2 mg·mL⁻¹.

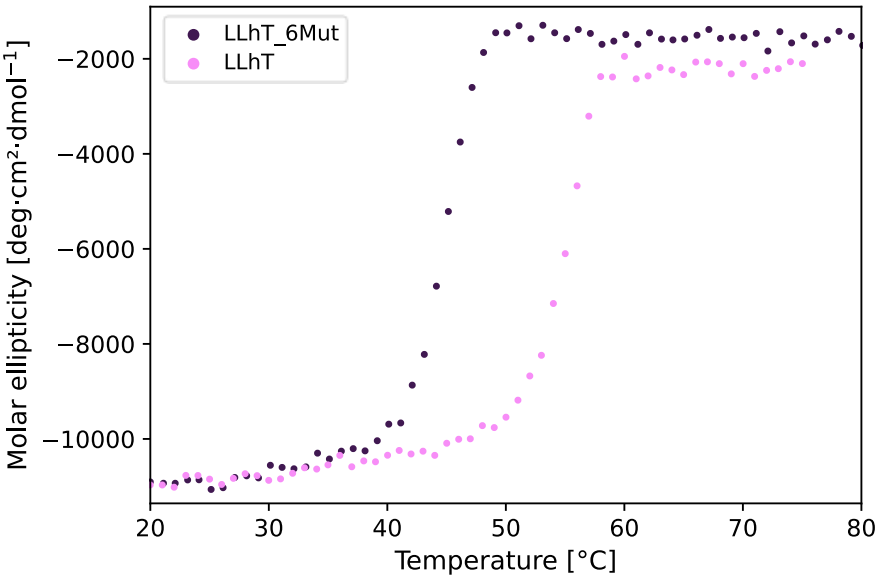


Figure 149. CD-based thermal unfolding analysis of monomeric LLhT and parental LLhT. Thermal unfolding of LLhT-6Mut (purple) and parental LLhT (pink) was monitored by the temperature-dependent change in molar ellipticity at 222 nm. Measurements were performed in Buffer C using a protein concentration of 0.2 mg·mL⁻¹.

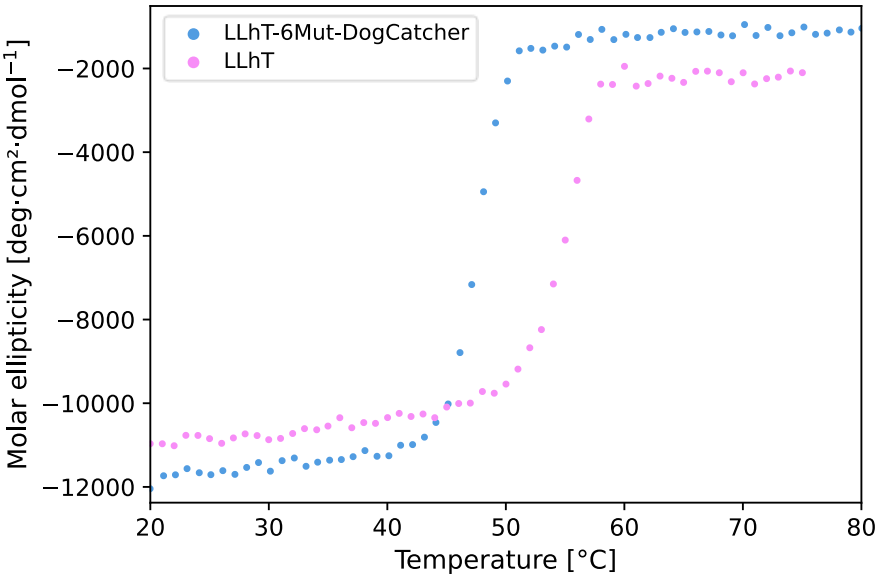


Figure 150. CD-based thermal unfolding analysis of monomeric LLhT-DogCatcher and parental LLhT. Thermal unfolding of LLhT-6Mut-DogCatcher (blue) and parental LLhT (pink) was monitored by the temperature-dependent change in molar ellipticity at 222 nm. Measurements were performed in Buffer C using a protein concentration of 0.2 mg·mL⁻¹.

Table 40. Apparent melting temperatures (T_{Melt}) derived from CD-based thermal unfolding experiments for LacI heterodimer with Linker extension and LLhT based variants. Values were determined as the transition midpoint of the temperature-dependent change in molar ellipticity at 222 nm.

Construct	**T_M [°C]**
LacI-Link1	43.1
LacI-Link2	45.5
Het-Link1-SpyTagxCatcher	45.6
Het-Link2-SpyTagxCatcher	47.5
LLhT	54.6
LLhF	n.d.
LLhT-Lev1	n.d.
LLhT-Lev7	42.2
LLhT-6Mut	44.5
LLhT-6Mut-DogCatcher	44.5

7.6 Thermal unfolding and Ligand binding analysis with Differential scanning calorimetry

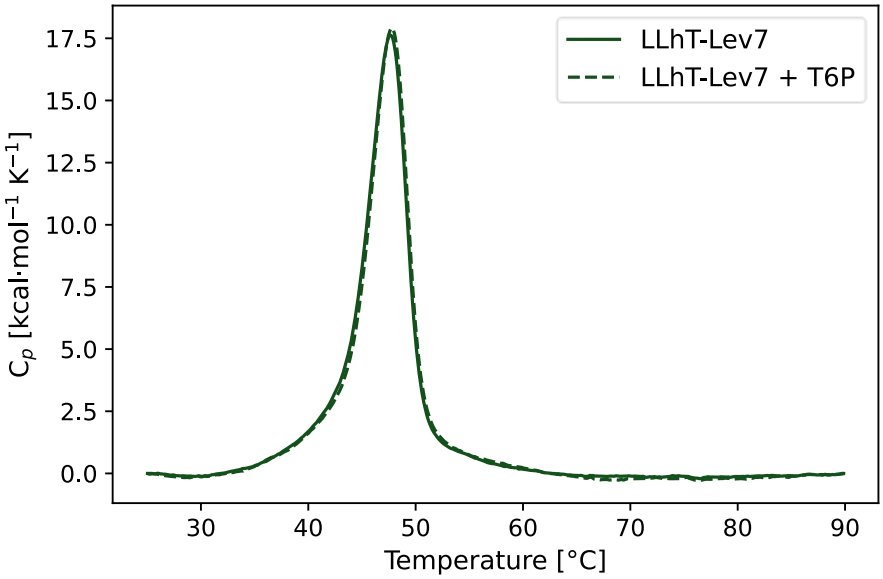


Figure 151. Differential scanning calorimetry analysis of LLhT-Lev7 in the absence and presence of T6P. DSC profiles of LLhT-Lev7 recorded without ligand (solid line) and in the presence of T6P (dashed line) are shown. Measurements were performed in Buffer C using a protein concentration of 0.5 mg·mL^{-1}.

