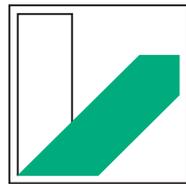


TIM Barrels 2.0

- Stabilizing and Diversifying a *De Novo* Designed Protein -



**UNIVERSITÄT
BAYREUTH**

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

Sina Elisabeth Kordes

aus Erlangen

Bayreuth, 2022

Die vorliegende Arbeit wurde in der Zeit von September 2017 bis März 2022 in Bayreuth am Lehrstuhl Biochemie unter Betreuung von Frau Prof. 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.).

Dissertation eingereicht am: 25.03.2022

Zulassung durch die Promotionskommission: 06.04.2022

Wissenschaftliches Kolloquium: 08.11.2022

Amtierender Dekan: Prof. Dr. Benedikt Westermann

Prüfungsausschuss:

Prof. Dr. Birte Höcker
Prof. Dr. Matthias Ullmann
Prof. Dr. Frank Hahn
Prof. Dr. Gerrit Begemann

Gutachterin
Gutachter
Vorsitz

*"After sleeping through a hundred million centuries
We have finally opened our eyes on a sumptuous planet
Sparkling with color, bountiful with life
Within decades we must close our eyes again
Isn't it a noble, an enlightened way of spending our brief
Time in the sun, to work at understanding the universe
And how we have come to wake up in it?"*

- The Greatest Show on Earth - Nightwish -

Contents

Abstract	i
Zusammenfassung	ii
1 Proteins	1
1.1 Protein Structure	1
1.2 Protein Stability	2
1.2.1 Stabilizing Forces in a Protein	2
1.2.2 Thermodynamics of Protein Stability	3
1.3 The TIM-Barrel Fold	4
1.3.1 Structure	5
1.3.2 Function	6
1.3.3 Stability Determinants	7
1.3.4 <i>De Novo</i> Design of a TIM Barrel	7
1.4 Coiled coils	8
1.4.1 Geometric and Sequence Parameterization	8
1.4.2 Biological Roles	9
1.4.3 Protein Design and Applications	10
2 The (Inverse) Protein Folding Problem	13
2.1 Mechanisms of Protein Folding	13
2.1.1 Folding Funnel and the Energy Landscape Model	14
2.2 Protein Structure Prediction	15
2.2.1 Template-based Modelling	15
2.2.2 Template-free Modelling	15
2.2.3 The Rosetta Algorithm	17
2.2.4 Recent Advances	17
2.3 Protein Design	18
2.3.1 <i>De Novo</i> Protein Design	18
2.3.2 Applications of Protein Design	19
2.3.3 Recent Advances and Perspectives	19
Synopsis	21
Author Contributions	22
3 Improving the Conformational Stability of a De Novo TIM Barrel	25
3.1 Project Objective	25
3.2 DeNovoTIMs - Improved Hydrophobic Packing	26
3.3 Introduction of a Salt Bridge Cluster	27
3.4 Conclusion	29
4 Extending the De Novo Designed TIM Barrel	31
4.1 Project Objective	31
4.2 Summary of Results and Conclusion	32
5 The Interplay of Protein Evolution, Folding and Design	35

6 Outlook	37
Bibliography	39
Manuscripts	51
Paper I: The stability landscape of <i>de novo</i> TIM barrels explored by a modular design approach	53
Paper II: A newly introduced salt bridge cluster improves structural and biophysical properties of <i>de novo</i> TIM barrels	105
Paper III: Extension of a <i>de novo</i> TIM barrel with rationally designed antiparallel coiled-coils .	133
Paper IV: Evolution, folding, and design of TIM barrels and related proteins	173
List of publications	187
Acknowledgements	189
(Eidesstattliche) Versicherungen und Erklärungen	190

Abstract

The TIM barrel is one of the oldest and most ubiquitous protein folds and is often referred to as the most successful in nature, due to its capability to host a diverse set of functions. This specific protein fold is characterized by a highly ordered topology with a central, eight-stranded, parallel β -barrel surrounded by eight α -helices. Interestingly, sequence analysis of natural TIM barrels revealed a low sequence conservation in this protein family despite a high structural similarity. Due to this high versatility in sequence and function the TIM barrel is a highly interesting target for protein design. The first validated *de novo* TIM barrel sTIM11 features a minimalistic structure of this protein fold. Due to its idealized topology it is an excellent system to study the folding determinants of the TIM-barrel fold and can also be used to create tailor-made enzymes. This work aims to improve this designed protein and create a set of diverse *de novo* TIM barrels. In the first part of this thesis, the original sTIM11 was stabilized by two different strategies. In a first approach, the hydrophobic packing of sTIM11 was improved by applying a fixed-backbone design strategy. In a modular approach initial stabilizing mutations in different regions of the barrel were identified and subsequently combined. This led to the construction of a large set of DeNovoTIMs with significantly improved stabilities and where crystal structures verified the formation of improved hydrophobic clusters. In a second approach, inspired by natural TIM barrels, a salt bridge network was installed in the β -barrel of three different *de novo* TIM barrels. Analysis of these salt bridge cluster variants revealed highly stabilizing, but also destabilizing effects. Structural analysis verified the formation of salt bridges but with various geometries ranging from single pair interactions to completely formed salt bridge networks. This highlights the challenges of designing salt bridges and especially salt bridge clusters. Even though all three analysed proteins have a highly similar fold, the influence on stability as well as the geometric formation of the salt bridges can vary significantly. In the second part of this work, the *de novo* TIM barrel is further diversified by the introduction of coiled coils into its $\beta\alpha$ -loops. Due to its minimalistic design principle, sTIM11 lacks any larger surface areas or cavities which can be utilized for installation of binding or catalytic sites. Therefore, the introduction of additional structural elements is a first step towards the creation of functional *de novo* TIM barrels. Using a multi-step design approach, a *de novo* designed antiparallel coiled coil was introduced into one and subsequently into a second $\beta\alpha$ -loop of the *de novo* TIM barrel creating a set of eight ccTIMs. Biochemical and biophysical analysis demonstrate the formation of additional α -helical elements with stabilizing interactions, which indicates a successful design. The research shown in this work created a large and diverse set of *de novo* TIM barrels which can be further utilized to build functional proteins and also to investigate folding determinants of TIM barrels.

Zusammenfassung

Das TIM-Barrel ist eine der ältesten und weit verbreitetsten Proteinfaltungen und wird häufig als eine der erfolgreichsten der Natur bezeichnet, da es eine Vielzahl an unterschiedlichen Funktionen durchführen kann. Diese Proteinfaltung ist durch eine hochgeordnete Topologie charakterisiert, die aus einem zentralen, achtsträngigen β -Barrel umgeben von acht α -Helices besteht. Trotz einer niedrigen Sequenzidentität besteht eine hohe strukturelle Übereinstimmung innerhalb dieser Proteinfamilien. Aufgrund dieser Vielseitigkeit in Struktur und Funktion ist das TIM-Barrel ein hochinteressantes Molekül für das Proteindesign. Das erste validierte *de novo* designte TIM-Barrel sTIM11 stellt eine minimalistische Variante dieser Proteinfaltung dar. Aufgrund seiner idealisierten Topologie ist es ein exzellentes System, um die bestimmenden Faktoren des TIM-Barrels zu untersuchen und kann verwendet werden um maßgeschneiderte Enzyme zu entwickeln. Ziel dieser Arbeit ist es, dieses designte Protein zu verbessern und die Diversität der *de novo* TIM-Barrels zu erhöhen. Im ersten Teil dieser Arbeit wurde das ursprüngliche sTIM11 mittels zwei verschiedener Strategien stabilisiert. Als erster Ansatz wurde die Packung des hydrophoben Kerns von sTIM11 mittels sogenanntem *fixed-backbone* Design verbessert. Dabei wurden zunächst stabilisierende Mutationen in einzelnen Bereichen des Barrels identifiziert und im Anschluss kombiniert. Dadurch wurde ein großes Set an DeNovoTIMs mit deutlich höheren Stabilitäten erzeugt und durch Kristallstrukturen die Ausbildung von verbesserten hydrophoben Clustern bestätigt. In einem zweiten Konzept wurde ein Salzbrückennetzwerk, welches durch natürlich vorkommende TIM-Barrels inspiriert war, in das β -Barrel dreier verschiedener *de novo* TIM-Barrels eingebaut. Die Analyse dieser Salzbrückencluster zeigte sowohl stark stabilisierende, als auch destabilisierende Effekte. Die Bildung der Salzbrücken konnte durch Strukturanalyse bestätigt werden, allerdings wurden sowohl paarweise Salzbrücken als auch vollständig ausgebildete Netzwerke beobachtet. Dies zeigt die Herausforderungen beim designen von Salzbrücken, und vor allem von komplexeren Salzbrückenclustern. Obwohl alle drei analysierten Proteine eine sehr ähnliche Faltung aufweisen, variiert der Einfluss der Salzbrücken auf die Stabilität aber auch deren Geometrie deutlich. Im zweiten Teil der Arbeit wurde das *de novo* TIM-Barrel durch den Einbau von Coiled-coils in die $\beta\alpha$ -Loops erweitert. Aufgrund der idealisierten Struktur von sTIM11 fehlen größere Oberflächenstrukturen oder Taschen, die für den Einbau von katalytischer Aktivität oder Bindung geeignet sind. Der Einbau zusätzlicher Strukturelemente ist daher der erste Schritt für die Etablierung eines funktionalisierten *de novo* TIM-Barrels. Mittels eines mehrstufigen Designs wurde ein *de novo* designtes, antiparalleles Coiled-coil zunächst in einen und dann in zwei der $\beta\alpha$ -Loops des *de novo* TIM-Barrels eingebaut und insgesamt acht verschiedene, sogenannte ccTIMs erzeugt. Das erfolgreiche Design wurde durch biochemische und biophysikalische Analysen bestätigt, welche die Ausbildung zusätzlicher α -helikaler Elemente mit stabilisierenden Wechselwirkungen zeigen. Durch die Forschungsergebnisse dieser Arbeit wurde ein breites Spektrum an *de novo* TIM-Barrels erzeugt, auf deren Basis funktionalisierte Proteine hergestellt und Faltungsdeterminanten der TIM-Barrels weiter untersucht werden können.

Chapter 1

Proteins

Protein science is a highly versatile research area, due to the complexity of these key elements of modern cellular life. Understanding these large biomolecules in detail provides crucial knowledge about cellular evolution, biological processes, and diseases, promoting innovations in biotechnology and biomedicine. The tasks of proteins are highly diverse, ranging from acting as structural components, responding to signals as molecular switches, and catalysing chemical reactions. Thereby, they control the growth of organism, form the basis of the metabolism and communicate with the environment. This high versatility and diversity is enabled by the building blocks of proteins, the amino acids, which are chemically highly diverse and thus ensure numerous different arrangements.

The central product of the genetic information encoded on the DNA of organisms are proteins. During biosynthesis DNA is transcribed into messenger RNA (mRNA) which is finally translated to amino acids. A triplet of three nucleotides encodes for one of the 20 proteinogenic amino acids, building the genetic code. Translation of the mRNA is catalyzed at the ribosome, where the translated amino acids are connected via peptide bonds to long polypeptides, creating ultimately the folded proteins.

1.1 Protein Structure

The great variety of the proteinogenic amino acids dictates through their diverse possible interactions and their miscellaneous chemical nature the structure and function of proteins. Therefore, the amino acid sequence of a protein, namely the *primary structure*, contains all information needed for a proteins' task. The native conformation of a protein is formed mainly through weak noncovalent interactions between the different amino acids and is organized in regular, local structures: the *secondary structure* elements. The most common ones were proposed already in 1951 by Linus Pauling on the basis of theoretical models: the α -helix and the β -sheets [1, 2] (Fig. 1.1). Less frequent elements are e.g. 3_{10} -helices, β -turns and omega loops. They are all stabilized by hydrogen bonds between backbone atoms with a specific pattern. In the prevalent α -helix hydrogen bonds are formed between the N-H group of residue n and the C=O group of residue $n+4$ pointing along the helix axis (Fig. 1.1A). The resulting right-handed helix has 3.6 residues per turn with side chains facing outwards. The β -sheets are stabilized through backbone hydrogen bonds between residues further apart in sequence. Two or more β -strands are arranged side by side in either parallel or antiparallel orientation to each other, interconnected through hydrogen bonds between N-H and C=O groups of the amino acid backbone (Fig. 1.1B). Due to steric effects all β -sheets have a right-handed twist, more pronounced for antiparallel ones. Secondary structure elements frequently form specific geometric arrangements with each other through side chain packing. These structural combinations are classified as supersecondary structure elements or motifs, and specific functions were associated to some of them. Typical ones are the $\beta\alpha\beta$ -motif, β -hairpin, β -barrels or coiled coils. The latter one will be described in

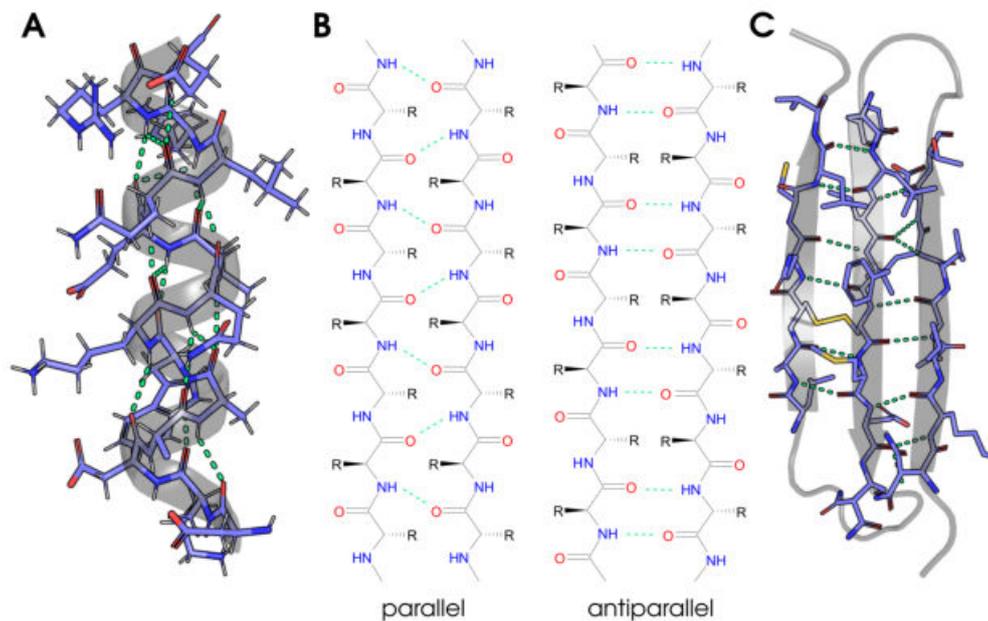


Figure 1.1: Secondary Structures. Most common secondary structure elements in proteins are shown. (A) Atomic representation of an α -helix is shown with stabilizing backbone hydrogen bonds highlighted in green. (B) Schematic representation of parallel and antiparallel β -strands with characteristic hydrogen bonds shown in green dotted lines. (C) Illustrative representation of an antiparallel β -sheet in atomic representation with hydrogen bonds shown in green.

detail in section 1.4.

In the native conformation of a protein the (super)secondary structure elements pack together and build a complex, compact fold. This three-dimensional arrangement is designated as *tertiary structure*. It is stabilized by weak, non-local interactions which typically results in the burial of hydrophobic side chains into a tightly packed protein core. Larger proteins often fold into multiple globular units: the domains. These fold autonomously and are self-stabilizing by the formation of individual hydrophobic cores. They are used in many protein classification systems and are believed to be evolutionary units. One of the most abundant and versatile protein domains in nature is the TIM-barrel fold, which is introduced in detail in section 1.3. Many proteins assemble into larger complexes, thereby acquiring their complete functional possibilities. This spatial structural organisation is referred to as the *quaternary structure* of a protein. Different compositions frequently occur: either two or more molecules of the same protein form dimers or homooligomers, or different proteins organize to heterooligomers or large protein complexes.

1.2 Protein Stability

In their functional form proteins fold typically into a single well-defined structure, the native state. To achieve and remain in this conformation different interactions have to be formed to favor it over unfolded or misfolded states. The stability of a protein is highly important for its optimal functioning and determines its resistance to high temperature, denaturant, proteases or pH changes.

1.2.1 Stabilizing Forces in a Protein

The native state is stabilized by a large number of interactions with different contributions to the overall stability. Following interactions and forces are described which are the main drivers of protein stability and define the secondary, tertiary and quaternary structure.

Hydrogen bonds are formed between partially charged donor and acceptor atoms or groups due to electrostatic attraction. In case a hydrogen atom is covalently bound to a highly electronegative donor, it obtains a partial positive charge due to the difference in electronegativity. Therefore, it interacts with the negative charge of unpaired electrons of an acceptor group and a hydrogen bond is formed. Typical donor and acceptor atoms are nitrogen (N), oxygen (O), and fluorine (F). Hydrogen bonds involve both backbone and side chain atoms and are therefore one of the most frequent stabilizing forces in proteins.

So called **Coulomb interactions** are present between charged groups and can either be attractive or repulsive. This force F is dependent on the distance r of the groups and their magnitude of charge, q_1 and q_2 , and is described in Coulombs law with k_e being Coulomb's constant:

$$E_{el} = \frac{1}{4\pi\epsilon_0} \frac{|q_1q_2|}{r^2} = \frac{1}{k_e} \frac{|q_1q_2|}{r^2} \quad (1.1)$$

In case both groups have the same magnitude the force between them is repulsive, in case they are of opposite charge they attract each other. These electrostatic interactions occur in proteins between charged amino acids. Special ionic or salt bridges can be formed when the distance between two charged residues is less than 4 Å.

Van der Waals forces act between all kind of atoms and are a combination of repulsive and attractive forces. The Pauli repulsion applies when two electron orbitals overlap and prevent the collapse of molecules. Transient dipoles due to fluctuation of electron clouds can induce complementary dipoles in proximal atoms, providing weak electrostatic attraction. As for each atom type the size of its electron cloud differs, the optimal contact distance for this type of interaction varies and is specified as the Van der Waals radius. At a certain distance the repulsion is minimal while the attractive force is maximal, which is termed the Van der Waals contact distance. The Lennard-Jones potential is a simplified model to describe this force mathematically, with r being the distance between an interacting pair, ϵ the dispersion energy, and σ the distance at which V equals zero.

$$V_{LJ}(r) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] \quad (1.2)$$

The **Hydrophobic effect** is the main driving force of protein folding and has a large contribution to the stability of a conformation. In the folded protein hydrophobic side chains are sequestered in the core, thereby avoiding unfavorable interactions with polar molecules of the surrounding solvent. As a consequence the entropy of the water molecules increases, as they don't form ordered cluster around the hydrophobic side chains.