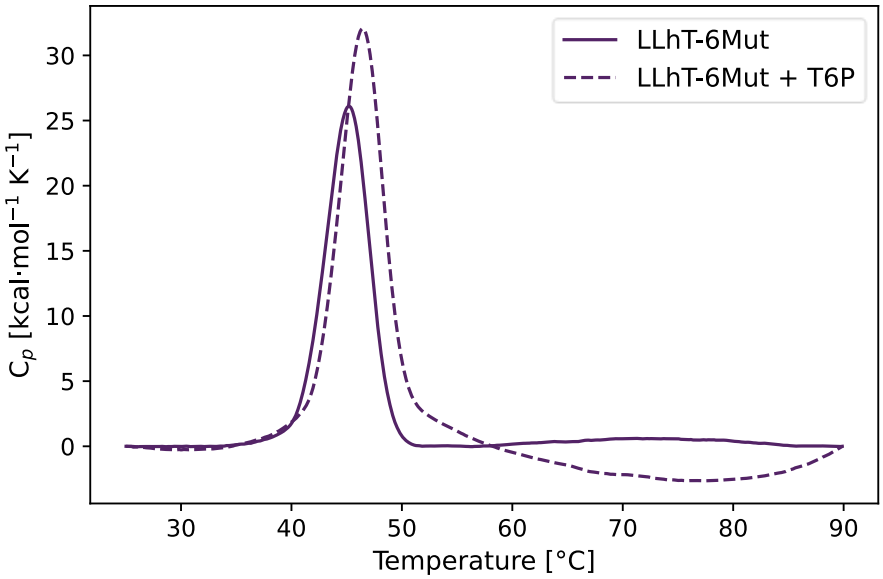


Figure 152. Differential scanning calorimetry analysis of LLhT-6Mut in the absence and presence of T6P. DSC profiles of LLhT-6Mut recorded without ligand (solid line) and in the presence of T6P (dashed line) are shown. Measurements were performed in Buffer C using a protein concentration of 0.5 mg·mL^{-1}.

Table 41. Thermal transition temperatures and unfolding enthalpies of LacI, LacI-Gentio and LLhT-variants determined by DSC. Listed thermal transition temperatures (T_M) and unfolding enthalpies (ΔH) were derived from DSC profiles recorded.

Construct	**T_M [°C]**	**ΔH [kcal·mol^{-1}]**
LacI	56.4	145
LacI + IPTG	70.4	211
LacI + Gentiobiose	60.5	170
LacI-Gentio	62.4	142
LacI-Gentio + Gentiobiose	62.1	114
LacI	57.7	152
LacI + IPTG	69.4	199
LacI + Trehalose-6-P	57.8	153
LLhT	57.9	114
LLhT + IPTG	57.6	74
LLhT + Trehalose-6P	60.3	293
LLhT-Lev7	47.6	100
LLhT-Lev7 + Trehalose-6P	47.8	100
LLhT-6Mut	45.2	131
LLhT-6Mut + Trehalose-6P	46.4	177

7.7 SEC-MALS evaluations determining the molecular weight

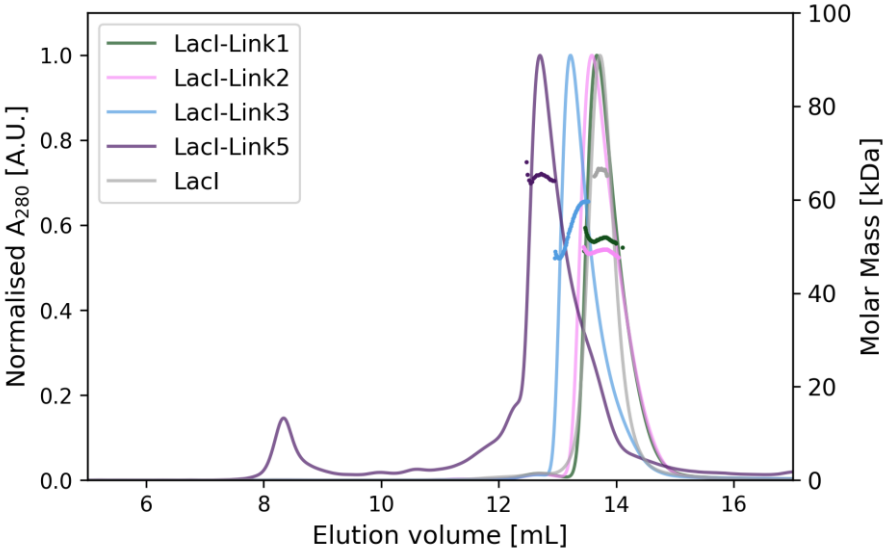


Figure 153. SEC-MALS characterisation of LacI-linker variants. Measurements were performed using an analytical Superdex 200 10/300 GL column. Absorbance at 280 nm is shown as a function of elution volume for LacI-Link1 (green), LacI-Link2 (pink), LacI-Link3 (blue), LacI-Link5 (purple), and wild-type LacI (grey), with the corresponding MALS-derived molar mass plotted on the secondary axis. The expected molecular weights are 88 kDa for the dimeric LacI linker variants and 72 kDa for dimeric wild-type LacI. Measurements were performed in Buffer C using a protein concentration of 0.5 mg·mL^{-1}.

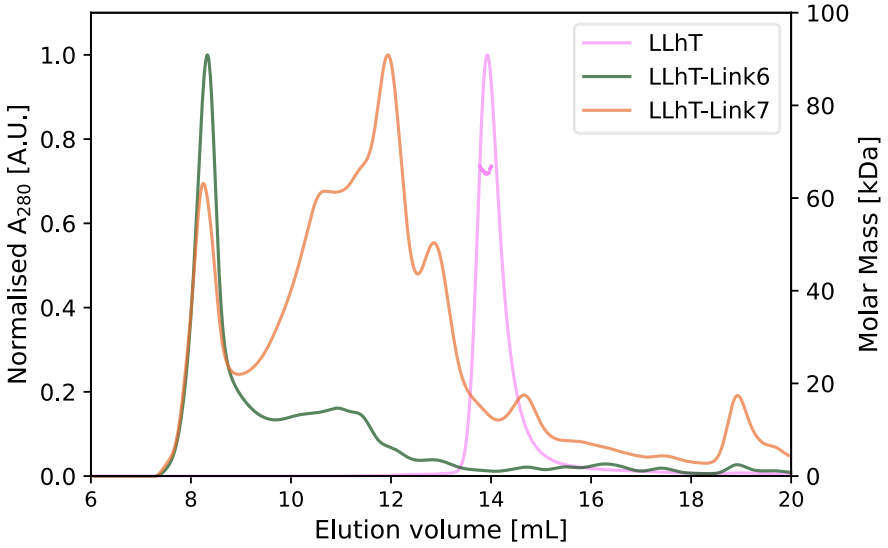


Figure 154. SEC-MALS analysis of LLhT DogTag insertion variants. Measurements were performed using an analytical Superdex 200 10/300 GL column. Normalised absorbance at 280 nm is plotted against elution volume for LLhT (pink, MW: 78 kDa), LLhT-Link6 (green, MW: 92 kDa), and LLhT-Link7 (orange, MW: 96 kDa). The corresponding MALS-derived molar masses are shown on the secondary axis, allowing comparison of the oligomeric state and hydrodynamic behaviour of the engineered variants with the parent protein. Measurements were performed in Buffer C using a protein concentration of 0.5 mg·mL^{-1}.

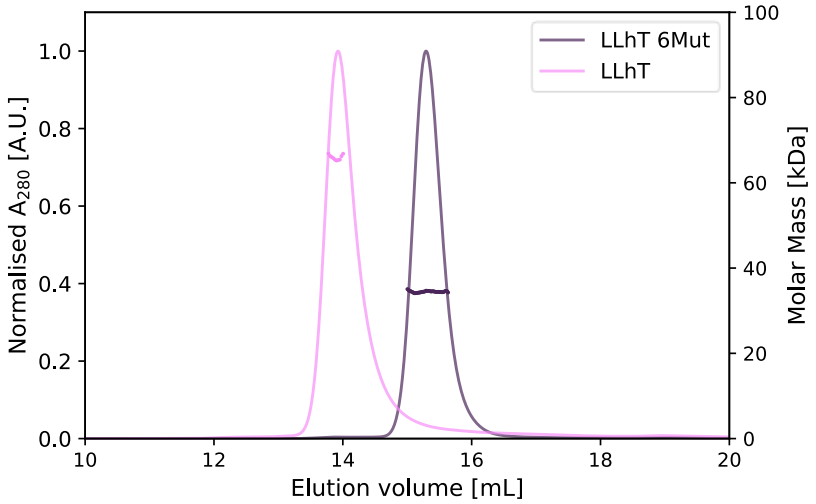


Figure 155. SEC-MALS characterisation of interface mutations of LLhT. Measurements were performed using an analytical Superdex 200 10/300 GL column. Absorbance at 280 nm is shown as a function of elution volume for LLhT-6Mut (purple, expected MW: 39 kDa) and parental LLhT (pink, expected MW: 78 kDa), with the corresponding MALS-derived molar mass plotted on the secondary axis. Measurements were performed in Buffer C using a protein concentration of 0.5 mg·mL^{-1}.

Table 42. Molecular weights determined by SEC-MALS for variants generated to achieve a heterodimeric LacI (Chapter 2.2).Expected molecular weights and corresponding oligomeric states are listed together with the experimentally determined molecular weights and associated measurement errors.

Construct	**Experimental MW [kDa]**	**Expected MW [kDa]**
LacI-WT	64.4 ± 0.2%	Dimer: 73.1
LacI-A1	34.9 ± 0.2%	Monomer: 40.7
LacI-B1	40.6 ± 2.7%	Monomer: 36.5
	70.0 ± 0.1%	
Co-expressed LacI-A1-B1	39.8 ± 0.5%	Heterodimer: 77.2
	93.1 ± 1.3%	
LacI-A2	36.2 ± 2.0%	Monomer: 37.4
	61.8 ± 1.7%	
LacI-B2	36.9 ± 4.5%	Monomer: 36.6
	74.6 ± 5.1%	
LacI-B3	39.4 ± 4.0%	Monomer: 36.5
	72.5 ± 2.1%	
LacI-splitGFP	190.4 ± 0.7%	Heterodimer: 105
		LacI-A2-splitGFP(10-11): 46.8
		LacI-B2-splitGFP(1-9): 61.4
LacI-Monomer-cleavedGFP	35.5 ± 2.7%	Monomer LacI-A1-cleav.: 39.1
	82.0 ± 2.6%	Monomer LacI-A1-cleav.: 39.3
LacI-Dimer-cleavedGFP	73.3 ± 2.0%	Dimer: 78.4
LacI-SpyTag	39.7 ± 1.5%	Monomer: 39.4
LacI-SpyCatcher	45.9 ± 1.3%	Monomer: 48.2
Het-SpyTagxCatcher	88.7 ± 1.4%	Heterodimer: 87.6

Table 43. Molecular weights determined by SEC-MALS for variants generated to generate a heterodimeric foot with LacI structural extension (Chapter 2.3 and Chapter 2.4). Expected molecular weights and corresponding oligomeric states are listed together with the experimentally determined molecular weights and associated measurement errors.

Construct	**Experimental MW [kDa]**	**Expected MW [kDa]**
LacI-Lev1	65.9 ± 6.0%	Monomer: 39.1
	50.1 ± 4.0%	
LacI-Lev2	72.7 ± 0.8%	Monomer: 38.8
LacI-Link1	51.5 ± 0.6%	Monomer: 44.1
LacI-Link2	49.2 ± 0.6%	Monomer: 44.8
LacI-Link3	54.3 ± 0.6%	Monomer: 43.9
LacI-Link4	n.d.	Monomer: 43.8
LacI-Link5	64.8 ± 0.8%	Monomer: 44.7
Het-Link1-SpyTagxCatcher	85.0 ± 3.6%	Heterodimer: 93.6
Het-Link2-SpyTagxCatcher	97.1 ± 1.1%	Heterodimer: 92.8
LacI*-DogCatcher	1st peak	Monomer: 48.8
	49.8 ± 1.7%	
	91.4 ± 1.9%	
	2nd peak	
	46.8 ± 0.6%	
	88.0 ± 1.2%	
LacI-A2-Link1-SpyCatcher-DogTag	1st peak	Monomer: 54.2
	51.1 ± 0.5%	
	99.3 ± 0.5%	
	2nd peak	
	51.4 ± 0.5%	
	95.1 ± 2.0%	
Het-Link1-DogTagxCatcher	99.5 ± 2.0%	Heterodimer: 103.1

Table 44. Molecular weights determined by SEC-MALS for variants generated to generate a heterodimeric foot based on LLhT (Chapter 2.5). Expected molecular weights and corresponding oligomeric states are listed together with the experimentally determined molecular weights and associated measurement errors.

Construct	Experimental MW [kDa]	Expected MW [kDa]
LLhF	83.6 ± 0.8%	Monomer: 40.0
LLhT	66.4 ± 1.3%	Monomer: 39.8
LLhT-Lev1	n.d.	Monomer: 39.0
LLhT-Lev7	66.4 ± 1.3%	Monomer: 36.0
LLhT-Link6	n.d.	Monomer: 48.8
LLhT-Link7	n.d.	Monomer: 46.5
LLhT-6Mut	34.5 ± 0.6%	Monomer: 39.8
LLhT-6Mut-DogCatcher	72.8 ± 0.1%	Monomer: 46.0

Table 45. Molecular weights determined by SEC-MALS for the clapping machine (Chapter 2.6). Expected molecular weights and corresponding oligomeric states are listed together with the experimentally determined molecular weights and associated measurement errors.

Construct	Experimental MW [kDa]	Expected MW [kDa]
CM-PotF- 1stPeak	98.1 ± 4.6%	Monomer: 55.6
	50.9 ± 3.0%	
CM-PotF- 2ndPeak	92.9 ± 2.4%	Monomer: 55.6
	47.9 ± 4.8%	

Table 46. Molecular weights determined by SEC-MALS for the assembly of the motor protein (Chapter 2.7).Expected molecular weights and corresponding oligomeric states are listed together with the experimentally determined molecular weights and associated measurement errors.

Construct	Experimental MW [kDa]	Expected MW [kDa]
CM-FootA-Link1	139.8 ± 1.1%	Assembly: 149.1
	53.5 ± 0.6%	
CM-FootA-Link2	145.1 ± 0.2%	Assembly: 148.3
	54.4 ± 0.5%	
CM-FootB	121.7 ± 0.5%	Assembly: 158.6
	57.5 ± 0.7%	
CM-FootA-Link1-FootB	340.8 ± 0.1%	Assembly: 251.4
	239.9 ± 0.2 %	
	125.3 ± 0.5%	
	56.5 ± 0.5%	
CM-FootA-Link2-FootB	345.6 ± 0.4%	Assembly: 250.6
	247.1 ± 0.4%	
	130.1 ± 0.6%	
	57.7 ± 0.7%	

7.8 DNA binding analysis with SPR

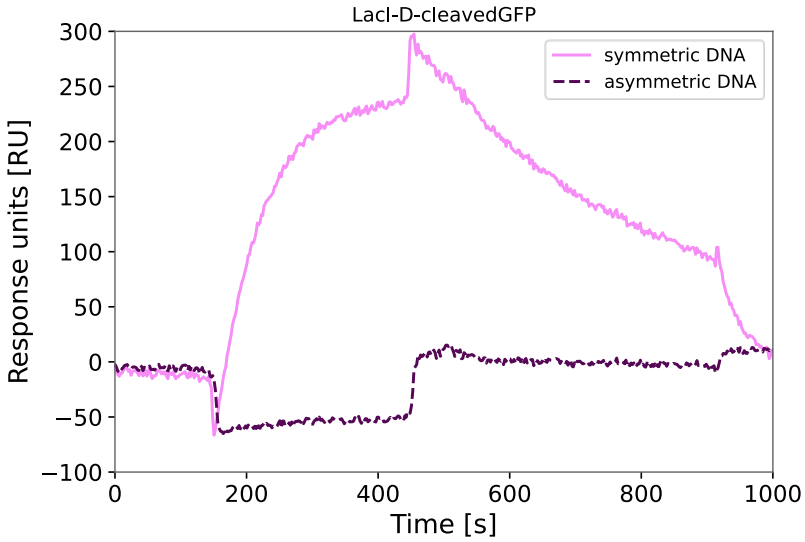


Figure 156. SPR analysis of LacI-D-cleavedGFP binding to symmetric DNA (pink) and asymmetric DNA (purple). Sensorgrams were recorded at room temperature with a constant flow rate of 30 µL·min⁻¹. Double-stranded oligonucleotides containing the cognate symmetric operator sequence were immobilised on the sensor surface, and protein variants were injected at a concentration of 1 µM. Association was monitored for 60 s, followed by a dissociation phase of at least 60 s. Measurements were conducted by Vanessa Troßmann.