Another major stabilizing interaction is the covalent **disulfide bond** formed between two sulfur groups (SH). In proteins they are solely formed between two cysteine residues as a result of oxidation in case they are brought in close proximity during folding.

1.2.2 Thermodynamics of Protein Stability

From a thermodynamic perspective protein stability is highly interesting due to its delicate balancing of stabilizing and destabilizing effects [3, 4]. The stability of a protein is defined as the net loss of free energy upon unfolding with $\Delta G_{folding}$ being described as the difference between the free energy of the native state G_{native} and unfolded states $G_{unfolded}$:

$$\Delta G_{folding} = G_{native} - G_{unfolded} \quad (1.3)$$

The free energy of a system is a combined effect of the systems entropy S and enthalpy H , leading to the following relation for the folding of a protein [5]:

$$\Delta G_{\text{folding}} = -RT \ln K_{\text{eq}} = \Delta H - T\Delta S \quad (1.4)$$

with K_{eq} the equilibrium constant, ΔH the enthalpy change, ΔS the entropy change, R the universal gas constant and T the absolute temperature. The enthalpy describes the internal atomic forces in a protein (Van der Waals interactions, electrostatic interactions, hydrogen bonds, etc) [6].

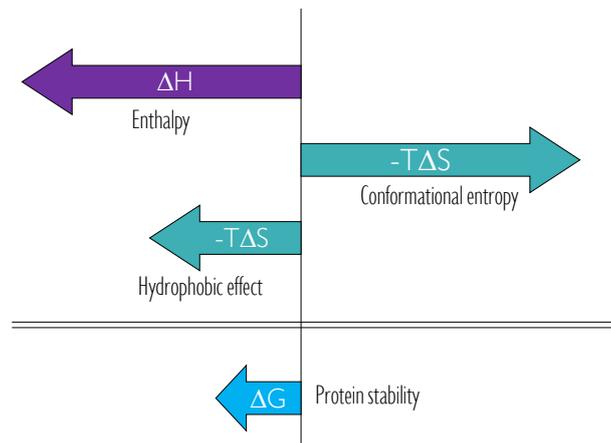


Figure 1.2: Thermodynamics of protein folding. The stability of a protein is described as the free energy change upon folding ΔG . This is influenced by the systems entropy S and enthalpy H . The enthalpy change has a large contribution on the protein stability by numerous stabilizing interactions. Upon folding the entropy of the protein itself decreases drastically, almost outperforming the enthalpy change. Nevertheless, the entropy of the surrounding solvent increases upon folding and has therefore a significant contribution to spontaneous folding. Here these thermodynamic contributions of enthalpy and entropy on the protein stability are shown, to exemplify the delicate balancing required for a stable, folded protein.

All these forces have only a small individual contribution of about $1 - 3 \text{ kcal mol}^{-1}$, but in a native fold a large number of interactions add up to contribute significantly to a proteins' stability [3] (Fig. 1.2). The contribution of entropy to the stability is more complex [7]. As proteins fold spontaneous into their native state the entropy of this system must increase. As the entropy describes the degree of disorder of a system, it seems to be contradictory in the context of protein folding. The conformational entropy of the peptide chain decreases significantly, having a similar or larger magnitude as the enthalpy change. Therefore, another entropic contribution has to exist to make protein folding a spontaneous process. This is explained by the hydrophobic effect. Upon folding the hydrophobic side chains pack into the proteins' interior, releasing water molecules, which cluster around them in the unfolded state. Therefore, the water molecules undergo an entropy increase, making an important contribution to the entropy of the system. These three terms - the internal interaction, the conformational entropy and the hydrophobic effect - drive the spontaneous folding of proteins and stabilize the native state (Fig. 1.2). Due to this delicate balancing of thermodynamics, the net free energy of a native protein is rather small with typical values of $5 - 10 \text{ kcal mol}^{-1}$ [3].

1.3 The TIM-Barrel Fold

One of the most investigated protein folds is the $(\beta\alpha)_8$ - or TIM-barrel fold. This canonical fold is characterized by a central eight-stranded, parallel β -barrel surrounded by eight α -helices, and was first described for the eponymous triose phosphate isomerase (TIM) (Fig. 1.3) [8]. Meanwhile, this fold was observed in more than 10 % of all structures deposited in the Protein Data Bank (PDB) and was classified to participate in enzymatic reactions assigned to five of the seven known enzyme classes [9]. Interestingly, a low sequence conservation was determined throughout this superfamily, despite a high structural similarity

[10]. Therefore, it is often referred to as the most common enzyme fold and is often designated as the most successful in nature. Due to its ubiquity, the TIM-barrel fold is a model system in different research areas of protein science: its evolutionary development was investigated in several studies, its folding mechanisms was examined, and stability determinants of this fold were characterized.

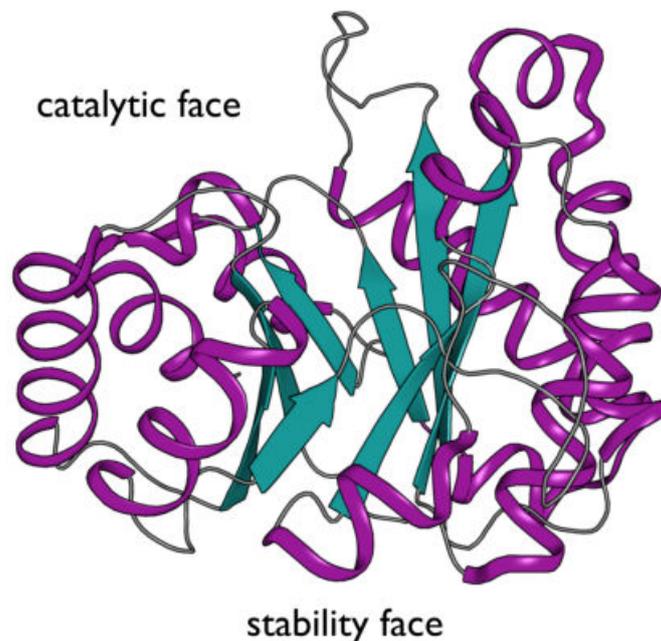


Figure 1.3: The TIM-barrel fold. The crystal structure of the eponymous $(\beta\alpha)_8$ /barrel triose phosphate isomerase (TIM) is shown to exemplify the ubiquitous TIM-barrel fold (PDB ID: 1TIM). An eight-stranded central β -barrel (in cyan) is surrounded by eight α -helices (in purple). At the N-terminal side of the β -barrel the stability face of the TIM barrel is found, comprising a large set of stabilizing interactions. On the opposite site the catalytic face is located, which is more diverse in different proteins and can comprise longer loops, secondary structure elements or larger insertions.

1.3.1 Structure

A typical TIM barrel is about 200 to 250 residues long and can function as a single domain, but also regularly forms homo-oligomers or is found in multi-enzyme complexes [11]. The topology can be separated in $\beta\alpha$ -units which are repeated eight times to form a closed barrel structure. The units are internally connected by $\beta\alpha$ -loops, whereas different units are connect by $\alpha\beta$ -loops. Parameterization of barrel structures led to the definition of the shear number s as important descriptor, which specifies the shift of residues when moving around the barrel by one turn [12]. For all TIM barrels a shear number of eight residues was determined [13]. Detailed analysis of the barrel core revealed a four-fold symmetry: side chains of the β -sheets point towards the barrel center in layers each constructed by four residues, either of the odd- or even-numbered β -strands [9].

A key feature of the TIM barrel, and maybe one of its secrets of success, is the spatial separation of stability and function. Analysis of sequence conservation and interactions showed a predominant stabilization of the fold through the 'stability face' comprising the hydrophobic barrel core, N-terminal ends of the β -strands and the connecting $\alpha\beta$ -loops [14, 15] (Fig. 1.3). On the opposite side of the barrel, invariably for all known $(\beta\alpha)_8$ -barrels, the C-terminal ends and the $\beta\alpha$ -loops contain the catalytic residues and generate the 'catalytic face' [16]. Analysis of the loops revealed a higher variability of $\beta\alpha$ -loops, as they vary in their length and frequently have larger insertions forming secondary structure elements or even domains [17, 18, 19]. This modularity makes it possible to change function without affecting stability and is a possible reason for the ubiquity of this fold making it a very interesting target for protein engineering and design. Different studies have shown that loops of natural TIM barrels can be redesigned without

compromising stability [20, 21, 22]. Despite the collective features of the $(\beta\alpha)_8$ -barrel this superfamily shows high structural diversity. Natural TIM barrels show frequently minor deviations from the topology like a reduced number of β -strands, inversion of secondary structure elements, terminal extensions or circular permutation.

1.3.2 Function

One of the most intriguing features of the TIM-barrel proteins is the high versatility of functions. In the structural classification database SCOP TIM barrels are classified into 33 unique superfamilies, each with a diverse functional range thereby covering enzyme activities from five of the seven enzyme commission (EC) classes [23]. About half of the known TIM barrels act as hydrolases with a large proportion of glycosidases, but also oxidoreductases, isomerases, lyases and transferases are frequently found [10]. Only a few TIM barrels without enzymatic activity are known, like the storage protein narbonin which provides source materials in developing seeds [24]. About 85 % of the catalyzed reactions are involved in energy metabolism, macromolecule metabolism, or small molecule metabolism [9]. The remaining functions are mostly unclassified or play important roles in transport or information pathways [10]. The broad range of functions is enabled through the use of several different cofactors, especially metals but also derivatives of nucleotides or amino acids [23]. A preference for negatively charged metabolites was determined and could be explained by the presence of a positive potential at the catalytic face created by α -helix microdipole induced electrostatic field patterns of the barrel [25, 11]. Additionally, side chains seem to focus this electrostatic field near the active site [9, 26].

Some of the most outstanding enzymes are TIM barrels. Starting with the eponymous TIM, a glycosidase, catalyzing the reversible isomerization of glyceraldehyde-3-phosphate (GAP) to dihydroxyacetone phosphate (DHAP), which was found in all domains of life [27]. Analysis of the enzymatic activity revealed an extremely efficient turnover from GAP to DHAP only limited by diffusion, making TIM a kinetically perfect enzyme [28]. In contrast, Rubisco is a very inefficient enzyme, which catalyzes the fixation of carbon dioxide in photosynthesis [29]. This enzyme is assembled from eight large and eight small subunits, where the large subunits are dimers of a β -sheet domain and a $(\beta\alpha)_8$ -barrel comprising the active site. The inefficiency is compensated by an exceptionally high concentration of about 250 mg mL^{-1} in the chloroplasts, making it the most abundant enzyme on earth [30]. A TIM barrel with a high clinical significance is Heparanase [31]. It breaks down heparin sulfate, a key component of the extracellular matrix, and was shown to be overexpressed in tumor cells and was identified as cancer marker with high linkage to cancer metastases. The increased breakdown of heparan sulfate results in release of stored growth factors. Therefore, heparanase is an interesting target for the development of anti-cancer drugs. Additionally, it was also linked to viral infections like COVID-19 and could be a possible drug target in viral infections [32].

An intensely studied enzyme is the imidazole glycerol phosphate synthase (IGPS), which forms a heterodimer in bacteria. HisH constitutes the glutaminase subunit with a flavodoxin-like fold, whereas HisF functions as synthase and has a $(\beta\alpha)_8$ -fold [33]. In a first step HisH catalyzes the hydrolysis of glutamine to glutamate and ammonia. Its active site is located at the dimer interface in proximity to the barrel center of HisF. Ammonia is then channeled through the β -barrel interior to the active site of HisF, where it reacts with N^3 -[(5'-phosphoribulose)]formimino]-5-aminoimidazole-4-carboxamide-ribonucleotide (PFRAR) to imidazole glycerol phosphate (IGP) and 5-aminoimidazole-4-carboxamide ribotide, both entering histidine or purine biosynthesis pathways. Both reactions are tightly coupled and an allosteric mechanism has been determined, which activates HisH only upon binding of PFRAR to HisF [34]. HisF developed to be a model system to understand the evolutionary history and relationships of TIM barrels and its stability determinants and folding pathway was investigated in detail.

1.3.3 Stability Determinants

The identification of explicit stabilizing elements in the $(\beta\alpha)_8$ -fold was the goal of several computational and mutagenesis studies. A large number of stabilizing residues was found in the β -sheet, which create an optimally packed barrel center and constitute the interface to the α -helices. Interestingly, the α -helices were not identified to be crucial for stability [35]. Polar residues in the protein interior forming buried salt bridges were also described to be essential for the formation of a stable barrel [36, 15]. Using a knowledge-based potential Wiederstein and Sippl [37] analysed amongst others the TIM-barrel fold and thereby supported previous studies which classified the $\beta\alpha$ -loops as catalytically important in contrast to the $\alpha\beta$ -loops, that contribute essentially to the stability only.

1.3.4 De Novo Design of a TIM Barrel

The ultimate test of our knowledge of proteins is the *de novo* design of a protein. A general view and recent advances of protein prediction and design are presented in Chapter 2. The TIM-barrel fold is a highly interesting target for protein design, due to its regular backbone geometry, its versatility in function, its sequence diversity and its high stability in general. Therefore, it was one of the first targets of protein design, but it took about 25 years until the first successful design was achieved. The first design attempts were done in the early 1990s: sequence-based designs were performed based on statistical analysis of a small set of natural TIM-barrel structures. The resulting proteins exhibited large amounts of secondary structure, but loose internal packing indicating the formation of molten-globule like states [38, 39, 40, 41, 42, 43]. With newly emerging computational methods, structure-based designs became possible, including all-atom models. Combined with an increasing set of natural TIM-barrel structures further design strategies were developed. Initially C_α -traces were created based on parameters established from natural structures, followed by an automated sequence design. The resulting proteins featured high secondary structure content and evidence of tertiary structure, but had an overall low solubility and conformational stability [44, 45]. Later on, one of those designs was optimized by directed evolution and a crystal structure was solved: it deviated from the intended $(\beta\alpha)_8$ -topology and revealed a Rossmann-like fold [46]. Detailed analysis of previous designs and a large set of natural TIM-barrels implied possible importance of polar residues in the core for protein folding [47]. Based on geometric and topological descriptors a protein backbone was created by assembling small fragments from known proteins and was applied to fixed backbone sequence design with specific restrictions including a high polarity for core residues. The resulting protein exhibited excellent solubility and folding characteristics, but again showed molten-globule like behaviour.

In the same year, 25 years after the first attempt, the mission of creating a *de novo* TIM barrel finally succeeded. Huang et al. [48] started with the definition of design principles. One major point in their design is the determination of the highest possible symmetry of a TIM barrel, which is a four-fold one. To achieve a shear number of 8, different β -strand compositions were tested. Based on the α/β -rules for protein loops from Koga et al. [49], which state that the first residue of the β -strands should point towards the barrel center, only one composition was feasible. Therefore, the required topology for the design was specified to be a repeat protein consisting of four identical $\beta\alpha\beta\alpha$ -subunits with a register shift of two residues. Following, the surrounding α -helices in one subunit had to be of different length and tilt. Based on these characteristics, an ideal backbone was created and applied to iterative sequence design. Additional restrictions were incorporated to ensure helical capping, good packing of the α/β -interface and satisfaction of hydrogen bonds in loop regions. From 22 tested designs five showed promising features and for one circular permuted version, sTIM11, a crystal structure was determined at a resolution of 2.0 Å (Fig. 1.4). Comparison with the initial model showed a high accuracy even at the side chain level, illustrating the successful design of a TIM barrel from scratch. Contrasting this successful design with previous ones, shows the importance to design specific side chain-backbone hydrogen bonding. Hydrophobic packing is obviously a key feature of proteins but the design of polar interactions seems to be essential for the

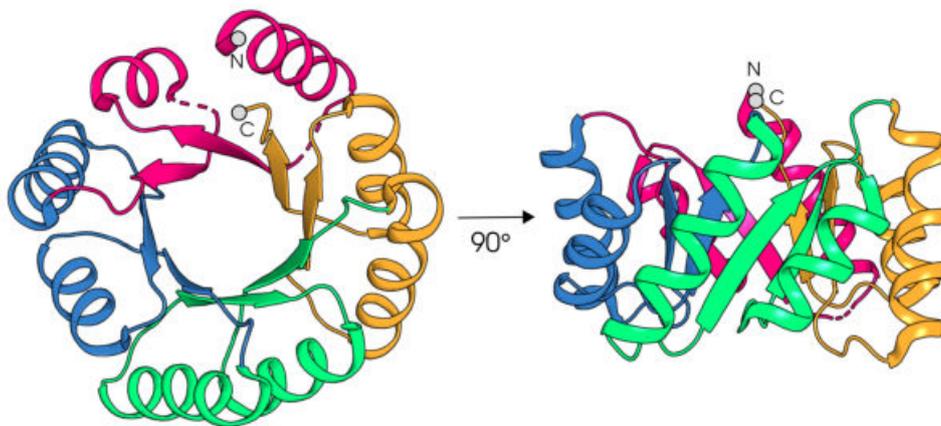


Figure 1.4: The de novo designed TIM barrel. The crystal structure of sTIM11 (PDB ID: 5BVL) is shown in top and side view. The four-fold symmetric quarters, consisting each of a $\beta_2\alpha_2$ -fragment, are shown in different colors. N and C termini are marked with grey circles.

formation of a specific, compact tertiary structure.

1.4 Coiled coils

Another highly structured, but functional diverse protein unit is the coiled-coil motif. In the early 1950s Linus Pauling and Francis Crick described it independently from each other solely based on theoretical models without knowledge of any amino acid sequence [50]. Nowadays it is one of the best studied protein folding motifs and the proposed parameterizations were tested by protein design leading to an excellent understanding of the sequence-structure relationship of these highly parameterizable super-secondary structures. Analysis of their occurrence in nature revealed their presence in all domains of life, being found in approximately 5 % of all proteins, making coiled coils ubiquitous with a broad functional range [51].

1.4.1 Geometric and Sequence Parameterization

Already in their initial proposals of a coiled coil Pauling and Crick defined important features and even provided a detailed packing description and parameterization of supercoiled α -helices [52, 53, 54]. Those were confirmed 30 years later by the solved structure of influenza virus hemagglutinin [55] and another 10 years later by the structure of the dimerization site of the DNA-binding protein GCN4 [56].

A coiled coil is formed by α -helices twisting around each other forming a left-handed supercoil (Fig. 1.5). Typically, they comprise two to six helices, which are arranged in either parallel or antiparallel orientation to each other. But also assemblies of more than 20 helices are present in nature [57, 58]. They assemble into homo- or heterooligomers and can be formed by helices either from separate chains or from a continuous polypeptide [59]. The various observed geometries and topologies are summarized in the periodic table of coiled-coil protein structures, which contains also several more complex architectures, which can not be described completely by the generic parameterization [59]. The following description of coiled coils will focus on two-stranded coiled coils, but higher order assemblies have comparable properties.

Characteristic for coiled coils is their specific mode of interaction via the so called "knobs-into-holes" packing, already described by Crick [54]. The knobs are side chains of one helix, which pack into so called

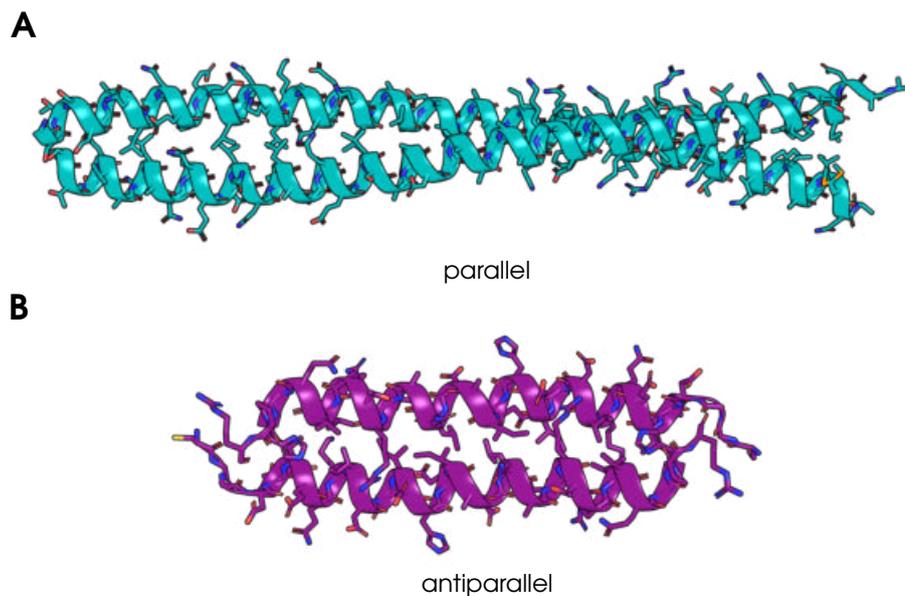


Figure 1.5: Coiled coils. Example structures for (A) parallel (PDB ID: 1UIX) and (B) antiparallel (PDB ID: 1R48) two-stranded coiled coils. The two helices form left-handed supercoils by burying hydrophobic residues at the interface.

holes, cavities formed by four surrounding residues of the facing helix, and vice versa. This packing can only be achieved through a periodic positioning of side chains at equivalent positions along the interface. This is accomplished through the supercoiling of the helices around each other: the typical periodicity of the α -helices is reduced from 3.6 to 3.5. Therefore, every seventh residue occupies an equivalent position along the helix axis. This structural periodicity is also reflected in the amino acid sequence, which exhibits a seven-residue long sequence pattern, the characteristic heptad repeat [60]. A typical notation of this pattern is $a-g$ and can be represented in a helical wheel diagram, displaying the position of each residue in context of the coiled-coil assembly (Fig. 1.6). The coiled-coil interface is constructed by a and d residues, which are frequently occupied by hydrophobic residues creating a hydrophobic core which is critical for assembly and stability. At the positions e and g charged amino acids frequently form interhelical salt bridges, shielding the hydrophobic core from solvent, driving oligomerization and modulating directionality. Remaining positions of the heptad repeat are solvent-exposed and are occupied by hydrophilic, helix promoting residues [61]. Focusing only on two-stranded coiled coils here, two different assemblies are possible: parallel and antiparallel. Depending on the orientation the interactions at the interface differ slightly. In a parallel assembly the a and d residues of each helix interact with each other and ionic interactions are formed between e and g residues (Fig. 1.6A). In contrast, the antiparallel orientation promotes interaction of the a residue of one helix with the d residue of the opposing helix and ionic interactions are formed between e or g residue pairs (Fig. 1.6B). The formation of a hydrophobic core through the interaction of a and d residues are the main driving force in coiled-coil assembly, which is supplemented by attractive forces between e and g residues [62]. Precise selection of the different interactions, both at the hydrophobic interface as well as of the surrounding residues, modulates orientation as well as the assembly state of the coiled coils.

1.4.2 Biological Roles

Even though coiled coils are simple in structure and have a limited sequence space due to the sequence periodicity, a wide range of different biological roles were assigned to them. Discontinuous coiled coils frequently facilitate oligomerization of proteins and are thereby a highly common motif in complex protein assemblies [63]. Due to their rigidity and length they are found in many fibrous proteins and mediate

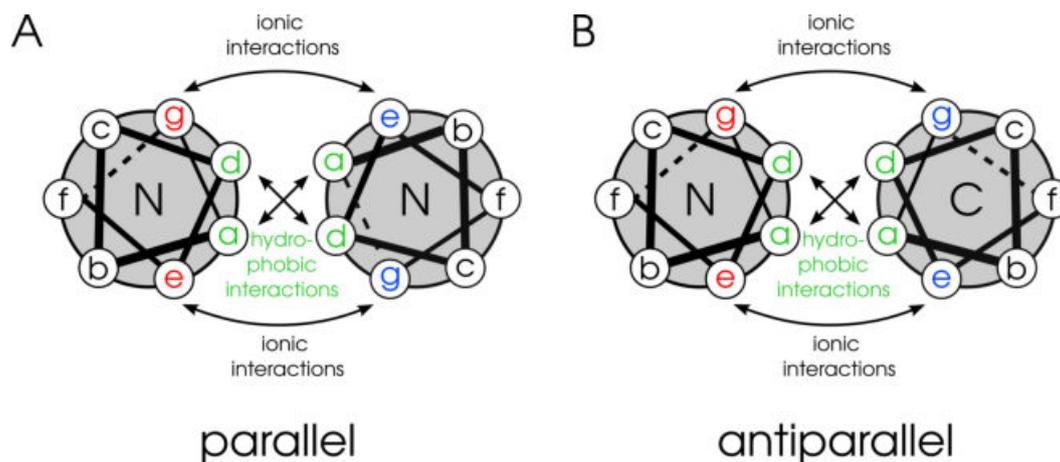


Figure 1.6: Helical wheel diagrams visualize the characteristic packing of coiled coils. The typical heptad repeat a-g is shown schematically in its three dimensional orientation in a coiled coil. The helical wheel diagrams show the interaction pattern for parallel (A) and antiparallel (B) two-stranded coiled coils. At the interface of the helices a and d residues form hydrophobic interactions and are flanked by ionic interactions formed between e and g residues. The exact interaction pattern for both coiled-coil types differ and the interhelical interactions are indicated with arrows.

mechanical strength in biological materials such as hair, nails and muscles [57]. Additionally, this property makes them excellent molecular spacer separating functional domains from each other or spanning the inner and outer membrane of bacteria [64]. As molecular spacer they are key players in cell division, tether vesicles and control DNA recognition. Also more complex functions are performed by coiled coils through transmission of structural changes along them and they are especially known for their importance in motor proteins, in which they are involved in complex allosteric communication [65]. A very special task is the molecular ruler: through the defined length it facilitates the production of e.g. antigens with a specific length [66]. Finally, for one coiled coil also a catalytic activity was verified [67].

1.4.3 Protein Design and Applications

One major model system in protein design is the coiled coil. Due to the simple interrelation between sequence and structure they were early protein design targets. Nevertheless, the precise modulation of coiled-coil assemblies needed many years of research to specify detailed rules to fine-tune oligomerization states and helix directions and revealed requirement for a fine balancing of interactions. Here, an overview of important design attempts is given for parallel and antiparallel two-stranded coiled coils, followed by examples how the acquired information were applied in the creation of larger assemblies and how they are utilized.

Parallel Coiled Coils

In the seminal works of O’Shea et al. [56] and Harbury et al. [68] the so called “*Harbury relationship*” was determined, which connects specific a and d residue combinations with different oligomerization states on the basis of a natural leucine-zipper coiled coil. These straightforward rules created the basis for several subsequent studies. In a later work Fletcher et al. [69] verified these rules in a context independent background with *de novo* designed proteins and created a basic toolkit of parallel coiled coils with targeted oligomerization states. The influence of intra- and interhelical salt bridges on parallel assemblies was dissected by Burkhard et al. [70] and revealed the possibility to fine-tune their stability. Gradišar and Jerala [71] showed through a negative design approach that specific positioning of Asn residues at the a or d position as well as rational arrangement of electrostatic interactions at e and g position can direct the orientation of the coiled coil and guide homo- or hetero oligomerization precisely. A critical element in protein design is to obtain stable proteins. Specific rules were identified for parallel coiled coils which can

modulate the stability of the assemblies [72, 73]. Further rules modulating the assembly were determined based on natural as well as *de novo* proteins, directing and improving prediction and design [50]. Nowadays a large set of successfully designed parallel two-stranded coiled coils is available with different topologies, sequences and stabilities.

Antiparallel Coiled Coils

Most studies on two-stranded coiled coil design focused on parallel arrangements, despite the majority of natural ones being antiparallel [74]. Only limited research has been done on understanding the determinants of antiparallel coiled-coil assemblies. First antiparallel designs were created based on previously designed parallel coiled coils constrained by disulfide bridges [75, 76]. A more precise and rational approach created a continuous antiparallel coiled coil [77]. This 56-residue polypeptide was based on specific hydrophobic interactions between *a* and *d* residues and the orientation was directed by ionic interactions between *e* and *g* residues. The design of an heterodimeric coiled coil without any covalent constraints directed the antiparallel orientation through specific interaction of two asparagines at the helix interface [78]. This favors the antiparallel orientation over the parallel, but in solution a mixture of both conformations was observed. In a similar approach, with addition of interhelical coulombic interactions, an exclusively antiparallel alignment was achieved [79]. Also the design of an homodimeric antiparallel coiled coil was enforced by a negative design approach. Rational placement of arginine residues in the hydrophobic interface suppresses the formation of parallel arrangements while favoring antiparallel. This alignment was further supported by specific interhelical salt bridges as well as specific combinations of alanine and isoleucine at the interface [80]. Additional interactions specifying the antiparallel assembly were dissected by Hadley et al. [74]: vertical interactions between e.g. *d'-d-d'* are important in modulating the helix directions. By applying many of the previously described restrictions and by including negative design features, a homodimeric antiparallel assembly was successfully created by Negron and Keating [81]. Experimental data for all these designs indicated the formation of an antiparallel two-stranded coiled coil, though no structures had been verified experimentally. Nevertheless, all these studies on two-stranded coiled coils helped to further expand our knowledge and established general design rules to create stable assemblies with targeted orientation and oligomeric state [82, 83, 50].

Beyond Two-Stranded Coiled-Coil Designs

Based on the achievements in two-stranded coiled-coil design, also several higher order assemblies with up to 12 strands were created [84, 85] and next was the challenge to create switchable assemblies. Coiled coils were designed that switch their oligomerization state depending on the pH [86], open up in the presence of a binding partner [87], or assemble upon metal binding [88]. Also, a first photoswitchable coiled coil was successfully designed with an N-terminally linked chromophore to induce disassembly upon irradiation [89]. Another increasing scope of application are nanomaterials or nanostructures. Large self-assembling filaments were created [90] and coiled coils were used as building blocks in the construction of molecular scaffolds [91, 92]. These protein origamis are highly versatile in their three dimensional shapes and can be applied in several ways [93]. On the basis of all these coiled-coil designs and the established toolbox of scaffolds, several different molecules for biomedical applications were created. Drug delivery systems, nanoparticles, hydrogels and many more are promising technologies which use coiled coils as structural units or regulatory elements [94, 95, 96]. This shows the extraordinary potential of this supersecondary structure element and highlights its versatility for different applications.

Chapter 2

The (Inverse) Protein Folding Problem

The general elements and parameters of proteins are already well understood, but the detailed knowledge about how the amino acid sequence exactly specifies the proteins' structure and function, and how nature finds this particular conformation in the large set of possible interactions is still challenging (Fig 2.1). This scientific issue is summarized in the **protein folding problem**, which can be distinguished into three main questions [97]:

1. Can we determine physicochemical properties which encode the folding of a protein?
2. What are the folding mechanisms proteins use in nature?
3. Based on these findings, is it possible to predict a proteins' structure only based on its amino acid sequence?

The first of these questions is universally recognized as understood: the forces which stabilize and direct protein folding are mainly hydrogen bonds, the hydrophobic effect, electrostatic interactions, van der Waals forces as well as disulfide bonds as described in section 1.2.1. Research in the field of protein folding and protein prediction is still a challenging field, but in the last years remarkable progress has been made and recent progressions allows the argumentation that the protein folding problem is solved. Nevertheless, new challenges are out there in protein science. With the detailed understanding of proteins, a new question came up: Can we create new proteins by ourselves solely based on physical determinants? This field of **protein design**, which can also be denoted as the **inverse protein folding problem**, is a fairly young research field and opens up many possibilities for protein science (Fig. 2.1).

2.1 Mechanisms of Protein Folding

If we are to understand protein folding, it is important to understand the challenge of it. Starting with a gedanken experiment, like Cyrus Levinthal did in 1969 [98]: Imagine a polypeptide chain with 100 amino acids and consider that every amino acid can adopt only three different configurations. Folding of this protein would have 3^{100} different conformations. Assuming that every 10^{-13} s a new conformation can be tested, it would take $5 \cdot 10^{34}$ seconds or $1.5 \cdot 10^{27}$ years to find every possible fold for this sequence. But in nature proteins fold within milliseconds to seconds into their native structure. Therefore, this *Levinthal paradox* raised the question how a protein can fold on such a short time scale and what mechanisms apply in nature to speed up folding. Obviously, this can not be achieved by random search and Levinthal proposed the existence of kinetically controlled folding pathways [99]. Since this question was raised, many different models and mechanisms for protein folding were proposed, the most common being described in the following.

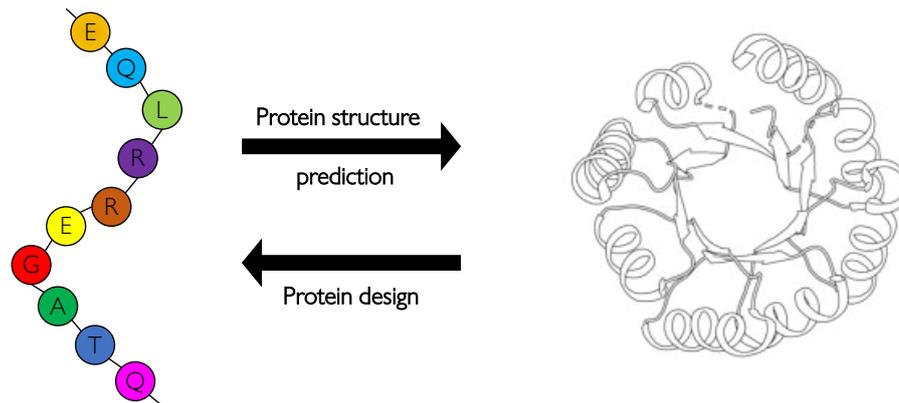


Figure 2.1: The Protein Folding Problem. Starting from a protein sequence (left) the protein folding problem aims to predict the three dimensional structure. Inversely, protein design launches from a determined protein conformation and finds the optimal amino acid sequence to stabilize this structure.

Already Levinthal suggested the gradual formation of local structures at the beginning of protein folding, which function as nuclei for further structure formation [100]. This nucleation growth model was extended to the framework model, implying the independent formation of secondary structural elements which finally organize to the native state [101]. Another hierarchical process is proposed in the diffusion-collision model, a variation of the framework model [102]. It implies diffusion as the driving force for the organization of local structures forming intermediates, which finally coalesce to the native structure. In contrast to these models, which imply the initial formation of secondary structures, the hydrophobic collapse model puts the tertiary structure first [103]. Driven by the hydrophobic effect the polypeptide chain collapses into a molten globule like state, which reorganizes later to build the native protein.

2.1.1 Folding Funnels and the Energy Landscape Model

The previously described models all assume a single folding pathway for each protein through defined intermediates. A more general view on protein folding was encouraged by Anfinsen's dogma. It states that the structure of a protein is solely defined by its amino acid sequence and that the native state is the global minimum of free energy [104]. Incorporating this into protein folding theory led to the folding funnel theory, which states that proteins find their native state by minimizing their free energy [105]. In this context, protein folding can be represented in folding funnels relating the possible conformations with its free energy (Fig. 2.2). In this kind of depiction the depth represents the energetic stabilization of the native state and the width represents the conformational entropy. These folding funnels are no flat 'golf courses' with a single energy minimum as a single pathway would produce, but rather are rough funnel-shaped energy landscapes with multiple energy minima directing the protein folding to its native state. The energy landscape theory proposes that proteins fold by the directional formation of either short- or long-range interactions or packing which guides the polypeptide along the funnel towards its global energy minimum [106]. On this way, also local energy minima can be occupied representing either on- or off-pathway intermediates. It is difficult to state, which of these models is the most accurate or if a combination of these describes reality, or if different models apply for different proteins. Nevertheless, with the progression of the energy landscape models new possibilities arose to model protein folding and to predict protein structures.

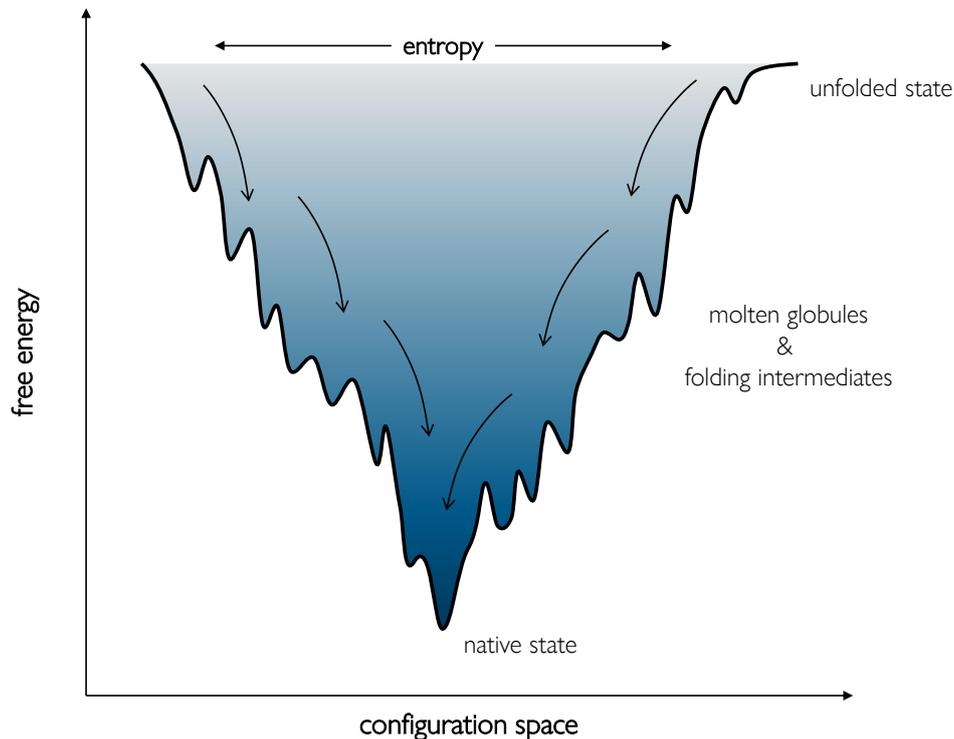


Figure 2.2: Protein Folding Funnel. A schematic representation of a protein folding funnel exemplifying the possible trajectories of an unfolded amino acid chain to its native state. Unfolded states have a high free energy and can occupy a large set of different configurations. Upon protein folding partially folded configurations are formed and can be trapped in local energy minima, which are associated with folding intermediates or molten globules. The completely folded, native structure is found at the global energy minimum.