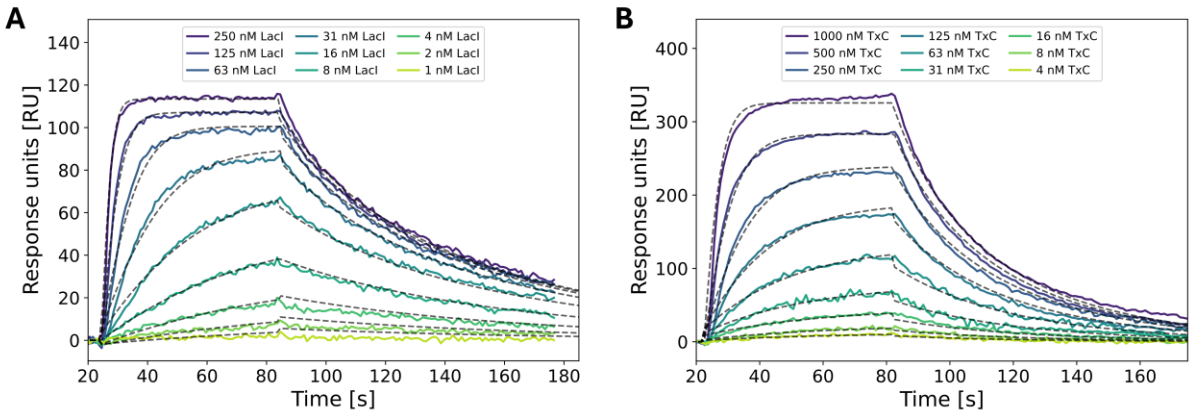


Figure 157. Concentration-dependent SPR analysis of LacI and Het-SpyTag×SpyCatcher binding to symmetric and asymmetric operator DNA. Sensorgrams were recorded at a constant flow rate of 30 µL·min⁻¹ at room temperature. Protein variants were injected over immobilised double-stranded oligonucleotides containing either the cognate symmetric operator site or an asymmetric operator site using two-fold serial dilutions. Protein association was monitored for 60 s, followed by a dissociation phase of at least 60 s. Reference- and buffer-corrected sensorgrams were fitted using a 1:1 binding model to determine the kinetic parameters k_{on}, k_{off} and K_D. Measurements were conducted by Vanessa Troßmann. Dotted lines represent the fit. A) SPR sensorgrams of LacI binding to symmetric operator DNA recorded using serial dilutions from 250 nM to 1 nM. B) SPR sensorgrams of Het-SpyTag×Catcher binding to asymmetric operator DNA recorded using serial dilutions from 1000 nM to 4 nM.

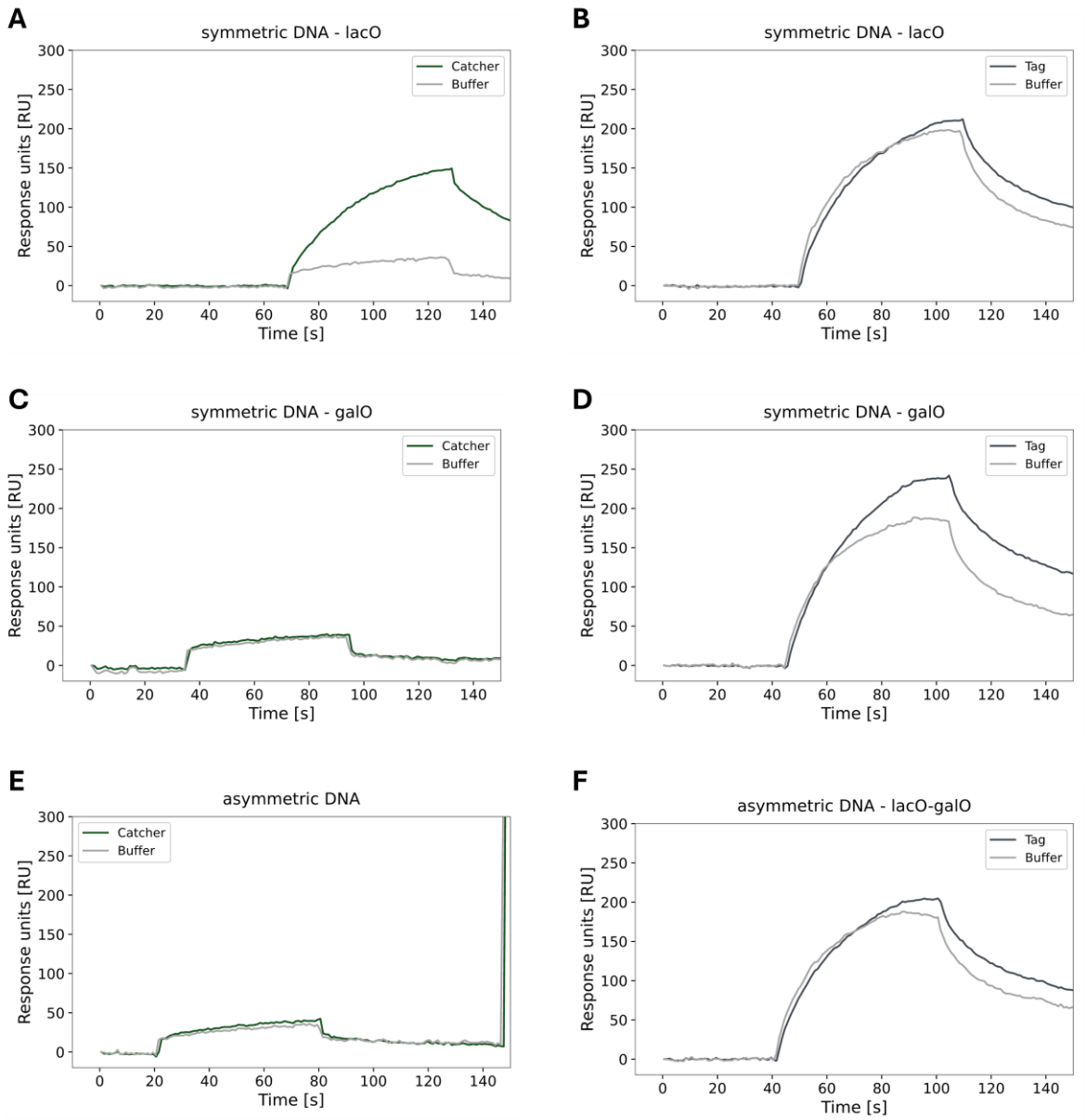


Figure 158. SPR analysis of LacI*-SpyCatcher and LacI-A2-SpyTag binding to different operator DNA sequences. Sensorgrams were recorded at room temperature with a constant flow rate of 30 µL·min⁻¹. LacI*-SpyCatcher (A, C, E) and LacI-A2-SpyTag (B, D, F) and were injected at a concentration of 1 µM over immobilised double-stranded oligonucleotides containing different operator arrangements. Measurements were conducted by Vanessa Troßmann. Protein-containing sample curves are shown in orange, while buffer/reference curves are shown in grey. A, B) Symmetric lacO DNA. C,D) Symmetric galO DNA. E,F) Asymmetric lacO–galO DNA.