2.2 Protein Structure Prediction

The thermodynamic hypothesis of Anfinsen is the basis for most protein structure prediction algorithms. By sampling alternative conformations and scoring them using appropriate scoring functions, the lowest energy state can be determined which is assumed to be the native state of the protein. Generally, two different approaches are distinguished: template-based and template-free strategies.

2.2.1 Template-based Modelling

Template-based modelling is built on the finding, that structure is more conserved than sequence [107]. Template structures for a target sequence are identified based on sequence identity, which are subsequently used to direct the structure prediction. A critical step is the selection of an appropriate template either by single sequence searches or by using sequence profiles to identify sequence-structure compatibility. Following, the target sequence is aligned to one or multiple template structures and deviations are identified. Using the known structures those variations, like mutations, deletions or insertions, are accounted by modelling. This strategy can be applied for about two third of known proteins, as a sequence identity of at least 20 % is required, although recent advances also softened this restriction [108].

2.2.2 Template-free Modelling

The prediction of a protein structure solely based on its sequence without any further information is the greatest challenge of the protein folding problem. In *ab initio* approaches a number of possible conformations is sampled followed by model selection based on scoring functions. Due to the large number

of possible conformations of a polypeptide chain, the conformational sampling is highly complex and therefore the exploration of low-energy states is performed by conformational search methods which are guided by energy functions.

Energy Functions

Generally two types of energy functions are distinguished. Physics-based functions determine the interaction between two atoms solely based on quantum mechanics and the coulomb potential. Due to the large number of atoms in a protein, this is very complex and computational expensive. To reduce this complexity atoms can be treated as point-particles interacting through defined potential forms creating so called force fields. Thereby, parameters like bond lengths, angles, van der Waals interactions or electrostatics are covered. While this already reduces the complexity drastically, the energy calculation of a medium-size protein is still impossible today. In knowledge-based energy functions, frequencies of observed structural features from known proteins are related to its energy. Typical terms are the interaction probability between two amino acids at a specific distance, secondary structure propensities and ideal geometries. Recent structure prediction methods often apply combinations of physics- and knowledge-based energy functions.

Conformational Search

The sampling of lowest-energy structures is directed through sampling methods, which define how new conformations are calculated based on a given state. In Molecular Dynamics (MD) simulations trajectories are created by applying Newton's equation of motion iteratively to all atoms, thereby changing the protein structure. This technique does not only determine a final structure but captures structure changes over time. Due to the size of a protein solving these equations for every containing atom is computationally highly expensive and only timescale of femtoseconds can be modelled. To simulate a complete protein folding requires times of at least micro- to milliseconds.

Monte Carlo based approaches sample the conformational space much more efficiently through indiscriminate changes of the protein structures with the application of a multi-step algorithm. In each step a random position in the polypeptide is randomly altered to produce a new conformation, which is only accepted if the energy is reduced for this state. As energy landscapes are rather rough this would lead to the trapping in a local energy-minimum and the global minimum wouldn't be sampled. Therefore, the Metropolis-Hasting criterion is typically included in the acceptance step. As long as the perturbation reduces the energy it will still be accepted, but if the energy is not reduced it will be accepted with a probability related to the Boltzmann-formula with k being the Boltzmann constant and T the temperature:

$$P = \exp\left(-\frac{\Delta E}{kT}\right) \quad (2.1)$$

Additionally, the implementation of Simulated Annealing, which applies initially a high temperature to enable large changes to sample a large conformational space, followed by step wise reduction of temperature, improves the sampling of lowest-energy states.

Even though no structural templates are applied, local structural features are typically predicted based on multiple sequence alignments and statistical analysis of protein structures. Therefore, often information such as secondary structure, residue-residue contacts and inter-residue distances direct the conformational search.

2.2.3 The Rosetta Algorithm

One widely used *ab initio* prediction program is Rosetta, which creates protein models by assembling small fragments of typical three and nine residues taken from known protein structures [109]. These fragments account for the structural bias of local sequences [110]. For each sequence stretch in the target protein 200 fragments are picked from known protein structures based on sequence similarity and prediction of local structural features. The conformational search commences with an extended polypeptide chain and folds the protein subsequently using a Monte Carlo based algorithm. Initially, for a randomly selected sequence stretch a fragment is chosen from the corresponding fragment library. Thereby, the torsion angles of this polypeptide section are replaced by the ones of the selected fragment and energies before and after the change are compared. According to the Metropolis criterion the step is either accepted or rejected. This is repeated 30 000 times with nine-residue fragments, followed by a more fine-grained sampling with 10 000 three-residue fragment insertions [111, 112]. The fragment assembly is performed with a centroid representation and creates only low resolution models of the protein backbone [108]. In an atom-detailed refinement the backbone is perturbed only slightly to maintain the global conformation and side chain conformations are sampled by a simulated annealing Monte Carlo algorithm. As final step the generated models are further minimized to find local energy minima by a gradient minimization approach. During the conformational search different energy terms are used due to the different representations of the polypeptide chain. In the initial coarse-grained fragment search a knowledge-based search function is applied, which considers electrostatics, hydrogen bonding, solvation and steric clashes based on probabilities derived from known protein structures. In the following high resolution refinement an all-atom energy function is applied, which combines physical-based energy terms as well as knowledge-based probabilities [113].

2.2.4 Recent Advances

Recent progress and innovations in protein structure prediction algorithms are frequently tested and compete in the biennial CASP ("Critical Assessment of protein Structure Prediction") workshop [114]. Beside Rosetta many other highly promising and successful algorithms are frequently presented and applied. In recent years the greatest progress was achieved in the prediction of long-range interactions based on evolutionary coupling. Even though this was tried for many years, the extraction of coevolutionary information from multiple sequence alignments and filtering indirect effects was only made possible through the combination of statistical methods with machine learning approaches [115]. In CASP13 a tremendous improvement in contact prediction was observed fueled by the inclusion of deep neural networks [116, 117]. Furthermore, a deep-learning approach was presented in the same year, which outperformed all other algorithms dramatically: AlphaFold is based on a convolutional neural network, which predicts inter-residue distances [118]. The conformational search is similarly to Rosetta based on a fragment assembly approach, but in contrast fragments are created *de novo* using deep learning algorithms and folding is performed by a simple gradient-descent algorithm. This work inspired the research in this field drastically and the application of machine learning was expanded. In the subsequent CASP14, an overall improvement of predictions was observed due to the prediction of interresidue orientations [119]. Nevertheless, one algorithm outperformed all other significantly and achieved unprecedented model accuracy: AlphaFold2 [120, 121]. In contrast to the preceding version, an end-to-end trained system is used, which incorporates several coupled attention-based neural networks [122, 123]. Meaning not only specific features are predicted and later incorporated in the conformational search, but the complete three-dimensional structure is sampled. Once the neural network converged, the model is refined using physical force fields to improve accuracy and stereochemical quality. With this the 50 year old protein folding problem is recognised to be solved by AlphaFold [124]. However, the field of protein prediction is not at its finish line. Recently the AlphaFold Protein Structure Database was released, which comprises all predicted protein structures of the known human proteome and other biologically-significant organisms and will be expanded to capture

all proteins we know [125]. This compendium of protein structures will enable further research promoting the understanding of diseases and providing even deeper understanding of proteins. Furthermore, a new challenge will be the fine-grained prediction of protein interactions, ligand binding and protein functions.

2.3 Protein Design

The design of proteins is a highly versatile and challenging field and is often designated as the inverse protein folding problem. In contrast to the prediction of a tertiary structure from the primary structure, the quest is to find an amino acid sequence which stabilizes a specific protein state. This adds additional complexity to the search problem as not only the flexibility of the single amino acids has to be considered but additionally the protein sequence is variable. During the design the protein sequence can be varied and at every position one of the 20 different amino acids can be placed, leading to a sequence space of 20^n for a protein of n amino acids. Considering a protein of a length of only 100 residues, this leads to a sequence search space of 20^{100} which exceeds even the number of particles in the universe. Therefore, it is even more important to have efficient search methods and evaluation tools to find a sequence which has an energy minimum compatible with the target structure.

2.3.1 De Novo Protein Design

Even though nature already supplies us with a tremendous variety of different proteins they only cover a tiny range of the possible conformational and sequence space of the protein universe. *De novo* protein design is the ultimate test of our knowledge of protein folding, as it targets the design of proteins solely based on our gained expertise of protein stabilizing forces and fold determinants without applying any specific sequence information from nature. Frequently, known protein folds are recapitulated to understand their determinants in more detail, but also the design of completely new proteins was accomplished.

A typical design strategy starts with the selection of a specific target framework and the creation of a backbone. Similar to structure prediction approaches, this can be done by assembly of protein fragments. In RosettaRemodel the backbone conformation is indicated by its secondary structure and additional definable constraints, like inter-residue distances [126]. Based on these definitions a backbone is assembled from fragments, which are picked based on their compatibility to the intended geometries. However, not all imaginable folds are actually realizable in nature, as a proper packing of the core and hydrogen bonding is necessary a folded and stable protein [127]. Therefore, the identification of physically realizable backbones is critical and different rules have been identified to guide the three-dimensional arrangement of secondary structure elements [49, 128]. Some protein folds can also be described through mathematical models, such as coiled coils [54]. Based on these parametrizations new arrangements can be created and applied to protein design.

The actual sequence design uses almost the same modelling tools as previously described for protein structure prediction. Typically, the sequence search is performed on a fixed backbone and only side chain conformations are sampled using rotamer libraries. Rosetta applies once more a Monte Carlo algorithm: randomly a position is mutated and additionally a random rotamer is selected, in the following this step is either accepted or rejected based on the Metropolis criterion [108]. However, as optimal side chain interactions are highly dependent on a precise three-dimensional orientation, often an iterative backbone sampling is included in the sampling strategy to find the optimal sequence [129]. Ranking and evaluation of the designed sequence and backbone is performed with the same all-atom energy function in Rosetta as applied in structure prediction, which combines physics- and knowledge-based terms [113]. Additionally to Rosetta many other protein design strategies have been developed also with great achievements [108].

2.3.2 Applications of Protein Design

The design of proteins, either *de novo* or engineering of existing ones, has a broad range of applications and opens up new possibilities in research, medicine and industry. Starting with the purpose and future of *de novo* designed proteins: they test our knowledge and understanding of protein folding and demonstrate which protein features need to be further investigated [130]. Furthermore, with the creation of new protein scaffolds tailor-made backbones are generated without any impairment of prior natural specificity [131]. One major task of protein design and engineering is the stabilization and optimization of existing proteins, either natural ones or previously designed ones. As natural proteins often have a low thermostability its increase is beneficial for applications in biotechnology. In contrast, designed proteins exhibit frequently a high thermal stability combined with a low conformational stability. Increasing both stabilities is a highly interesting field, again teaching us a deeper understanding of stabilizing forces and providing improved protein scaffolds for new applications [132]. Creation and alteration of ligand binding led to the development of imaging reagents and diverse biosensors [133, 134]. The design of protein switches created molecules, which undergo large conformational changes upon either ligand binding or environmental changes [135, 136]. This requires a multi-state design approach including negative and positive design of specific features. Beyond sole binding of ligands is the design of novel enzymes [137]. The creation of a catalytic mechanism requires the binding of not only one molecule, but must consider the substrate, stabilize transitions states and have a reduced affinity to the product. As such it is a highly complex system. Nevertheless, some *de novo* enzymes could be designed, catalysing new reactions not observed in nature [138, 139, 140]. Furthermore, protein design can also be applied to control protein-protein interactions. Modulation of binding specificity was used to create protein logic gates which can control gene transcription [141]. Additionally, protein design has a wide application range in the development of therapeutics. Targeting specific disease-related proteins, antibodies can be designed to disrupt relevant interactions. Antibodies and other proteins can be designed and optimized to engage specific protein targets to intervene in disease-associated pathways. One highly promising research created bispecific antibodies which can be applied in immunotherapy to target cancer cells more specifically [142]. Recently, the design of a miniprotein inhibiting the SARS-CoV-2 spike protein highlighted the possibilities of this research area [143]. These examples for the utilization of protein design in different research areas is only a little excerpt of the constantly extending range of applications.

2.3.3 Recent Advances and Perspectives

As protein structure prediction and design are highly correlated with each other, noteworthy advances have been achieved in both areas in the last years. Overall in the last decade the complexity of the designed proteins increased, including larger proteins with more diverse secondary structure arrangement [133, 48]. A major contribution to this is the significant technological progress of our time enabling the solution of more complex computational problems. Therefore, more detailed energy functions could be developed, which enable the treatment of proteins on atomic level leading also to improved algorithms and reduced requirements for user-defined constraints to achieve successful designs.

As recent years have shown that the design of protein structure is highly successful new perspectives are the design of functions and the creation of novel enzymes. Therefore, especially the implementation of multi-state design algorithms is a great achievement and might fuel further research towards tailor-made functional proteins. Finally, as the use of deep learning techniques has pushed protein structure prediction tremendously, a similar influence could be expected on protein design. First approaches have already started to use neural networks in the design of favorable sequences, and it will be intriguing to see how it will influence future research [144, 129].

Synopsis

With the successful *de novo* design of the TIM barrel sTIM11 an excellent scaffold for tailor-made proteins was created. Using this idealized protein as foundation the research presented in this thesis aimed to improve its stability determinants and further diversify the set of *de novo* TIM barrels.

The objective of Paper I and II is the stabilization of the initial design. Thereby, the stability determinants of this outstanding protein fold were explored and a set of highly stabilized proteins was created. A short introduction into fine-tuning protein stability and an overview of the results is given in Chapter 3 - "Improving the Conformational Stability of a *De Novo* TIM Barrel".

With the aim to increase the surface of the *de novo* TIM barrel and creating shielded areas, sTIM11 was further diversified by insertion of antiparallel coiled coils into $\beta\alpha$ -loops. Stabilizing interactions determined in Paper II were also incorporated into these designs to achieve a more stable TIM-barrel scaffold. This work is summarized in Chapter 4 - "Extending the *De Novo* Designed TIM Barrel" and Paper III includes the detailed results and discussion.

The review, Paper IV, discusses the importance of an interconnected analysis of evolution, folding and design for a thorough understanding of proteins based on the research on TIM barrels. In the context of Paper I, II and III especially the detailed understanding of folding and stability as well as its value for the design of new proteins is highlighted. A short outline is given in Chapter 5 with a focus on the implications for protein design.

Author Contributions

This thesis is based on the following publications, referred to by their roman numerals.

Paper I: The stability landscape of de novo TIM barrels explored by a modular design approach

S. Romero-Romero, M. Costas, D.-A. Silva Manzano, S. Kordes, E. Rojas-Ortega, C. Tapia, Y. Guerra, S. Shanmugaratnam, A. Rodriguez-Romero, D. Baker, B. Höcker, D. A. Fernandez-Velasco
Journal of Molecular Biology, 2021, 433(18)

In this work I performed structural characterization of DeNovoTIMs by screening crystallization conditions for several proteins, optimization of crystallization, data acquisition, data processing and refinement. Furthermore, I supported the manuscript drafting. S. R.-R. performed protein production, DSC measurements, chemical unfolding, data analysis, participated in designing the study, participated in manuscript drafting and wrote the manuscript. M.C. participated in data analysis, supervised and financially supported DSC measurements D.-A. S. M. developed Rosetta code. E. R.-O. and C.T. performed repetitions of chemical unfolding and DSC measurements. Y.G. screened crystal conditions for some DeNovoTIMs. S.S. participated in structure determination with suggestions for crystal and data improvement and supported data collection of DeNovoTIM13. A. R.-R. provided crystallography sources and materials as well as X-ray beamtime for crystal testing. D. B. provided financial sources, supervision in Rosetta design and supported the study design. B.H. provided financial sources, supervision and participated in conceptualization and manuscript drafting. D. A. F.-V. performed thermodynamic data analysis, provided financial support, supervision, design of initial constructs and preliminary characterization, participated in manuscript drafting, designed this study.

Paper II: A newly introduced salt bridge cluster improves structural and biophysical properties of de novo TIM barrels

S. Kordes*, S. Romero-Romero*, L. Lutz, B. Höcker
Protein Science, 2022, 31, pp. 513-527

*authors contributed equally to this work

In this work I created the protein constructs, established expression and purification protocols, performed biochemical and stability analysis, did the crystallization and structure determination and wrote parts of the manuscript. S.R.R. performed biochemical and stability analysis, participated in crystallization and structure determination and wrote parts of the manuscript. L.L. performed biochemical and stability analysis under my supervision. B.H. designed the study, provided financial support, and supervision.

Paper III: Extension of a de novo TIM barrel with rationally designed antiparallel coiled-coils

S. Kordes, M. Flecks, L. Lutz, B. Höcker
Manuscript

In this work I participated in the study design, performed the Rosetta design, did the cloning of the constructs, established expression and purification protocols, performed biochemical analysis and attempted different structure determination approaches. Furthermore, I wrote the manuscript and did supervision. M.F. participated in establishing expression and purification, biochemical analysis and structure determination under my supervision. L. L. participated in establishing expression and purification

and biochemical analysis under my supervision. B. H. provided the study design, financial support and supervision.

Paper iv: Evolution, folding, and design of TIM barrels and related proteins

S. Romero-Romero*, S. Kordes*, F. Michel*, B. Höcker
Current Opinion in Structural Biology, 2021, 68, pp. 94-104
*authors contributed equally to this work

For this review I participated in concept design, wrote parts of the manuscript, specifically the design section and conclusion. S.R.R., F.M. and B.H. participated in the concept design and wrote parts of the review.