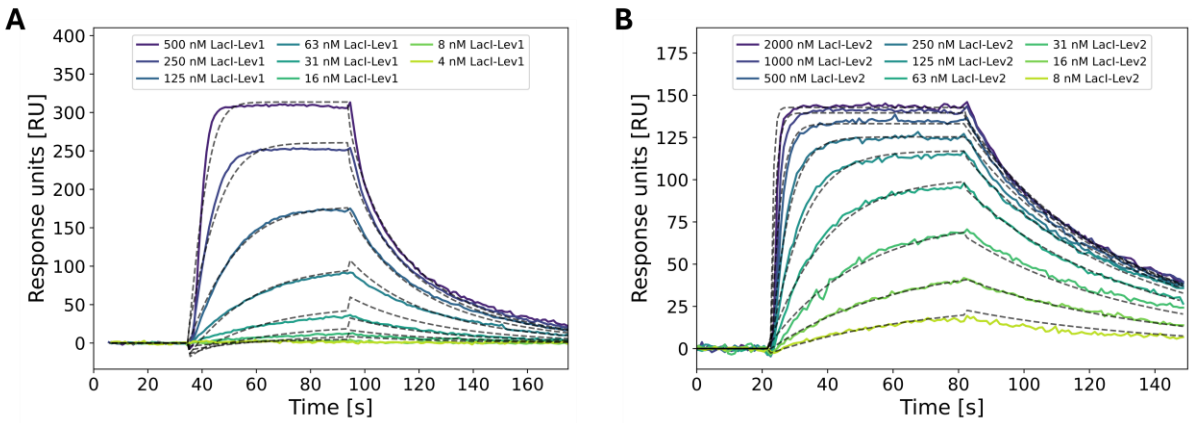


Figure 159. Concentration-dependent SPR analysis of LacI-Lev1 and LacI-Lev2 binding to symmetric operator DNA. Sensorgrams were recorded at a constant flow rate of 30 µL·min^{-1} at room temperature. Protein variants were injected over immobilised double-stranded oligonucleotides containing the cognate symmetric operator site using two-fold serial dilutions. Protein association was monitored for 60 s, followed by a dissociation phase of at least 60 s. Reference- and buffer-corrected sensorgrams were fitted using a 1:1 binding model to determine the kinetic parameters k_{on}, k_{off} and K_D. Measurements were conducted by Vanessa Troßmann. Dotted lines represent the fit. A) SPR sensorgrams of LacI-Lev1 using serial dilutions from 500 nM to 4 nM. B) SPR sensorgrams of LacI-Lev2 recorded using serial dilutions from 2000 nM to 8 nM.

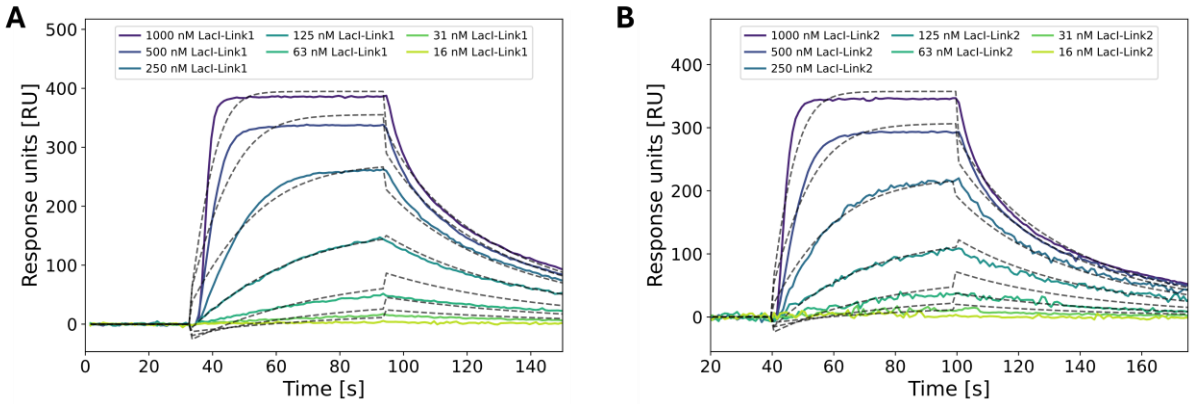


Figure 160. Concentration-dependent SPR analysis of LacI-Link1 and LacI-Link2 binding to symmetric operator DNA. Sensorgrams were recorded at a constant flow rate of 30 µL·min^{-1} at room temperature. Protein variants were injected over immobilised double-stranded oligonucleotides containing the cognate symmetric operator site using two-fold serial dilutions. Protein association was monitored for 60 s, followed by a dissociation phase of at least 60 s. Reference- and buffer-corrected sensorgrams were fitted using a 1:1 binding model to determine the kinetic parameters k_{on}, k_{off} and K_D. Measurements were conducted by Vanessa Troßmann. Dotted lines represent the fit. A) SPR sensorgrams of LacI-Link1 using serial dilutions from 1000 nM to 16 nM. B) SPR sensorgrams of LacI-Link2 recorded using serial dilutions from 1000 nM to 16 nM.

Table 47. Kinetic parameters of DNA binding for LacI linker variants were determined by SPR. Association rate constants (k_{on}), dissociation rate constants (k_{off}), and apparent dissociation constants (K_D) were derived from fits of the corresponding SPR sensorgrams.

Parameter	LacI-Link1	LacI-Link2	LacI
DNA	symmetric	symmetric	symmetric
K_D [nM]	145.0 ± 12.5	82.5 ± 2.3	9.5 ± 0.1
k_{on} [M^{-1} s^{-1}]	(2.28 ± 0.50)·10^5	(3.73 ± 0.89)·10^5	(2.60 ± 0.08)·10^6
k_{off} [s^{-1}]	0.033 ± 0.010	0.031 ± 0.008	0.026 ± 0.002

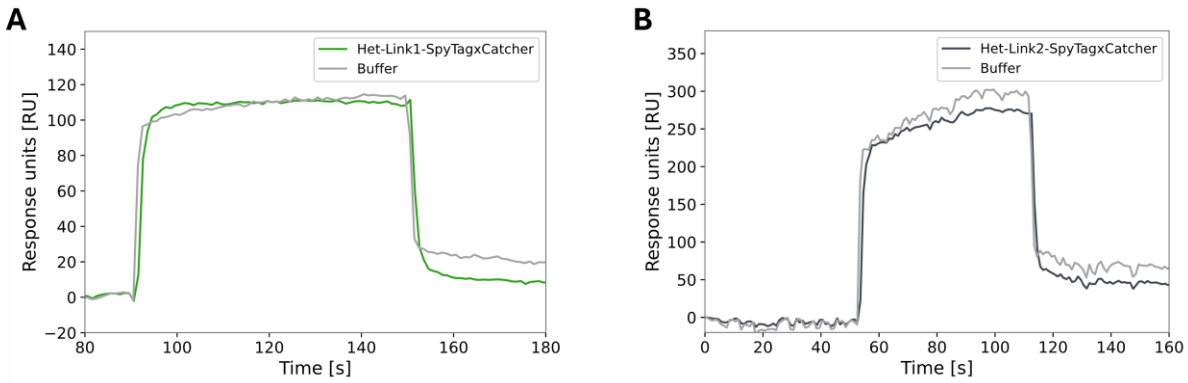


Figure 161. SPR analysis of Het-Link1-SpyTagxCatcher and Het-Link2-SpyTagxCatcher binding to asymmetric operator DNA. Sensorgrams were recorded at room temperature with a constant flow rate of 30 µL·min⁻¹. Het-Link1-SpyTagxCatcher and Het-Link2-SpyTagxCatcher were injected at a concentration of 3 µM over immobilised double-stranded oligonucleotides containing different operator arrangements. Measurements were conducted by Vanessa Troßmann. Protein-containing sample curves are shown in orange, while buffer/reference curves are shown in grey. A) Het-Link1-SpyTagxCatcher. B) Het-Link2-SpyTagxCatcher.