Chapter 3

Improving the Conformational Stability of a De Novo TIM Barrel

A proteins' stability determines its ability to perform a specific function, the timescale on which it is stable and the pathway it takes upon folding. Therefore, an exact balancing of forces is crucial for proper functioning of proteins. Detailed analysis of natural proteins revealed only a marginal stability of active protein conformations of 5 to 15 kcal mol⁻¹ [145]. A protein under physiological conditions is predominantly folded in its active conformation, but undergoes certain molecular fluctuations thereby producing partially or completely unfolded states. The stability of the active state is therefore referred to as the conformational stability, which compares the folded to the unfolded state. The main indicators for the conformational stability are the thermal melting point T_m and the free-energy change upon unfolding ΔG . Both parameters can be determined experimentally and are interconnected through the Gibbs Helmholtz equation, which expresses dependency of ΔG on the temperature.

3.1 Project Objective

As protein stability affects not only the resistance to temperature, denaturant, proteases or pH variations, but also influences solubility, *in vivo* yields and half life, its adjustment is highly interesting for biotechnological applications [132]. Even though rationalization of protein stabilization is difficult, several strategies were developed to modulate the stability landscape of existing proteins. In general, a protein can be stabilized either by altering the entropy or enthalpy. Entropy-driven stabilization aims to restrict the conformational space, whereas the enthalpy-driven approaches are more rational and intent to optimize specific interactions [146]. An early strategy for this purpose was the substitution of glycines with alanines. Due to the reduced flexibility upon mutation the chain entropy of the unfolded states is reduced and the equilibrium is shifted towards the active conformation [147, 148]. In a similar manner, the introduction of proline residues restricts the backbone geometry and as a consequence decreases the entropy of unfolding [149, 150]. The installation of additional disulfide bonds is one of the most frequent approaches and can have entropic as well as enthalpic effects on the stability. The induced alterations are more complex ranging from a modified hydrophobicity of all states over reduced flexibility near the covalent linkage in the native state to the generation of residual structure in the denatured state [151, 152]. Additionally, more rational approaches aim to tune the enthalpic effect using information of the three-dimensional structure. Incorporation of additional hydrophobic residues on the protein surface improves hydration and thereby influences stability [146]. Also the introduction of rational mutations to generate specific interactions or to improve the hydrophobic packing are frequently applied in protein stability design studies [153]. The emerging protein structure prediction and design methodologies also fuelled the design of stabilized proteins and the launch of energy functions made it possible to rank different mutants and determine the potentially most stabilizing variants *in silico*.

Nevertheless, stabilization of proteins is not that simple due to the complex synergy of multiple interactions in a three-dimensional structure. Even though the creation of proteins from scratch is often successful and provides us with stable, well-folded proteins with nature-like features, detailed analysis of several *de novo* proteins revealed non-natural folding landscapes and stability curves. Interestingly, most designed proteins exhibit a high thermal resistance, but combined with low ΔG values as compared to natural proteins. This observation is probably due to the fact that most design approaches focus on a positive design strategy to stabilize the native state [132]. Thereby, the energy gap between the folded and unfolded state is comparably small and the free energy upon unfolding stays low. This is also observed for the *de novo* designed TIM barrel, which was introduced in section 1.3: Detailed analysis of sTIM11 revealed a high thermal stability ($T_m = 80^\circ\text{C}$) combined with a low free-energy of unfolding ($\Delta G_{25^\circ\text{C}} = 4 \text{ kcal mol}^{-1}$) [48]. Concluding, the conformational stability of sTIM11 is significantly lower than determined for natural TIM barrels. Hence, the objective of this project was to fine-tune and increase the stability of sTIM11 in order to create suitable proteins for later diversification and ultimate creation of functional TIM barrels. For this purpose, two different approaches were used to improve the stability determinants of *de novo* TIM barrels. First, using a modular design strategy the hydrophobic packing of sTIM11 was increased step wise (Paper I). Following up on this, a salt bridge cluster was introduced into different *de novo* TIM barrels to combine the effect (Paper II).

3.2 DeNovoTIMs - Improved Hydrophobic Packing

A rather recently described stabilizing element in protein structures are hydrophobic clusters, which are formed by cooperative, long-range interactions in the proteins hydrophobic core [154]. They create crucial stabilization centers comprising hydrophobic or aromatic amino acids. Their implications for protein stabilization and utilization in stability design were described for several different protein folds [155, 156, 157, 158]. Analysis of natural TIM barrels also emphasized their importance for stability and folding of this specific fold [159, 160, 161].

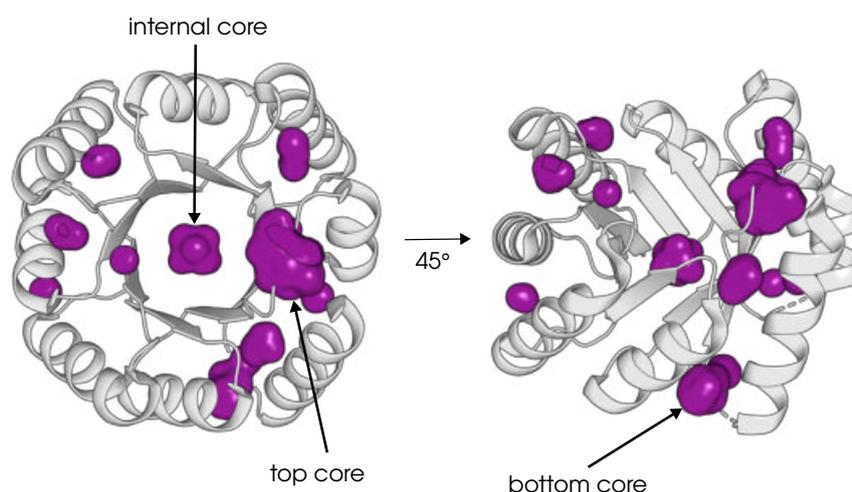


Figure 3.1: Cavities in sTIM11. The crystal structure of sTIM11 (PDB ID: 5BVL) is shown from top view and rotated by 45° . Identified internal cavities are shown in purple spheres and targeted cores are marked with arrows indicating the used nomenclature.

Detailed analysis of sTIM11 revealed a diminished hydrophobic packing in the various regions of the TIM barrel manifested in small cavities in the hydrophobic core. Three separate regions were determined to have a reduced packing: the internal core - located at the inner face of the central β -barrel -,

the bottom core - formed at the N-terminal side of the β -sheets at the interface to the α -helices -, and similarly on the opposite site of the TIM barrel the top core - situated between the C-terminal ends of the β -strand and the surrounding α -helices (Fig. 3.1). In a preceding adjustment of sTIM11 two cysteine residues which did not form the intended disulfide bridge were mutated producing sTIM11noCys. Following, the core regions were targeted in a fixed-backbone design using Rosetta aiming to increase the hydrophobic packing by applying specific constraints. In this first design round the different regions were targeted separately and the most stabilizing design variants were determined experimentally. Finally, the best designs were combined to produce double- and triple-region designs to ultimately create a family of fourteen DeNovoTIMs. Thermodynamic analysis of all designs revealed a wide coverage of thermal and conformational stabilities with T_m values ranging from 47 to 109 °C and $\Delta G_{25^\circ\text{C}}$ between 1.5 and 23.6 kcal mol⁻¹. Furthermore, the DeNovoTIM family exhibits new combinations of thermodynamic parameters as revealed by comparison with natural proteins. Due to the modular design approach it was possible to specify the contribution of each region and an interdependency of the different regions for the stabilization was determined. This epistasis effect was associated with the extension of hydrophobic clusters in the double- and triple-region designs. The determination of crystal structures for three proteins of this collection supported this assumption. For DeNovoTIM13, a double-region design with improved stability features, the high resolution structure revealed an increase of the hydrophobic clusters by 60 % in comparison to sTIM11. This highlights that the modular design strategy successfully improved the hydrophobic packing, stabilized the *de novo* TIM barrel significantly and additionally provided a new set of proteins to investigate stability and folding determinants.

3.3 Introduction of a Salt Bridge Cluster

The effect of salt bridges on the stability of proteins is highly complex and prediction of their contribution is still challenging. These particular interactions are a combination of hydrogen and ionic bonds and are formed between oppositely charged residues which are in close proximity to each other [162]. Typically a distance below 4 Å between donor and acceptor atoms is considered as cut off. In proteins salt bridges are formed between the negatively charged Asp or Glu and positively charged Arg, Lys or His. Besides their potential stabilizing effects they are also important for creation of conformational specificity and can be critical for optimal positioning of functional groups [163, 164, 165]. The contribution of salt bridges to stability is difficult to predict and is not necessarily a favorable one [166, 167, 168]. For their formation, a specific geometry of both residues to each other is required. This involves the ordering of the protein and also desolvation of the side chains, which are both entropic and enthalpic expensive [169]. Therefore, the favorable ionic interaction of the salt bridge has to counterbalance these unfavorable effects and often is only marginally stabilizing. As the burial of charged residues is quite expensive, most salt bridges are surface exposed. Nevertheless, buried salt bridges are observed and were proven to be crucial for conformational specificity of the native state. A possible way found in nature to increase the stabilizing effect of buried salt bridges is the formation of so called networks or clusters [166, 170, 171]. Often a local salt bridge interacts with one distant residue, but also more complex networks were identified in natural proteins. It was found that networked salt bridges are a viable way to interconnect secondary structure elements distant in sequence. Due to their complex geometry salt bridge clusters are even more difficult to predict and design.

The analysis of natural TIM barrels determined several conserved salt bridges being crucial for stability and function [172, 173]. An overall statistical investigation of stabilizing interactions throughout this fold identified an over representation of polar residues and salt bridges in the inner β -barrel [15, 174]. Especially, larger salt-bridge networks were found to stabilize the β -barrel through long range electrostatic interactions. These clusters were found in different subfamilies and are frequently conserved across them. The exact role of these electrostatic interactions remains unidentified but a possible association with improved foldability was proposed [47]. One specific natural protein, which comprises a highly coordinated

salt bridge cluster in the inner barrel, is HisF, a subunit of the indole-3-glycerol phosphate synthase (IGPS) [33]. This cyclic salt bridge is formed by four residues located at the N-terminal end of the β -strands (Fig. 3.2B) [175]. Despite its stabilizing function it also acts as a gate for the ammonia tunnel located in the barrel center. One residue is believed to alter its conformation upon gate opening and in the following enables ammonia transfer [176]. Its high potential to stabilize this fold was also determined in a study which aimed to recapitulate the evolution of TIM barrels by duplication. Fusion of two C-terminal halves of HisF was performed and was stabilized by the introduction of the natural salt bridge cluster [177, 178].

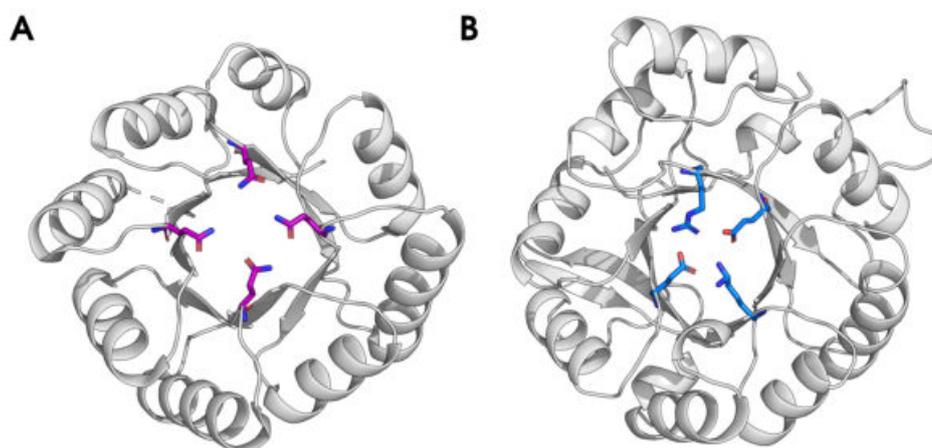


Figure 3.2: Salt Bridge Clusters in TIM barrels (A) Crystal structure of sTIM11noCys (6YQY) is shown from bottom view. The four symmetry-related glutamine residues are shown in stick representation. These positions were mutated for the creation of a salt bridge network. (B) Crystal structure of HisF (PDB ID: 1THF) is shown with residues at the N-terminal face of the β -sheet forming a salt bridge network shown as sticks.

On the basis of this natural salt bridge network we aimed to stabilize the *de novo* TIM barrels structure by introduction of an analogously constructed cluster. Inspection of the sTIM11noCys evinced four symmetry-related glutamine residue at the N-terminal side of the β -barrel in a similar geometric position as the respective residues in HisF (Fig. 3.2). Therefore, these four residues were mutated alternating to Arg and Glu to construct a salt bridge cluster. In order to investigate the effect of this salt bridge cluster in the context of a similar topology but various stability backgrounds we selected three different proteins of the previously created set of *de novo* TIM barrels: sTIM11noCys, the cysteine free variant of the original sTIM11; DeNovoTIM6, comprising several stabilizing mutations in the previously described bottom core; and DeNovoTIM13, a stabilized version with mutations in the top and bottom region. The created proteins were entitled sTIM11noCys-SB, DeNovoTIM6-SB and DeNovoTIM13-SB. The effect of the salt bridge cluster on these three proteins was investigated and showed high divergence. For sTIM11noCys-SB and DeNovoTIM6-SB explicit stabilization by $1.6 \text{ kcal mol}^{-1}$ and $1.9 \text{ kcal mol}^{-1}$ was identified, respectively. In contrast, DeNovoTIM13-SB is destabilized by $-1.3 \text{ kcal mol}^{-1}$ but the introduced salt bridge cluster altered the thermal unfolding process to be reversible. As the base design DeNovoTIM13 is already highly stabilized with a ΔG of $9.5 \text{ kcal mol}^{-1}$, we speculate that the stabilizing effects of the previously improved hydrophobic packing and the introduced salt bridge cluster cancel each other out. Another interesting observation was the influence of the salt bridge cluster on the crystallization behaviour of these proteins. Especially for sTIM11 and DeNovoTIM6 it was challenging to achieve diffracting crystals previously. The introduced cluster enhanced the crystallizability of both proteins and higher resolution structures could be solved. Analysis of high resolution structures of all three proteins revealed dissimilar geometries for each, ranging from a complete absence of a cluster in DeNovoTIM6-SB to highly coordinated networks in sTIM11noCys-SB and DeNovoTIM13-SB. Comparison of the determined geometries with the measured stabilizing effects revealed no correlation. Collectively, this research highlights the complexity of salt bridges in proteins: despite a further stabilization and structural improvement of the

de novo TIM-barrel collection, the exact prediction of the determined geometries was not possible. Taken together, it emphasizes the huge potential of salt bridges and especially salt bridge clusters to stabilize proteins, but also points out the deficiency in the exact understanding and design of salt bridges.

3.4 Conclusion

The stabilization of the *de novo* TIM barrel was achieved in two studies with significantly different approaches. Both, increase of the hydrophobic packing and introduction of a salt bridge cluster, altered the stability of this particular protein significantly and intriguing observations were made. This large family of *de novo* TIM barrel opens up new possibilities to study the stability landscape of this fold in more detail and the members with diverse stability features can be used to create proteins with tailor-made functions.

Chapter 4

Extending the De Novo Designed TIM Barrel

In the beginning, the main objective of *de novo* protein design was to examine the understanding of protein folding, but soon researchers realized the vast potential of these new molecules and astonishing progress has been achieved in the creation of functional *de novo* proteins recently. One category of proteins, which is highly over-represented in this large set of designed proteins, are coiled-coil and helical bundle proteins, mostly due to their excellent perception and control of oligomerization and interaction. In recent years, small-molecule binding helical bundles were reported [179, 180], and led to the establishment of biosensors [134]. Additionally, protein switches and logical circuits on the basis of coiled-coil proteins opened up new possibilities in biotechnology [181, 182, 183] and finally also *de novo* enzymes were designed [184, 140, 185]. Another category, highly successful in different applications, are mini-proteins which are defined as short polypeptides with less than 40 amino acids. Despite their small size, they are adopting a defined tertiary structure with at least two secondary structure elements, thereby comprising a hydrophobic core and exhibit cooperative folding [186]. The application of mini-proteins as therapeutics was already shown some years ago and most recently they were applied as inhibitors for SARS-CoV-2 [187, 143]. Even though in nature larger globular proteins are more frequent the design of these complex structures is challenging due to the larger structure and sequence space. Most globular *de novo* proteins were designed as idealized matches of their natural counterparts, such as idealized β -barrel, β -sheet and TIM-barrel proteins [133, 188, 48], but also new topologies were created [131]. Some of these idealized proteins were functionalized, for example by incorporation of fluorescence activation into the *de novo* β -barrel [133]. More recent approaches already embedded the functional motif in the initial design approach and successfully created new biosensors [189].

4.1 Project Objective

Through the *de novo* design of the idealized TIM barrel sTIM11 the main determinants of this ubiquitous fold could be identified. Our research on further stabilizing this fold, which is outlined in the previous chapter, resulted in optimized variants and diversified this *de novo* protein family. The next ambitious goal in designing TIM-barrel proteins is the incorporation of binding sites and ultimately the creation of a *de novo* enzyme. As sTIM11 is an idealized, minimalistic scaffold, it lacks cavities or pockets which could be utilized for this approach. Therefore, an initial step towards the functionalization of *de novo* TIM barrels is to increase their surface area and to create larger cavities. Studies on natural TIM barrels specified that active and binding sites are located at the top of the barrel, precisely at the C-terminal face of the β -strands and the continuing $\beta\alpha$ -loops (Fig. 1.3). However, during the design of sTIM11 these regions were mostly ignored resulting in short loops (Fig. 4.1A). This emerges even more when sTIM11 is compared to natural TIM barrels (Fig. 4.1B). The later ones often comprise extended unstructured as well as structured loops, and even complete domain insertions in these regions. Evidently, the extension and diversification of the

$\beta\alpha$ -loops are the first step to functionalize the *de novo* TIM barrels.

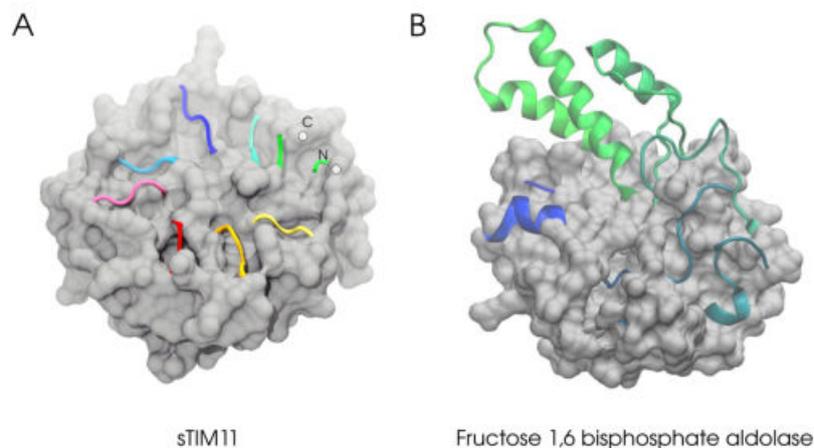


Figure 4.1: Surface of TIM barrel proteins Shown in grey is the surface of the TIM-barrel topology of (A) the *de novo* TIM barrel sTIM11 and in (B) the natural TIM barrel Fructose 1,6 biphosphate aldolase (PDB ID: 1RV8 [190]). In cartoon representation the $\beta\alpha$ -loops are shown.