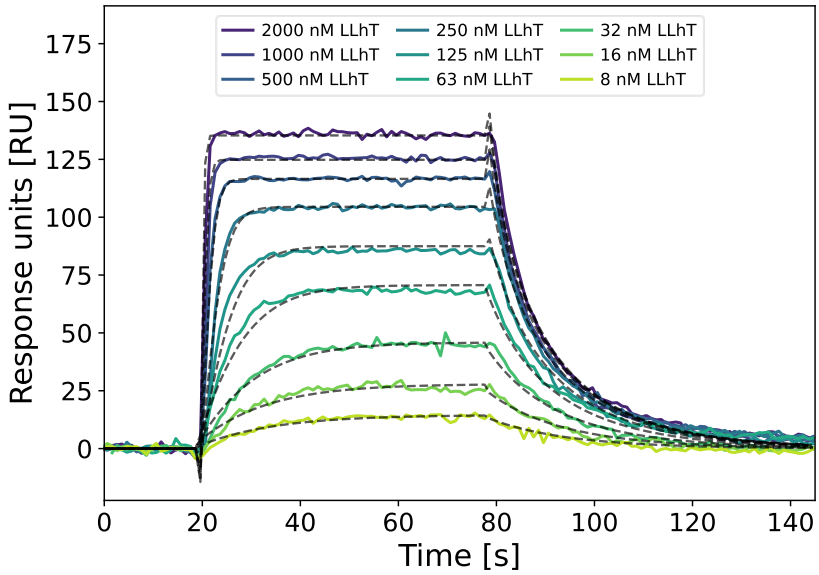


Figure 162. Concentration-dependent SPR analysis of LLhT binding to symmetric operator DNA. SPR sensorgrams of LLhT using serial dilutions from 2000 nM to 8 nM. Sensorgrams were recorded at a constant flow rate of 30 µL·min⁻¹ at room temperature. Protein variants were injected over immobilised double-stranded oligonucleotides containing the cognate symmetric operator site using two-fold serial dilutions. Protein association was monitored for 60 s, followed by a dissociation phase of at least 60 s. Measurements were conducted by Vanessa Troßmann. Reference- and buffer-corrected sensorgrams were fitted using a 1:1 binding model to determine the kinetic parameters k_{on}, k_{off} and K_D. Dotted lines represent the fitted curves.

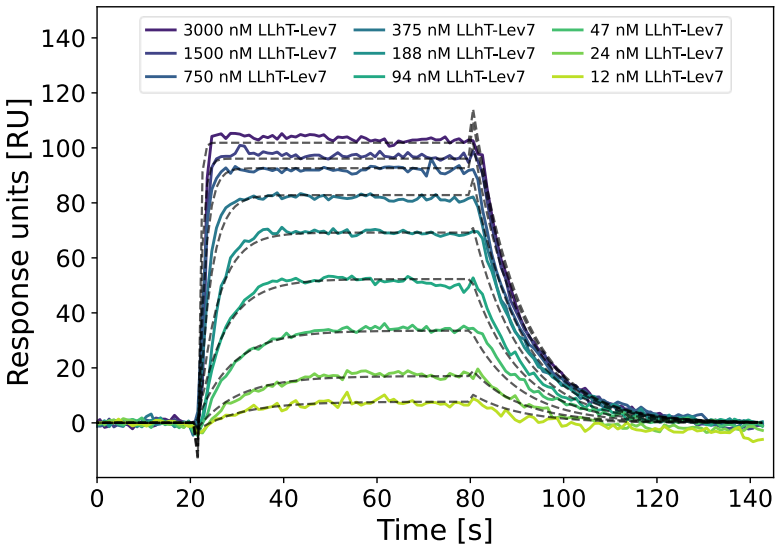


Figure 163. Concentration-dependent SPR analysis of LLhT-Lev7 binding to symmetric operator DNA. SPR sensorgrams of LLhT-Lev7 using serial dilutions from 2000 nM to 8 nM. Sensorgrams were recorded at a constant flow rate of 30 µL·min⁻¹ at room temperature. Protein variants were injected over immobilised double-stranded oligonucleotides containing the cognate symmetric operator site using two-fold serial dilutions. Protein association was monitored for 60 s, followed by a dissociation phase of at least 60 s. Measurements were conducted by Vanessa Troßmann. Reference- and buffer-corrected sensorgrams were fitted using a 1:1 binding model to determine the kinetic parameters k_{on}, k_{off} and K_D. Dotted lines represent the fitted curves.

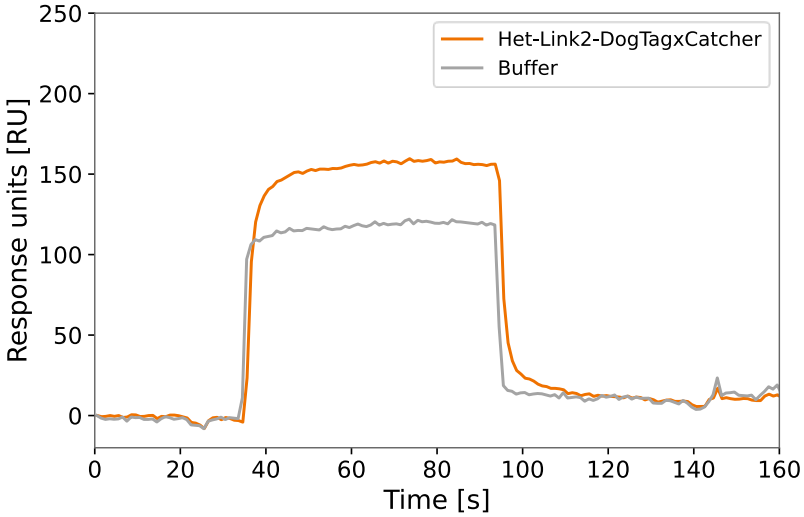


Figure 164. SPR analysis of Het-Link1-DogTagxCatcher binding to different operator DNA sequences. Sensorgrams were recorded at room temperature with a constant flow rate of 30 µL·min⁻¹. Het-Link1-DogTagxCatcher was injected at a concentration of 3 µM over immobilised double-stranded oligonucleotides containing different operator arrangements. Protein-containing sample curves are shown in orange, while buffer/reference curves are shown in grey. Measurements were conducted by Vanessa Troßmann.

7.9 Additional crystallographic data and evaluations

Table 48. X-ray crystallographic data collection and refinement statistics of PBP1. The table summarises the quality of diffraction data, refinement statistics, model composition, stereochemical validation, B-factor distribution, and TLS refinement parameters.

Parameter	PBP1
Wavelength	0.9184
Resolution range	33.889 – 1.250 (1.295 – 1.250)
Space group	I 2 2 2
Unit cell	90.642 102.08 149.05 90 90 90
Total reflections	2452978 (390774)
Unique reflections	185165 (29667)
Multiplicity	13.25 (13.17)
Completeness (%)	99.72 (98.93)
Mean I/sigma(I)	10.77 (0.45)
Wilson B-factor	22.68
R-meas	0.130 (5.198)
CC1/2	0.999 (0.374)
Reflections used in refinement	189134 (18649)
Reflections used for R-free	2096
R-work	0.1919 (0.6556)
R-free	0.2114 (0.6401)
Number of non-hydrogen atoms	5044
macromolecules	4420
ligands	15
solvent	609
Protein residues	596
RMS(bonds)	0.005
RMS(angles)	0.77
Ramachandran favored (%)	96.45

Parameter	PBP1
Ramachandran allowed (%)	3.21
Ramachandran outliers (%)	0.34
Rotamer outliers (%)	0.43
Clashscore	1.80
Average B-factor	1.80
macromolecules	30.1
ligands	28.77
solvent	27.54
Number of TLS groups	0

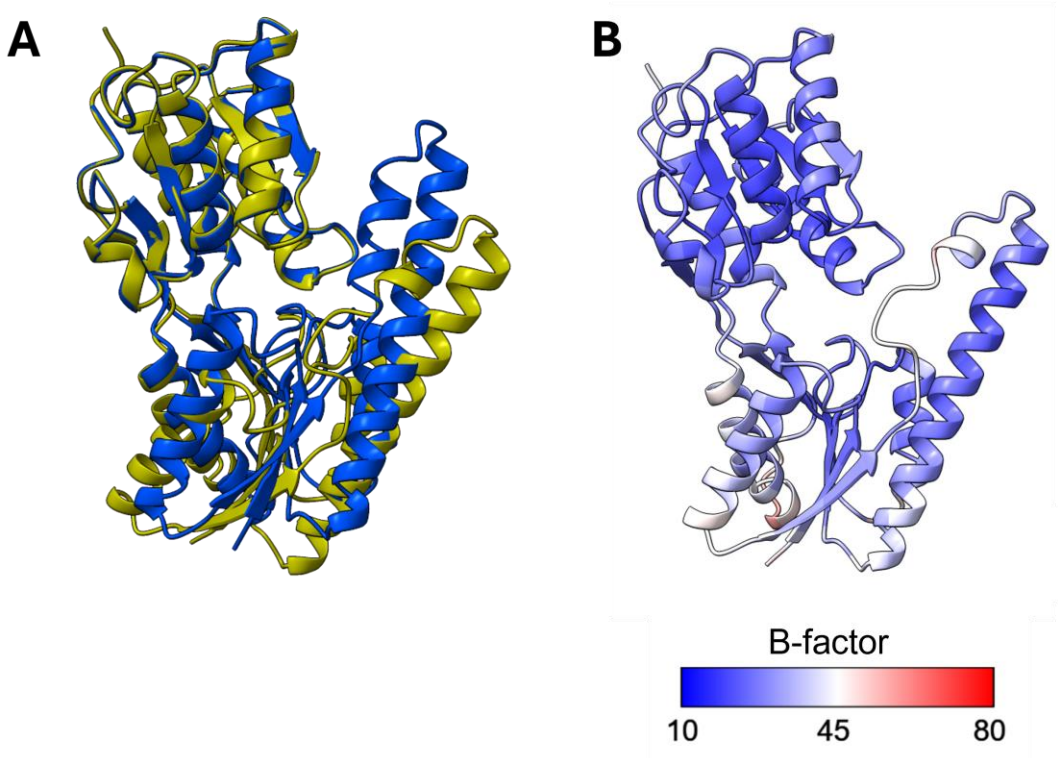


Figure 165. Analysis of the crystal structure of chain B A) Superposition of the AlphaFold2 model (blue) and the experimentally determined structure of PBP1 chain B (yellow). B) Crystallographic B-factor distribution mapped onto chain B, highlighting regions of increased local disorder or flexibility.

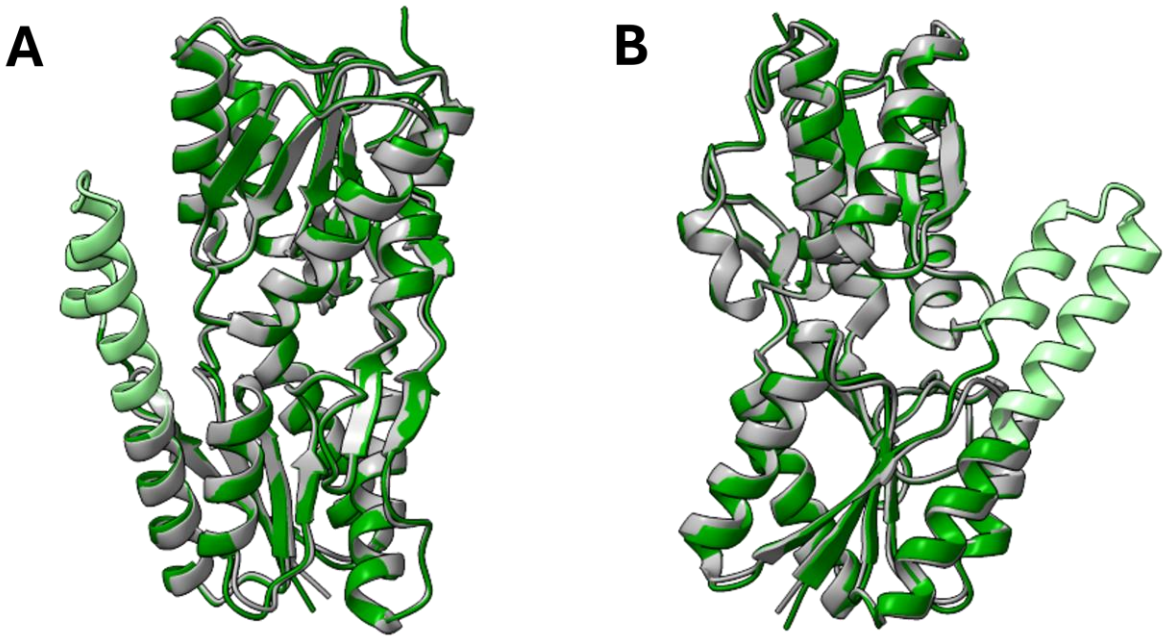


Figure 166. Structural comparison of PBP1 with wild-type LacI. A), B) Superposition of the experimentally determined crystal structure of PBP1 (green) with wild-type LacI (grey, PDB:2P9H[30]) in two orientations. The introduced structural extension in PBP1 is highlighted in light green. The comparison shows that the overall fold of the LacI-derived effector-binding domain is largely retained, while the extension protrudes from the core structure at the engineered insertion site.

7.10 Fluorescence measurements

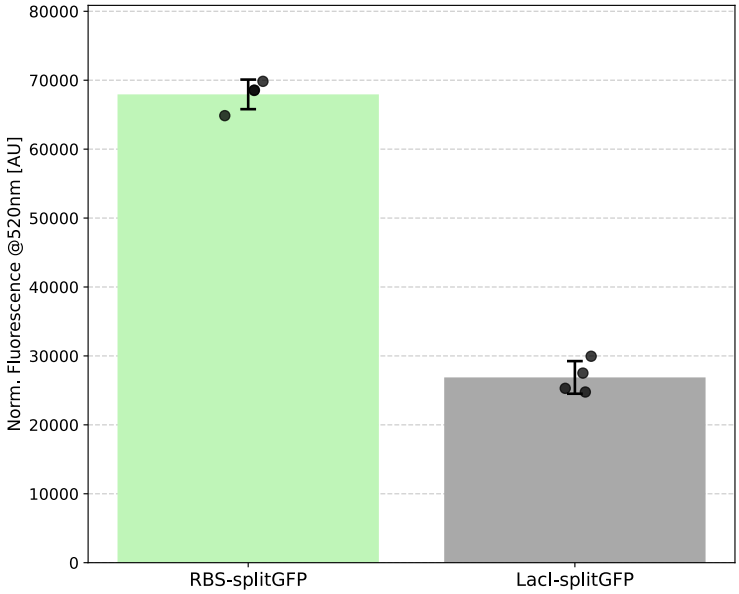


Figure 167. OD-normalised GFP fluorescence of splitGFP control constructs. Normalised GFP fluorescence at 520 nm was determined for RBS-splitGFP and LacI-splitGFP. The lower fluorescence signal observed for LacI-splitGFP indicates reduced splitGFP maturation or complementation compared with the RBS-splitGFP control. Bars represent mean values, individual measurements are shown as black dots, and error bars indicate the standard deviation.