In previous works, these regions were already submitted to various design approaches. Wiese et al. [191] introduced a small helix into one $\beta\alpha$ -loop by applying a design pipeline on a small $\beta\alpha\beta$ fragment of the four-fold symmetric protein. However, the determined crystal structure differed from the predicted Rosetta model: the ensuing α -helix is shifted significantly to the bottom of the TIM barrel. Nevertheless, this work successfully inserted a small structural element into sTIM11noCys without influencing the stability, producing a first diversified *de novo* TIM barrel with increased surface and demonstrating the applicability of the $\beta\alpha$ -loops for this purpose. In another work, a complete domain was inserted into the $\beta\alpha$ loop of DeNovoTIM15, a further refined DeNovoTIM variant [192]. Availing the ability of DeNovo-TIM half barrels for homo-dimerization, half-DeNovoTIM15 was fused with a *de novo* ferredoxin thereby creating an enclosed chamber between the two domains. Further engineering of this region enabled the binding of lanthanide ions with high affinity, producing the first functionalized *de novo* TIM barrel. Carrying the diversification of *de novo* TIM barrels forward we aimed to introduce larger structured elements into the $\beta\alpha$ -loops of sTIM11noCys thereby creating pockets and interaction areas for later functionalization. Due to the well understood determinants of coiled-coil structures and their diverse applicability, these supersecondary structure elements are excellent candidates for this purpose.

4.2 Summary of Results and Conclusion

Using a multistep design approach antiparallel coiled coils were introduced into one or two $\beta\alpha$ -loops of sTIM11noCys. Commencing with a previously designed sequence for an antiparallel coiled coil, an *ab initio* model was created and incorporated into the *de novo* TIM barrel. Following adjustments of the linker regions and sequence optimization were performed to generate a set of four coiled-coil TIM barrels (ccTIMs) with one coiled coil. After initial characterization one coiled coil outperformed the other designs and was deployed to create ccTIMs with two coiled coils in opposing loops. The four resulting variants differ from each other by an insertion of three residues (Gly-Leu-Glu, GLE) in another $\beta\alpha$ -loop, the introduction of a potentially stabilizing salt-bridge cluster (SB) as described in Paper II [193] or a combination of both.

Detailed analysis of a set of five ccTIMs indicated the formation of the coiled coil for three of these proteins and therefore implies a successful design. The leading single coiled-coil design, ccTIM2, as well as two double coiled-coil designs, ccTIM5-GLE and ccTIM5-GLE-SB, have CD spectra with increased α -helical signal compatible with the predicted models. Stability analysis by Differential Scanning Calorime-

try (DSC) and chemical unfolding showed improved stability features for these three designs indicating the formation of additional stabilizing interactions. Based on these experimental data we can assume, that the introduced sequences form stable α -helical secondary structure without impairment of the TIM-barrel scaffold. We suppose a coiled-coil arrangement similar to our designs, nevertheless verification by a three-dimensional structure is required. Despite attempting NMR, Small Angle X-ray scattering and X-ray crystallography, we did not obtain structural information yet to ultimately verify the designs. Even though diffracting crystals were acquired for ccTIM5-GLE, they were not reproducible and data quality was not sufficient for structure determination.

Nevertheless, with the presented results we have explicit data supporting our design goal to further diversify the *de novo* TIM-barrel family by increasing its surface area and building accessible cavities. Already with these ccTIM proteins the design of functional *de novo* TIM barrels can be promoted. However, the presented design procedure provides also opportunities to create other extended *de novo* TIM-barrel variants. By introducing two coiled coils into different $\beta\alpha$ -loops, we proved the feasibility of incorporating different modules into the *de novo* TIM barrel without affecting the scaffolds' stability. Therefore, it would be an interesting approach to combine this work with the previous work from Wiese et al. [191] and to introduce the reported short α -helix as well as the coiled coils to create an even more diverse protein. Furthermore, to extend this modular strategy and utilize the four-fold symmetry of the *de novo* TIM barrel the creation of a ccTIM with four coiled coil inserted in all four symmetry related $\beta\alpha$ -loops would spawn a remarkable protein 4.2). First design steps were already performed and showed the compatibility of this intended topology with the *de novo* TIM barrel. With the insertion of four coiled coils a large cavity in the center between them is formed which could be utilized for the creation of a large hydrophobic pocket for ligand binding and catalytic activities (Figure 4.2). In order to advance this idea, an explicit sequence redesign has to be performed to establish a stabilizing network spanning the four coiled coils to create a stable, folded protein. In summary, our approach of inserting coiled coils into the *de novo* TIM barrel conceived a first set of promising proteins for further designs and opened up new avenues for further diversification of this remarkable protein family.

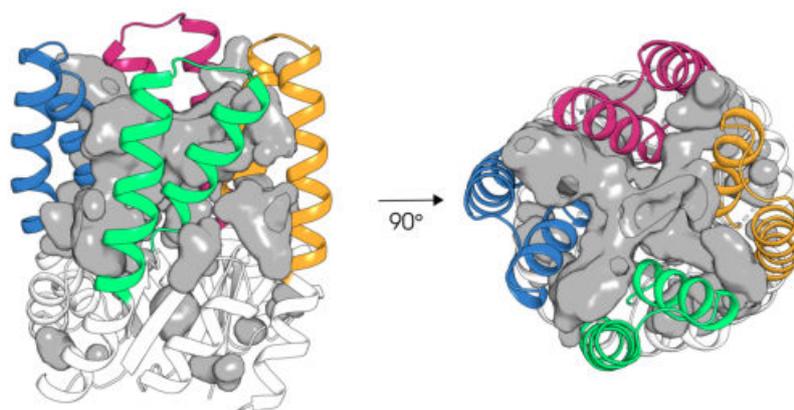


Figure 4.2: Model of a four-fold coiled-coil TIM barrel In cartoon representation the TIM barrel is shown in white and four introduced coiled coils are highlighted in the colors orange, pink, blue and green. Potential cavities are shown as surface in grey.

Chapter 5

The Interplay of Protein Evolution, Folding and Design

In protein design the knowledge of several decades of research are brought together to be tested and utilized to generate new proteins. With the exploration of mechanisms in protein evolution interconnections between protein folds were determined, unraveling new routes of protein fold development. Thereby, different evolutionary units have been identified which were on the one hand tested by design approaches but also employed in the creation of new proteins [194, 195]. Additionally, understanding protein evolution gives crucial information about protein folding. As modern, complex proteins have evolved from smaller structural units through recombination, duplication and mutation the investigation of evolutionary related proteins can provide us with specific folding determinants [196, 197]. Information about crucial stabilizing interactions or folding pathways can be tested by protein engineering and design and are also important to be considered in new design approaches.

In this review, we use the TIM-barrel fold to highlight the inter-connectivity between these three research areas. As the TIM barrel is supposed to be one of the oldest and ubiquitous protein folds in nature its evolutionary history was studied extensively [10, 9]. Furthermore, its ubiquity and functional diversity has drawn attention on its folding pathways and stability, and several studies have investigated its folding determinants [198, 23, 199]. This deep knowledge gained through the study of evolution and folding fuelled the design of the *de novo* TIM barrel. Even though, the design of this protein fold was a long standing challenge it also provides us with a great overview how recent findings from different research areas improved this design and finally facilitated the success of the design of the four-fold symmetric TIM barrel sTIM11. Generally speaking, this work aims to provide a wider picture of protein research and highlights how evolution, folding and design improves our understanding of proteins in general.

Chapter 6

Outlook

In this work a family of diversified *de novo* TIM barrels was created with various characteristics. On the one hand the DeNovoTIMs exhibit a wide variety of thermodynamic and conformational stability and thereby also explore new regions of the stability landscape of proteins. Therefore, they are highly interesting to identify modulators of protein stability and to investigate the design of highly stable *de novo* proteins. Additionally, the here applied strategy to determine stabilizing mutations for single areas and in the following combine different stabilized regions is a promising strategy for larger proteins. It would be interesting to test this approach also with other domains or even multi-domain proteins. The introduction of salt bridges into a subset of *de novo* TIM barrels showed the challenge in designing networked interactions, but demonstrated also their relevance in proteins. Even though the utilized base designs have a highly similar structure the geometries of the introduced salt bridges vary significantly and additionally show various contributions to stability. Nevertheless, by introducing only one salt bridge cluster the stability of the *de novo* TIM barrels could be improved significantly by up to 2 kcal mol^{-1} . Furthermore, evaluation of the salt bridge TIM barrels using Rosetta highlighted the challenge to precisely model cluster geometries. Detailed understanding of salt bridges in natural and synthetic proteins will improve the design of such networked interactions and thereby support the construction of stable proteins.

The designed ccTIMs extend this set of stabilized *de novo* TIM barrels to a more diverse family. Antiparallel coiled coils were introduced into the $\beta\alpha$ -loops of the *de novo* TIM barrel with the goal to increase the surface and create shielded areas. The experimental data of three designed ccTIMs support the successful design and prove the formation of additional, stable secondary structure elements that increase the surface area. Even though no experimental structural data could be achieved and no explicit evidence for created cavities are present, the ccTIMs provide an excellent basis for the construction of functionalized *de novo* TIM barrels. Further diversification could be initiated by the introduction of additional coiled coils in the remaining two $\beta\alpha$ -loops, which would lead to an exceptional protein with a large cavity suitable for ligand binding or catalytic activity. Additionally, the here provided designs could be combined with previously reported designs from Caldwell et al. [192] or Wiese et al. [191] creating an even more diverse set of *de novo* TIM barrels.

Bibliography

- [1] Linus Pauling, Robert B. Corey, and H. R. Branson. The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences of the United States of America*, 37(4):205–211, 1951. ISSN 00278424. doi: 10.1073/pnas.37.4.205.
- [2] Linus Pauling and Robert B. Corey. Configurations of Polypeptide Chains With Favored Orientations Around Single Bonds: Two New Pleated Sheets. *Proceedings of the National Academy of Sciences*, 37(11):729–740, 11 1951. ISSN 0027-8424. doi: 10.1073/pnas.37.11.729.
- [3] Gregory A. Petsko and Dagmar. Ringe. *Protein structure and function*. New Science Press, 2004. ISBN 0953918149.
- [4] Jiacheng Li, Chengyu Hou, Xiaoliang Ma, Shuai Guo, Hongchi Zhang, Liping Shi, Chenchen Liao, Bing Zheng, Lin Ye, Lin Yang, and Xiaodong He. Entropy-enthalpy compensations fold proteins in precise ways. *International Journal of Molecular Sciences*, 22(17), 9 2021. ISSN 14220067. doi: 10.3390/ijms22179653.
- [5] Alan Fersht and others. *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*. Macmillan, 1999.
- [6] Vincent J. Hilser, Javier Gómez, and Ernesto Freire. The enthalpy change in protein folding and binding: Refinement of parameters for structure-based calculations. *Proteins: Structure, Function and Genetics*, 26(2): 123–133, 1996. ISSN 08873585. doi: 10.1002/(SICI)1097-0134(199610)26:2<123::AID-PROT2>3.0.CO;2-H.
- [7] Lavi S. Bigman and Yaakov Levy. Entropic Contributions to Protein Stability. *Israel Journal of Chemistry*, 60 (7):705–712, 7 2020. ISSN 1869-5868. doi: 10.1002/IJCH.202000032.
- [8] D. W. Banner, A. C. Bloomer, G. A. Petsko, D. C. Phillips, C. I. Pogson, I. A. Wilson, P. H. Corran, A. J. Furth, J. D. Milman, R. E. Offord, J. D. Priddle, and S. G. Waley. Structure of chicken muscle triose phosphate isomerase determined crystallographically at 2.5Å resolution: Using amino acid sequence data. *Nature*, 255(5510):609–614, 1975. ISSN 00280836. doi: 10.1038/255609a0.
- [9] Reinhard Sterner and Birte Höcker. Catalytic versatility, stability, and evolution of the ($\beta\alpha$)8-barrel enzyme fold. *Chemical Reviews*, 105(11):4038–4055, 2005. ISSN 00092665. doi: 10.1021/cr030191z.
- [10] Nozomi Nagano, Christine A Orenge, and Janet M Thornton. One Fold with Many Functions : The Evolutionary Relationships between TIM Barrel Families Based on their Sequences , Structures and Functions. *Journal of Molecular Biology*, 2836(02):741–765, 2002. doi: 10.1016/S0022-2836(02)00649-6.
- [11] R. K. Wierenga. The TIM-barrel fold: A versatile framework for efficient enzymes. *FEBS Letters*, 492(3): 193–198, 3 2001. ISSN 00145793. doi: 10.1016/S0014-5793(01)02236-0.
- [12] Alexey G Murzin, Arthur M Lesk, and Cyrus Chothia. Principles determining the structure of β -sheet barrels in proteins II. The observed structures. *Journal of Molecular Biology*, 236(5):1382–1400, 1994. ISSN 00222836. doi: 10.1016/0022-2836(94)90065-5.
- [13] Nozomi Nagano, E. Gail Hutchinson, and Janet M. Thornton. Barrel structures in proteins: Automatic identification and classification including a sequence analysis of TIM barrels. *Protein Science*, 8(10):2072–2084, 1999. ISSN 09618368. doi: 10.1110/ps.8.10.2072.
- [14] Roman Urfer and Kasper Kirschner. The importance of surface loops for stabilizing an eightfold $\beta\alpha$ barrel protein. *Protein Science*, 1(1):31–45, 1992. ISSN 1469896X. doi: 10.1002/pro.5560010105.
- [15] M. S. Vijayabaskar and Saraswathi Vishveshwara. Insights into the Fold Organization of TIM Barrel from Interaction Energy Based Structure Networks. *PLoS Computational Biology*, 8(5):e1002505, 5 2012. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1002505.
- [16] Ralf Thoma, Michael Hennig, Reinhard Sterner, and Kasper Kirschner. Structure and function of mutationally generated monomers of dimeric phosphoribosylanthranilate isomerase from *Thermotoga maritima*. *Structure*, 8(3):265–276, 3 2000. ISSN 09692126. doi: 10.1016/S0969-2126(00)00106-4.

- [17] Dominique Maes, Johan P Zeelen, Narmada Thanki, Nicola Beaucamp, Marco Alvarez, Thi Minh Hoa Dao, Jan Backmann, Joseph A Martial, Lode Wyns, Rainer Jaenicke, and Rik K Wierenga. The crystal structure of triosephosphate isomerase (TIM) from *Thermotoga maritima*: A comparative thermostability structural analysis of ten different TIM structures. *Proteins: Structure, Function and Genetics*, 37(3):441–453, 1999. ISSN 08873585. doi: 10.1002/(SICI)1097-0134(19991115)37:3<441::AID-PROT11>3.0.CO;2-7.
- [18] D. Lang, R. Thoma, M. Henn-Sax, R. Sterner, and M. Wilmanns. Structural evidence for evolution of the β/α barrel scaffold by gene duplication and fusion. *Science*, 289(5484):1546–1550, 2000. ISSN 00368075. doi: 10.1126/science.289.5484.1546.
- [19] Todd M. Larsen, L. Timothy Laughlin, Hazel M. Holden, Ivan Rayment, and George H. Reed. Structure of Rabbit Muscle Pyruvate Kinase Complexed with Mn^{2+} , K^{+} , and Pyruvate. *Biochemistry*, 33(20):6301–6309, 5 1994. ISSN 15204995. doi: 10.1021/bi00186a033.
- [20] Brian V Norledge, Anne M Lambeir, Ruben A Abagyan, Antje Rottmann, Anna M Fernandez, Vladimir V Filimonov, Martin G Peter, and Rik K Wierenga. Modeling, mutagenesis, and structural studies on the fully conserved phosphate-binding loop (Loop 8) of triosephosphate isomerase: Toward a new substrate specificity. *Proteins: Structure, Function, and Bioinformatics*, 42(3):383–389, 2 2001. ISSN 0887-3585. doi: [https://doi.org/10.1002/1097-0134\(20010215\)42:3<383::AID-PROT80>3.0.CO;2-G](https://doi.org/10.1002/1097-0134(20010215)42:3<383::AID-PROT80>3.0.CO;2-G).
- [21] Adrián Ochoa-leyva, Xavier Soberón, Filiberto Sánchez, Martha Argüello, Gabriela Montero-morán, and Gloria Saab-rincón. Protein Design through Systematic Catalytic Loop Exchange in the $(\beta/\alpha)8$ Fold. *Journal of Molecular Biology*, 387(4):949–964, 2009. ISSN 0022-2836. doi: 10.1016/j.jmb.2009.02.022.
- [22] Adrián Ochoa-Leyva, Francisco Barona-Gómez, Gloria Saab-Rincón, Karina Verdel-Aranda, Filiberto Sánchez, and Xavier Soberón. Exploring the Structure–Function Loop Adaptability of a $(\beta/\alpha)8$ -Barrel Enzyme through Loop Swapping and Hinge Variability. *Journal of Molecular Biology*, 411(1):143–157, 2011. ISSN 0022-2836. doi: <http://dx.doi.org/10.1016/j.jmb.2011.05.027>.
- [23] Aaron David Goldman, Joshua T. Beatty, and Laura F. Landweber. The TIM Barrel Architecture Facilitated the Early Evolution of Protein-Mediated Metabolism. *Journal of Molecular Evolution*, 82(1):17–26, 1 2016. ISSN 00222844. doi: 10.1007/s00239-015-9722-8.
- [24] Michael Hennig, Bernhard Schlesier, Zbigniew Dauter, Sabine Pfeffer, Christian Betzel, Wolfgang E. Höhne, and Keith S. Wilson. A TIM barrel protein without enzymatic activity? Crystal-structure of narbonin at 1.8 Å resolution. *FEBS Letters*, 306(1):80–84, 7 1992. ISSN 00145793. doi: 10.1016/0014-5793(92)80842-5.
- [25] W. G.J. Hol, P. T. Van Duijnen, and H. J.C. Berendsen. The α -helix dipole and the properties of proteins. *Nature*, 273(5662):443–446, 6 1978. ISSN 00280836. doi: 10.1038/273443a0.
- [26] Soumya Raychaudhuri, Fayyaz Younas, P. Andrew Karplus, Carlos H. Faerman, and Daniel R. Ripoll. Backbone makes a significant contribution to the electrostatics of α/β -barrel proteins. *Protein Science*, 6(9):1849–1857, 9 1997. ISSN 09618368. doi: 10.1002/pro.5560060905.
- [27] R. K. Wierenga, E. G. Kapetaniou, and R. Venkatesan. Triosephosphate isomerase: A highly evolved biocatalyst. *Cellular and Molecular Life Sciences*, 67(23):3961–3982, 12 2010. ISSN 1420682X. doi: 10.1007/s00018-010-0473-9.
- [28] Jingyi Xiang, Ju Yeon Jung, and Nicole S. Sampson. Entropy effects on protein hinges: The reaction catalyzed by triosephosphate isomerase. *Biochemistry*, 43(36):11436–11445, 9 2004. ISSN 00062960. doi: 10.1021/bi049208d.
- [29] Inger Andersson. Catalysis and regulation in Rubisco. *Journal of Experimental Botany*, 59(7):1555–1568, 5 2008. ISSN 00220957. doi: 10.1093/jxb/ern091.
- [30] R. John Ellis. The most abundant protein in the world. *Trends in Biochemical Sciences*, 4(11):241–244, 11 1979. ISSN 09680004. doi: 10.1016/0968-0004(79)90212-3.
- [31] Deirdre R. Coombe and Neha S. Gandhi. Heparanase: A Challenging Cancer Drug Target. *Frontiers in Oncology*, 9:1316, 11 2019. ISSN 2234943X. doi: 10.3389/fonc.2019.01316.
- [32] Safa Kinaneh, Iyad Khamaysi, Tony Karram, and Shadi Hamoud. Heparanase as a potential player in SARS-CoV-2 infection and induced coagulopathy. *Bioscience Reports*, 41(7):20210290, 7 2021. ISSN 15734935. doi: 10.1042/BSR20210290.
- [33] Barnali N. Chaudhuri, Stephanie C. Lange, Rebecca S. Myers, Sridar V. Chittur, V. Jo Davisson, and Janet L. Smith. Crystal structure of imidazole glycerol phosphate synthase: A tunnel through a $(\beta/\alpha)8$ barrel joins two active sites. *Structure*, 9(10):987–997, 10 2001. ISSN 09692126. doi: 10.1016/S0969-2126(01)00661-X.
- [34] Ivan Rivalta, Mohammad M. Sultan, Ning Shiuan Lee, Gregory A. Manley, J. Patrick Loria, and Victor S. Batista. Allosteric pathways in imidazole glycerol phosphate synthase. *Proceedings of the National Academy of Sciences of the United States of America*, 109(22):E1428–E1436, 5 2012. ISSN 00278424. doi: 10.1073/pnas.1120536109.

- [35] M. Michael Gromiha, Gerard Pujadas, Csaba Magyar, Samuel Selvaraj, and Istvan Simon. Locating the Stabilizing Residues in (α/β)₈ Barrel Proteins Based on Hydrophobicity, Long-Range Interactions, and Sequence Conservation. *Proteins: Structure, Function and Genetics*, 55(2):316–329, 5 2004. ISSN 08873585. doi: 10.1002/prot.20052.
- [36] J. A. Silverman, R. Balakrishnan, and P. B. Harbury. Reverse engineering the (β/α)₈ barrel fold. *Proceedings of the National Academy of Sciences of the United States of America*, 98(6):3092–3097, 3 2001. ISSN 00278424. doi: 10.1073/pnas.041613598.
- [37] Markus Wiederstein and Manfred J. Sippl. Protein sequence randomization: Efficient estimation of protein stability using knowledge-based potentials. *Journal of Molecular Biology*, 345(5):1199–1212, 2 2005. ISSN 00222836. doi: 10.1016/j.jmb.2004.11.012.
- [38] Karine Goraj, André Renard, and Joseph A. Martial. Synthesis, purification and initial structural characterization of octarellin, a de novo polypeptide modelled on the α/β -barrel proteins. *Protein Engineering, Design and Selection*, 3(4):259–266, 1990. ISSN 17410126. doi: 10.1093/protein/3.4.259.
- [39] Marc Beauregard, Karine Goraj, Vincent Goffin, Karel Heremans, Eric Goormaghtigh, Jean Marie Ruyschaert, and Joseph A. Martial. Spectroscopic investigation of structure in octarellin (a de novo protein designed to adopt the α/β -barred packing). *Protein Engineering, Design and Selection*, 4(7):745–749, 10 1991. ISSN 17410126. doi: 10.1093/protein/4.7.745.
- [40] Toshiki Tanaka, Hiromi Kimura, Mayumi Hayashi, Yosinori Fujiyoshi, Ken-Ichi Ken-Ichi Fukuhara, and Haruki Nakamura. Characteristics of a de novo designed protein. *Protein Science*, 3(3):419–427, 3 1994. ISSN 1469896X. doi: 10.1002/pro.5560030306.
- [41] Toshiki Tanaka, Mayumi Hayashi, Hiromi Kimura, Motohisa Oobatake, and Haruki Nakamura. De novo design and creation of a stable artificial protein. *Biophysical Chemistry*, 50(1-2):47–61, 5 1994. ISSN 03014622. doi: 10.1016/0301-4622(94)85019-4.
- [42] Toshiki Tanaka, Yutaka Kuroda, Hiromi Kimura, Shun ichi Kidokoro, and Haruki Nakamura. Cooperative deformation of a de novo designed protein. *Protein Engineering Design and Selection*, 7(8):969–976, 8 1994. ISSN 1469896X. doi: 10.1002/pro.5560030306.
- [43] Annick Houbrechts, Benoît Moreau, Ruben Abagyan, Véronique Mainfroid, Gisèle Préaux, Alain Lamproye, Alain Poncin, Erik Goormaghtigh, Jean Marie Ruyschaert, Joseph A. Martial, and Karine Goraj. Second-generation octarellins: Two new de novo (β/α)₈ polypeptides designed for investigating the influence of β -residue packing on the α/β -barrel structure stability. *Protein Engineering, Design and Selection*, 8(3):249–259, 1995. ISSN 17410126. doi: 10.1093/protein/8.3.249.
- [44] F. Offredi, F. Dubail, P. Kischel, K. Sarinski, A. S. Stern, C. Van de Weerd, J. C. Hoch, C. Prosperi, J. M. François, S. L. Mayo, and J. A. Martial. De novo backbone and sequence design of an idealized α/β -barrel protein: Evidence of stable tertiary structure. *Journal of Molecular Biology*, 325(1):163–174, 1 2003. ISSN 00222836. doi: 10.1016/S0022-2836(02)01206-8.
- [45] Maximiliano Figueroa, Nicolas Oliveira, Annabelle Lejeune, Kristian W. Kaufmann, Brent M. Dorr, André Matagne, Joseph A. Martial, Jens Meiler, and Cécile Van de Weerd. Octarellin VI: using rosetta to design a putative artificial (β/α)₈ protein. *PloS One*, 8(8):e71858, 8 2013. ISSN 19326203. doi: 10.1371/journal.pone.0071858.
- [46] Maximiliano Figueroa, Mike Sleutel, Marylene Vandevenne, Gregory Parvizi, Sophie Attout, Olivier Jacquin, Julie Vandenameele, Axel W. Fischer, Christian Damblon, Erik Goormaghtigh, Marie Valerio-Lepiniec, Agathe Urvoas, Dominique Durand, Els Pardon, Jan Steyaert, Philippe Minard, Dominique Maes, Jens Meiler, André Matagne, Joseph A. Martial, and Cécile Van de Weerd. The unexpected structure of the designed protein Octarellin V.1 forms a challenge for protein structure prediction tools. *Journal of Structural Biology*, 195(1):19–30, 2016. ISSN 10958657. doi: 10.1016/j.jsb.2016.05.004.
- [47] Deepesh Nagarajan, Geeta Deka, and Megha Rao. Design of symmetric TIM barrel proteins from first principles. *BMC Biochemistry*, 16(1):1–22, 8 2015. ISSN 14712091. doi: 10.1186/s12858-015-0047-4.
- [48] Po-Ssu Ssu Huang, Kaspar Feldmeier, Fabio Parmeggiani, D. Alejandro Fernandez Velasco Fernandez, Birte Höcker, David Baker, D Alejandro Fernandez Velasco, Birte Höcker, and David Baker. De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nature Chemical Biology*, 12(1):29–34, 2015. ISSN 1552-4450. doi: 10.1038/nchembio.1966.
- [49] Nobuyasu Koga, Rie Tatsumi-Koga, Gaohua Liu, Rong Xiao, Thomas B. Acton, Gaetano T. Montelione, and David Baker. Principles for designing ideal protein structures. *Nature*, 491(7423):222–227, 2012. ISSN 00280836. doi: 10.1038/nature11600.
- [50] Derek N. Woolfson. Coiled-coil design: Updated and upgraded. *Sub-Cellular Biochemistry*, 82:35–61, 1 2017. ISSN 03060225. doi: 10.1007/978-3-319-49674-0{_}2.

- [51] Jinfeng Liu and Burkhard Rost. Comparing function and structure between entire proteomes. *Protein Science*, 10(10):1970–1979, 10 2001. ISSN 09618368. doi: 10.1110/ps.10101.
- [52] Linus Pauling and Robert B. Corey. Compound Helical Configurations of Polypeptide Chains: Structure of Proteins of the α -Keratin Type. *Nature*, 171(4341):59–61, 1953. ISSN 00280836. doi: 10.1038/171059a0.
- [53] F. H.C. Crick. Is α -keratin a coiled coil? *Nature*, 170(4334):882–883, 1952. ISSN 00280836. doi: 10.1038/170882b0.
- [54] F. H. C. Crick. The packing of α -helices: simple coiled-coils. *Acta Crystallographica*, 6(8):689–697, 9 1953. ISSN 0365110X. doi: 10.1107/s0365110x53001964.
- [55] I. A. Wilson, J. J. Skehel, and D. C. Wiley. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. *Nature*, 289(5796):366–373, 1981. ISSN 00280836. doi: 10.1038/289366a0.
- [56] Erin K. O’Shea, Juli D. Klemm, Peter S. Kim, and Tom Alber. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science*, 254(5031):539–544, 10 1991. ISSN 00368075. doi: 10.1126/science.1948029.
- [57] Andrei N. Lupas and Jens Bassler. Coiled Coils – A Model System for the 21st Century. *Trends in Biochemical Sciences*, 42(2):130–140, 2017. ISSN 13624326. doi: 10.1016/j.tibs.2016.10.007.
- [58] Liam C. Welsh, Martyn F. Symmons, and D. A. Marvin. The molecular structure and structural transition of the α -helical capsid in filamentous bacteriophage pfl. *Acta Crystallographica Section D: Biological Crystallography*, 56(2):137–150, 2 2000. ISSN 09074449. doi: 10.1107/S0907444999015334.
- [59] Efrosini Moutevelis and Derek N. Woolfson. A Periodic Table of Coiled-Coil Protein Structures. *Journal of Molecular Biology*, 385(3):726–732, 1 2009. ISSN 00222836. doi: 10.1016/j.jmb.2008.11.028.
- [60] Andrei N. Lupas. Coiled coils: new structures and new functions. *Trends in Biochemical Sciences*, 21(10):375–382, 10 1996. ISSN 09680004. doi: 10.1016/s0968-0004(96)10052-9.
- [61] Jody M. Mason and Katja M. Arndt. Coiled coil domains: Stability, specificity, and biological implications. *ChemBioChem*, 5(2):170–176, 2 2004. ISSN 14394227. doi: 10.1002/cbic.200300781.
- [62] Linda Truebestein and Thomas A. Leonard. Coiled-coils: The long and short of it. *BioEssays*, 38(9):903–916, 9 2016. ISSN 15211878. doi: 10.1002/bies.201600062.
- [63] William H. Landschulz, Peter F. Johnson, and Steven L. McKnight. The leucine zipper: A hypothetical structure common to a new class of DNA binding proteins. *Science*, 240(4860):1759–1764, 1988. ISSN 00368075. doi: 10.1126/science.3289117.
- [64] Glenn E. White and Harold P. Erickson. Sequence divergence of coiled coils-structural rods, myosin filament packing, and the extraordinary conservation of cohesins. *Journal of Structural Biology*, 154(2):111–121, 5 2006. ISSN 10478477. doi: 10.1016/j.jsb.2006.01.001.
- [65] Andrew P. Carter, Aristides G. Diamant, and Linas Urnavicius. How dynein and dynactin transport cargos: A structural perspective. *Current Opinion in Structural Biology*, 37:62–70, 4 2016. ISSN 1879033X. doi: 10.1016/j.sbi.2015.12.003.
- [66] Gregor Hagelueken, Bradley R. Clarke, Hexian Huang, Anne Tuukkanen, Iulia Danciu, Dmitri I. Svergun, Rohanah Hussain, Huanting Liu, Chris Whitfield, and James H. Naismith. A coiled-coil domain acts as a molecular ruler to regulate O-antigen chain length in lipopolysaccharide. *Nature Structural and Molecular Biology*, 22(1):50–56, 1 2015. ISSN 15459985. doi: 10.1038/nsmb.2935.
- [67] Gang Dong, Martina Medkova, Peter Novick, and Karin M. Reinisch. A Catalytic Coiled Coil: Structural Insights into the Activation of the Rab GTPase Sec4p by Sec2p. *Molecular Cell*, 25(3):455–462, 2 2007. ISSN 10972765. doi: 10.1016/j.molcel.2007.01.013.
- [68] P. Harbury, T Zhang, P. Kim, T Alber, J. Engel, D. Baker, and Derek N. Woolfson. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science*, 262(5138):1401–1407, 11 1993. ISSN 0036-8075. doi: 10.1126/science.8248779.
- [69] Jordan M. Fletcher, Aimee L. Boyle, Marc Bruning, SGail J. Bartlett, Thomas L. Vincent, Nathan R. Zaccari, Craig T. Armstrong, Elizabeth H.C. Bromley, Paula J. Booth, R. Leo Brady, Andrew R. Thomson, and Derek N. Woolfson. A basis set of de novo coiled-Coil peptide oligomers for rational protein design and synthetic biology. *ACS Synthetic Biology*, 1(6):240–250, 6 2012. ISSN 21615063. doi: 10.1021/sb300028q.
- [70] Peter Burkhard, Sergei Ivaninskii, and Ariel Lustig. Improving coiled-coil stability by optimizing ionic interactions. *Journal of Molecular Biology*, 318(3):901–910, 5 2002. ISSN 00222836. doi: 10.1016/S0022-2836(02)00114-6.
- [71] Helena Gradišar and Roman Jerala. De novo design of orthogonal peptide pairs forming parallel coiled-coil heterodimers. *Journal of Peptide Science*, 17(2):100–106, 2 2011. ISSN 10752617. doi: 10.1002/psc.1331.

- [72] Asha Acharya, Sergei B Ruvinov, Jozsef Gal, Jonathan R Moll, and Charles Vinson. A Heterodimerizing Leucine Zipper Coiled Coil System for Examining the Specificity of a Position Interactions : Amino Acids I , V , L , N , A , and K. *Biochemistry*, 41:14122–14131, 2002. doi: 10.1021/bi020486r.
- [73] Igor Drobnak, Helena Gradišar, Ajasja Ljubetič, Estera Merljak, and Roman Jerala. Modulation of Coiled-Coil Dimer Stability through Surface Residues while Preserving Pairing Specificity. *Journal of the American Chemical Society*, 139(24):8229–8236, 6 2017. ISSN 15205126. doi: 10.1021/jacs.7b01690.
- [74] Erik B Hadley, Oliver D Testa, Derek N Woolfson, and Samuel H Gellman. Preferred side-chain constellations at antiparallel coiled-coil interfaces. *Proceedings of the National Academy of Sciences of the United States of America*, 105(2):530–5, 2008. doi: 10.1073/pnas.0709068105.
- [75] Oscar D Monera, Nian E Zhou, Cyril M Kay, and Robert S Hodgess. Comparison of Antiparallel and Parallel Two-stranded α -Helical Coiled-coils. *The Journal of biological chemistry*, 268(26):19218–19227, 1993. ISSN 00219258. doi: 10.1016/S0021-9258(19)36502-0.
- [76] Oscar D. Monera, Cyril M. Kay, and Robert S. Hodges. Electrostatic Interactions Control the Parallel and Antiparallel Orientation of α -Helical Chains in Two-Stranded α -Helical Coiled-Coils. *Biochemistry*, 33(13):3862–3871, 4 1994. ISSN 15204995. doi: 10.1021/bi00179a010.
- [77] David G. Myszka and Irwin M Chaiken. Design and Characterization of an Intramolecular Antiparallel Coiled-Coil Peptide. *Biochemistry*, 33(9):2363–2372, 1994. ISSN 0006-2960. doi: 10.1021/bi00175a003.
- [78] Martha G. Oakley and Peter S. Kim. A buried polar interaction can direct the relative orientation of helices in a coiled coil. *Biochemistry*, 37(36):12603–12610, 9 1998. ISSN 00062960. doi: 10.1021/bi981269m.
- [79] D. L. McClain, H. L. Woods, and M. G. Oakley. Design and characterization of a heterodimeric coiled coil that forms exclusively with an antiparallel relative helix orientation. *Journal of the American Chemical Society*, 123(13):3151–3152, 2001. ISSN 00027863. doi: 10.1021/ja004099l.
- [80] Daniel G. Gurnon, Jennifer A. Whitaker, and Martha G. Oakley. Design and characterization of a homodimeric antiparallel coiled coil. *Journal of the American Chemical Society*, 125(25):7518–7519, 2003. ISSN 00027863. doi: 10.1021/ja0357590.
- [81] Christopher Negron and Amy E. Keating. A set of computationally designed orthogonal antiparallel homodimers that expands the synthetic coiled-coil toolkit. *Journal of the American Chemical Society*, 136(47):16544–16556, 11 2014. ISSN 15205126. doi: 10.1021/ja507847t.
- [82] M. G. Oakley and J. J. Hollenbeck. The design of antiparallel coiled coils. *Current Opinion in Structural Biology*, 11(4):450–457, 2001. ISSN 0959440X. doi: 10.1016/S0959-440X(00)00232-3.
- [83] By Derek N Woolfson. The Design Of Coiled-Coil Structures And Assemblies. *Advances in Protein Chemistry*, 70(04):79–112, 2005. doi: 10.1016/S0065-3233(04)70004-2.
- [84] Po-ssu Huang, Gustav Oberdorfer, Chunfu Xu, Xue Y Pei, and Brent L Nannenga. High thermodynamic stability of parametrically designed helical bundles. *Science*, 346(6208):481–486, 2014. doi: 10.1126/science.1253596.29.
- [85] Derek N Woolfson, Gail J Bartlett, Marc Bruning, and Andrew R Thomson. New currency for old rope: from coiled-coil assemblies to α -helical barrels. *Current Opinion in Structural Biology*, 22(4):432–441, 8 2012. ISSN 0959-440X. doi: 10.1016/J.SBI.2012.03.002.
- [86] Robert Lizatović, Oskar Aurelius, Olof Stenström, Torbjörn Drakenberg, Mikael Akke, Derek T. Logan, and Ingemar André. A De Novo Designed Coiled-Coil Peptide with a Reversible pH-Induced Oligomerization Switch. *Structure*, 24(6):946–955, 6 2016. ISSN 18784186. doi: 10.1016/j.str.2016.03.027.
- [87] Carolin Mueller and Tom N. Grossmann. Coiled-Coil Peptide Beacon: A Tunable Conformational Switch for Protein Detection. *Angewandte Chemie - International Edition*, 57(52):17079–17083, 12 2018. ISSN 15213773. doi: 10.1002/anie.201811515.
- [88] Jana Aupič, Fabio Lapenta, and Roman Jerala. SwitCCCh: Metal-Site Design for Controlling the Assembly of a Coiled-Coil Homodimer. *ChemBioChem*, 19(23):2453–2457, 12 2018. ISSN 14397633. doi: 10.1002/cbic.201800578.
- [89] Justin M. Torner and Paramjit S. Arora. Conformational control in a photoswitchable coiled coil. *Chemical Communications*, 57(12):1442–1445, 2 2021. ISSN 1364548X. doi: 10.1039/d0cc08318f.
- [90] Hao Shen, Jorge A. Fallas, Eric Lynch, William Sheffler, Bradley Parry, Nicholas Jannetty, Justin Decarreau, Michael Wagenbach, Juan Jesus Vicente, Jiajun Chen, Lei Wang, Quinton Dowling, Gustav Oberdorfer, Lance Stewart, Linda Wordeman, James De Yoreo, Christine Jacobs-Wagner, Justin Kollman, and David Baker. De novo design of self-assembling helical protein filaments. *Science (New York, N.Y.)*, 362(6415):705–709, 11 2018. ISSN 1095-9203. doi: 10.1126/science.aau3775.
- [91] Ajasja Ljubetič, Fabio Lapenta, Helena Gradišar, Igor Drobnak, Jana Aupič, Žiga Strmšek, Duško Lainšček, Iva Hafner-Bratkovič, Andreja Majerle, Nuša Krivec, Mojca Benčina, Tomaž Pisanski, Tanja Čirkovič