7.11 Wildtype PotF

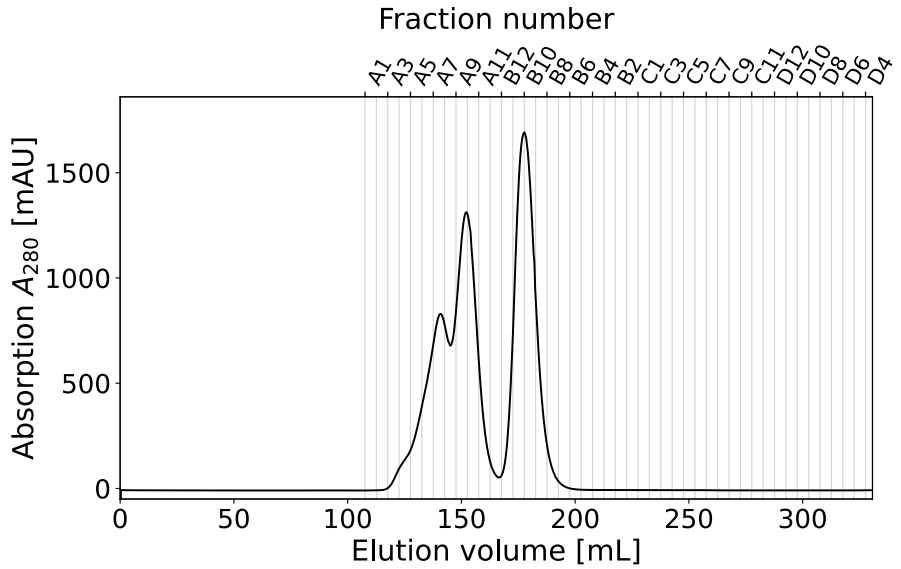


Figure 168. SEC chromatogram of PotF after Ni^{2+}-IMAC. Shown is the absorption at 280 nm against the elution volume. Purifications were conducted by Pascal Kröger and Sooruban Shanmugaratnam.

Acknowledgement

First and foremost, I would like to express my sincere gratitude to my supervisor, **Birte Höcker**, for giving me the opportunity to work on this fascinating project. I am deeply thankful for her continuous support, her guidance whenever questions arose, and for providing me with the opportunity to grow both scientifically and personally throughout this work. I greatly appreciate the time, trust, and encouragement she invested in me during the course of this thesis.

I would also like to thank the members of my thesis committee for the time and effort they devoted to evaluating my dissertation.

My sincere thanks go to the international **ArtMotor team**, including the members of **Heiner Linke's** and **Paul Curmi's** labs, both past and present. The interdisciplinary nature of this collaboration made discussions particularly inspiring and valuable. Exchanging ideas with researchers from very different scientific backgrounds was always enriching and helped me to view my work from new perspectives. I am especially grateful for the supportive atmosphere and for the openness toward my ideas throughout the project. I also appreciate the financial support received from the ERC.

I would also like to thank the **ArtMotor team** for the excellent team spirit and the many stimulating discussions while working together toward such an ambitious goal. I am especially grateful to **Vanessa**, particularly for her support in the later stages of this work. For conducting and helping with the SPR measurements, without which important parts of this thesis would have been missing. I also deeply appreciate her guidance during the writing of this dissertation.

Furthermore, I would like to thank all members of the **Höcker group** for their discussions, encouragement, and practical support. I am especially grateful for their flexibility whenever I needed to carry out experiments at short notice or required substantial resources. I also truly appreciated sharing the excitement of finally obtaining meaningful results after three years of work.

A very special thank you goes to **Sabrina**. I truly enjoyed our early-morning conversations, which often made even difficult days a little easier. Thank you for being there during the hard times, for always supporting me, and for always trying to understand the many different projects that were running in parallel. Knowing that I could count on your support made it much easier to come to the lab every morning, especially on the more challenging days.

Finally, I would like to thank the four most important people in my life.

Jonas, thank you for always motivating me during the hardest times. Whenever I was frustrated, doubted myself, or felt like giving up, you were always there for me. Thank you for listening so patiently to all my thoughts and frustrations over dinner, even when my research probably sounded like a different language. Your constant encouragement, patience, and belief in me meant more than I can express.

Mama and Papa, thank you for always being there for me, no matter what. You have supported me from the very beginning and helped me become the person I am today. I am deeply grateful for all the time, energy, and care you invested in my future. Beyond all practical support, your emotional support has been truly invaluable. Thank you for always being just a phone call or short car ride away whenever I was struggling, for encouraging me when things felt difficult, and for staying by my side throughout this journey.

And finally, my twin sister **Eva**, thank you for being part of my life from the very beginning or even before the beginning, in a way. Having you by my side is something truly special, and I could not have made it to this point without you. Growing up with you shaped me in countless ways and gave me a companion who always understood me in a way no one else could. Thank you for supporting me throughout this entire journey, and especially during my studies, when you were always there to encourage me and help me move forward. I am particularly grateful for the time you spent proofreading so many of my earlier works and helping me improve my English. But even more than that, thank you for your constant presence, your understanding, and your unwavering support. Knowing that I could always turn to you, whether for advice, reassurance, or simply comfort, has meant the world to me. Having you as my sister is one of the greatest gifts in my life.

(Eidesstattliche) Versicherungen und Erklärungen

(§ 8 Satz 2 Nr. 3 PromO Fakultät)

Hiermit versichere ich eidesstattlich, dass die Arbeit selbstständig verfasst und keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt habe (vgl. Art. 97 Abs. 1 Satz 8 BayHIG)

(§ 8 Satz 2 Nr. 3 PromO Fakultät)

Hiermit erkläre ich, dass ich die Dissertation nicht bereits zur Erlangung eines akademischen Grades eingereicht habe und dass ich nicht bereits diese oder eine gleichartige Doktorprüfung endgültig nicht bestanden habe.

(§ 8 Satz 2 Nr. 4 PromO Fakultät)

Hiermit erkläre ich, dass ich Hilfe von gewerblichen Promotionsberatern bzw. -vermittlern oder ähnlichen Dienstleistern weder bisher in Anspruch genommen habe noch künftig in Anspruch nehmen werde.

(§ 8 Satz 2 Nr. 7 PromO Fakultät)

Hiermit erkläre ich mein Einverständnis, dass die elektronische Fassung der Dissertation unter Wahrung meiner Urheberrechte und des Datenschutzes einer gesonderten Überprüfung unterzogen werden kann.

(§ 8 Satz 2 Nr. 7 PromO Fakultät)

Hiermit erkläre ich mein Einverständnis, dass bei Verdacht wissenschaftlichen Fehlverhaltens Ermittlungen durch universitätsinterne Organe der wissenschaftlichen Selbstkontrolle stattfinden können.

Bayreuth, den …… ……

Anna Zink