- Veličković, Adam Round, José María Carazo, Roberto Melero, and Roman Jerala. Design of coiled-coil protein-origami cages that self-assemble in vitro and in vivo. *Nature Biotechnology*, 35(11):1094–1101, 10 2017. ISSN 15461696. doi: 10.1038/nbt.3994.
- [92] Fabio Lapenta, Jana Aupič, Žiga Strmšek, and Roman Jerala. Coiled coil protein origami: From modular design principles towards biotechnological applications. *Chemical Society Reviews*, 47(10):3530–3542, 5 2018. ISSN 14604744. doi: 10.1039/c7cs00822h.
- [93] Andreja Majerle, San Hadzi, Jana Aupič, Tadej Satler, Fabio Lapenta, Žiga Strmšek, Jurij Lah, Remy Loris, and Roman Jerala. A nanobody toolbox targeting dimeric coiled-coil modules for functionalization of designed protein origami structures. *Proceedings of the National Academy of Sciences of the United States of America*, 118(17), 4 2021. ISSN 10916490. doi: 10.1073/pnas.2021899118.
- [94] Johanna Utterström, Sajjad Naeimipour, Robert Selegård, and Daniel Aili. Coiled coil-based therapeutics and drug delivery systems. *Advanced Drug Delivery Reviews*, 170:26–43, 3 2021. ISSN 18728294. doi: 10.1016/j.addr.2020.12.012.
- [95] Lindsay K. Hill, Michael Meleties, Priya Katyal, Xuan Xie, Erika Delgado-Fukushima, Teeba Jihad, Che Fu Liu, Sean O’Neill, Raymond S. Tu, P. Douglas Renfrew, Richard Bonneau, Youssef Z. Wadghiri, and Jin Kim Montclare. Thermoresponsive Protein-Engineered Coiled-Coil Hydrogel for Sustained Small Molecule Release. *Biomacromolecules*, 20(9):3340–3351, 9 2019. ISSN 15264602. doi: 10.1021/acs.biomac.9b00107.
- [96] Jordan M. Fletcher, Robert L. Harniman, Frederick R.H. Barnes, Aimee L. Boyle, Andrew Collins, Judith Mantell, Thomas H. Sharp, Massimo Antognozzi, Paula J. Booth, Noah Linden, Mervyn J. Miles, Richard B. Sessions, Paul Verkade, and Derek N. Woolfson. Self-assembling cages from coiled-coil peptide modules. *Science*, 340(6132):595–599, 5 2013. ISSN 10959203. doi: 10.1126/science.1233936.
- [97] Ken A. Dill and Justin L. MacCallum. The Protein-Folding Problem, 50 Years On. *Science*, 338(6110):1042–1046, 8 2012. ISSN 0036-8075. doi: 10.1126/science.321.5890.784.
- [98] Cyrus Levinthal. How to fold graciously. *Mössbauer Spectroscopy in Biological Systems Proceedings*, 24(41):22–24, 1969.
- [99] Cyrus Levinthal. Are there pathways for protein folding? *Journal de Chimie Physique*, 65:44–45, 1968. ISSN 0021-7689. doi: 10.1051/jcp/1968650044.
- [100] D. B. Wetlaufer. Nucleation, rapid folding, and globular intrachain regions in proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 70(3):697–701, 1973. ISSN 00278424. doi: 10.1073/pnas.70.3.697.
- [101] O. B. Ptitsyn and A. A. Rashin. A model of myoglobin self-organization. *Biophysical Chemistry*, 3(1):1–20, 2 1975. ISSN 03014622. doi: 10.1016/0301-4622(75)80033-0.
- [102] Martin Karplus and David L. Weaver. Diffusion–collision model for protein folding. *Biopolymers*, 18(6):1421–1437, 6 1979. ISSN 10970282. doi: 10.1002/bip.1979.360180608.
- [103] Ken A. Dill. Theory for the Folding and Stability of Globular Proteins. *Biochemistry*, 24(6):1501–1509, 3 1985. ISSN 15204995. doi: 10.1021/bi00327a032.
- [104] Christian B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181(4096):223–230, 7 1973. ISSN 00368075. doi: 10.1126/science.181.4096.223.
- [105] P. E. Leopold, M. Montal, and J. N. Onuchic. Protein folding funnels: A kinetic approach to the sequence-structure relationship. *Proceedings of the National Academy of Sciences of the United States of America*, 89(18):8721–8725, 9 1992. ISSN 00278424. doi: 10.1073/pnas.89.18.8721.
- [106] José Nelson Onuchic, Zaida Luthey-Schulten, and Peter G. Wolynes. Theory of Protein Folding: The Energy Landscape Perspective. *Annual Review of Physical Chemistry*, 48(1):545–600, 1997. ISSN 0066426X. doi: 10.1146/annurev.physchem.48.1.545.
- [107] M. Bajaj and T. Blundell. Evolution and the tertiary structure of proteins. *Annual review of biophysics and bioengineering*, 13:453–492, 11 1984. ISSN 00846589. doi: 10.1146/annurev.bb.13.060184.002321.
- [108] Brian Kuhlman and Philip Bradley. Advances in protein structure prediction and design. *Nature Reviews Molecular Cell Biology*, 20(11):681–697, 11 2019. ISSN 14710080. doi: 10.1038/s41580-019-0163-x.
- [109] Kim T. Simons, Charles Kooperberg, Enoch Huang, and David Baker. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *Journal of Molecular Biology*, 268(1):209–225, 4 1997. ISSN 00222836. doi: 10.1006/jmbi.1997.0959.
- [110] Christopher Bystroff, Kim T. Simonst, Karen F Han, and David Baker. Local sequence-structure correlations in proteins. *Current Opinion in Biotechnology*, 7:417–421, 1996.
- [111] Carol A. Rohl, Charlie E.M. M Strauss, Kira M.S. S Misura, and David Baker. Protein Structure Prediction Using Rosetta. *Methods in Enzymology*, 383(2003):66–93, 1 2004. ISSN 00766879. doi: 10.1016/S0076-6879(04)83004-0.

- [112] Kristian W Kaufmann, Gordon H Lemmon, Samuel L Deluca, Jonathan H Sheehan, and Jens Meiler. Practically Useful : What the R OSETTA Protein Modeling Suite Can Do for You. *Biochemistry*, 49:2987–2998, 2010. doi: 10.1021/bi902153g.
- [113] Rebecca F Alford, Andrew Leaver-Fay, Jeliasko R. Jeliaskov, Matthew J. O’Meara, Frank P. DiMaio, Hahnbeom Park, Maxim V. Shapovalov, P. Douglas Renfrew, Vikram K. Mulligan, Kalli Kappel, Jason W. Labonte, Michael S. Pacella, Richard Bonneau, Philip Bradley, Roland L. Dunbrack, Rhiju Das, David Baker, Brian Kuhlman, Tanja Kortemme, Jeffrey J. Gray, Matthew J. O’Meara, Frank P. DiMaio, Hahnbeom Park, Maxim V. Shapovalov, P. Douglas Renfrew, Vikram K. Mulligan, Kalli Kappel, Jason W. Labonte, Michael S. Pacella, Richard Bonneau, Philip Bradley, Roland L. Dunbrack, Rhiju Das, David Baker, Brian Kuhlman, Tanja Kortemme, and Jeffrey J. Gray. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *Journal of Chemical Theory and Computation*, 13(6):3031–3048, 6 2017. ISSN 15499626. doi: 10.1021/acs.jctc.7b00125.
- [114] John Moulton, Jan T. Pedersen, Richard Judson, and Krzysztof Fidelis. A large-scale experiment to assess protein structure prediction methods. *Proteins: Structure, Function, and Bioinformatics*, 23(3):ii–iv, 1995. ISSN 10970134. doi: 10.1002/prot.340230303.
- [115] Sergey Ovchinnikov, David E. Kim, Ray Yu Rwei Wang, Yuan Liu, Frank Dimaio, and David Baker. Improved de novo structure prediction in CASP11 by incorporating coevolution information into Rosetta. *Proteins: Structure, Function and Bioinformatics*, 84(1):67–75, 9 2016. ISSN 10970134. doi: 10.1002/prot.24974.
- [116] Andriy Kryshchak, Torsten Schwede, Maya Topf, Krzysztof Fidelis, and John Moulton. Critical assessment of methods of protein structure prediction (CASP)—Round XIII, 12 2019. ISSN 10970134.
- [117] Joerg Schaarschmidt, Bohdan Monastyrskyy, Andriy Kryshchak, and Alexandre M.J.J. Bonvin. Assessment of contact predictions in CASP12: Co-evolution and deep learning coming of age. *Proteins: Structure, Function and Bioinformatics*, 86:51–66, 3 2018. ISSN 10970134. doi: 10.1002/prot.25407.
- [118] Andrew W. Senior, Richard Evans, John Jumper, James Kirkpatrick, Laurent Sifre, Tim Green, Chongli Qin, Augustin Židek, Alexander W.R. Nelson, Alex Bridgland, Hugo Penedones, Stig Petersen, Karen Simonyan, Steve Crossan, Pushmeet Kohli, David T. Jones, David Silver, Koray Kavukcuoglu, and Demis Hassabis. Improved protein structure prediction using potentials from deep learning. *Nature*, 577(7792):706–710, 1 2020. ISSN 14764687. doi: 10.1038/s41586-019-1923-7.
- [119] Jianyi Yang, Ivan Anishchenko, Hahnbeom Park, Zhenling Peng, Sergey Ovchinnikov, and David Baker. Improved protein structure prediction using predicted interresidue orientations. *Proceedings of the National Academy of Sciences of the United States of America*, 117(3):1496–1503, 1 2020. ISSN 10916490. doi: 10.1073/pnas.1914677117.
- [120] Andriy Kryshchak, Torsten Schwede, Maya Topf, Krzysztof Fidelis, and John Moulton. Critical assessment of methods of protein structure prediction (CASP)—Round XIV. *Proteins: Structure, Function and Bioinformatics*, 12 2021. ISSN 10970134. doi: 10.1002/PROT.26237/FORMAT/PDF.
- [121] José M. Pereira, Maria Vieira, and Sérgio M. Santos. Step-by-step design of proteins for small molecule interaction: A review on recent milestones. *Protein Science*, 30(8):1502–1520, 8 2021. ISSN 1469-896X. doi: 10.1002/PRO.4098.
- [122] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Židek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A.A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 7 2021. ISSN 14764687. doi: 10.1038/s41586-021-03819-2.
- [123] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Židek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, David Silver, Oriol Vinyals, Andrew W Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Applying and improving AlphaFold at CASP14. *Proteins: Structure, Function, and Bioinformatics*, 89(12):1711–1721, 12 2021. ISSN 0887-3585. doi: 10.1002/prot.26257.
- [124] Ewen Callaway. ‘It will change everything’: DeepMind’s AI makes gigantic leap in solving protein structures, 12 2020. ISSN 14764687.
- [125] DeepMind and EMBL-EBI. AlphaFold Protein Structure Database, 2021.

- [126] Po Ssu Huang, Yih En Andrew Ban, Florian Richter, Ingemar Andre, Robert Vernon, William R. Schief, and David Baker. RosettaRemodel: A generalized framework for flexible backbone protein design. *PLoS ONE*, 6 (8), 2011. ISSN 19326203. doi: 10.1371/journal.pone.0024109.
- [127] Po-Ssu Huang, Scott E. Boyken, and David Baker. The coming of age of de novo protein design. *Nature*, 537(7620):320–327, 2016. ISSN 0028-0836. doi: 10.1038/nature19946.
- [128] Yu Ru Lin, Nobuyasu Koga, Rie Tatsumi-Koga, Gaohua Liu, Amanda F Clouser, Gaetano T Montelione, David Baker, and William F. DeGrado. Control over overall shape and size in de novo designed proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 112(40):E5478–E5485, 10 2015. ISSN 10916490. doi: 10.1073/pnas.1509508112.
- [129] Robin Pearce and Yang Zhang. Deep learning techniques have significantly impacted protein structure prediction and protein design. *Current Opinion in Structural Biology*, 68:194–207, 6 2021. ISSN 0959-440X. doi: 10.1016/J.SBI.2021.01.007.
- [130] David Baker. What has de novo protein design taught us about protein folding and biophysics? *Protein Science*, 28(4):678–683, 4 2019. ISSN 1469-896X. doi: 10.1002/PRO.3588.
- [131] Xingjie Pan, Michael C. Thompson, Yang Zhang, Lin Liu, James S. Fraser, Mark J. S. Kelly, and Tanja Kortemme. Expanding the space of protein geometries by computational design of de novo fold families. *Science*, 369(6507):1132–1136, 8 2020. ISSN 0036-8075. doi: 10.1126/science.abc0881.
- [132] Adi Goldenzweig and Sarel J. Fleishman. Principles of Protein Stability and Their Application in Computational Design. *Annual Review of Biochemistry*, 87:105–129, 6 2018. doi: 10.1146/ANNUREV-BIOCHEM-062917-012102.
- [133] Jiayi Dou, Anastassia A. Vorobieva, William Sheffler, Lindsey A. Doyle, Hahnbeom Park, Matthew J. Bick, Binchen Mao, Glenna W. Foight, Min Yen Lee, Lauren A. Gagnon, Lauren Carter, Banumathi Sankaran, Sergey Ovchinnikov, Enrique Marcos, Po Ssu Huang, Joshua C. Vaughan, Barry L. Stoddard, and David Baker. De novo design of a fluorescence-activating β -barrel. *Nature*, 561(7724):485–491, 9 2018. ISSN 14764687. doi: 10.1038/s41586-018-0509-0.
- [134] Alfredo Quijano-Rubio, Hsien Wei Yeh, Jooyoung Park, Hansol Lee, Robert A. Langan, Scott E. Boyken, Marc J. Lajoie, Longxing Cao, Cameron M. Chow, Marcos C. Miranda, Jimin Wi, Hyo Jeong Hong, Lance Stewart, Byung Ha Oh, and David Baker. De novo design of modular and tunable protein biosensors. *Nature*, 591(7850):482–487, 1 2021. ISSN 14764687. doi: 10.1038/s41586-021-03258-z.
- [135] Sailan Shui, Pablo Gainza, Leo Scheller, Che Yang, Yoichi Kurumida, Stéphane Rosset, Sandrine Georgeon, Raphaël B. Di Roberto, Rocío Castellanos-Rueda, Sai T. Reddy, and Bruno E. Correia. A rational blueprint for the design of chemically-controlled protein switches. *Nature Communications*, 12(1):1–12, 10 2021. ISSN 20411723. doi: 10.1038/s41467-021-25735-9.
- [136] Kathy Y. Wei, Danai Moschidi, Matthew J. Bick, Santrupti Nerli, Andrew C. McShan, Lauren P. Carter, Po Ssu Huang, Daniel A. Fletcher, Nikolaos G. Sgourakis, Scott E. Boyken, and David Baker. Computational design of a protein family that adopts two well-defined and structurally divergent de novo folds. *bioRxiv*, page 597161, 4 2019. ISSN 2692-8205. doi: 10.1101/597161.
- [137] Valerie Vaissier Welborn and Teresa Head-Gordon. Computational Design of Synthetic Enzymes. *Chemical Reviews*, 119(11):6613–6630, 6 2019. ISSN 15206890. doi: 10.1021/acs.chemrev.8b00399.
- [138] Justin B. Siegel, Alexandre Zanghellini, Helena M. Lovick, Gert Kiss, Abigail R. Lambert, Jennifer L. St.Clair, Jasmine L. Gallaher, Donald Hilvert, Michael H. Gelb, Barry L. Stoddard, Kendall N. Houk, Forrest E. Michael, and David Baker. Computational design of an enzyme catalyst for a stereoselective bimolecular diels-alder reaction. *Science*, 329(5989):309–313, 7 2010. ISSN 00368075. doi: 10.1126/science.1190239.
- [139] Ann E. Donnelly, Grant S. Murphy, Katherine M. Digianantonio, and Michael H. Hecht. A de novo enzyme catalyzes a life-sustaining reaction in *Escherichia coli*. *Nature Chemical Biology*, 14(3):253–255, 1 2018. ISSN 15524469. doi: 10.1038/nchembio.2550.
- [140] Daniel W. Watkins, Jonathan M.X. Jenkins, Katie J. Grayson, Nicola Wood, Jack W. Steventon, Kristian K. Le Vay, Matthew I. Goodwin, Anna S. Mullen, Henry J. Bailey, Matthew P. Crump, Fraser MacMillan, Adrian J. Mulholland, Gus Cameron, Richard B. Sessions, Stephen Mann, and J. L. Ross Anderson. Construction and in vivo assembly of a catalytically proficient and hyperthermostable de novo enzyme. *Nature Communications*, 8(1):1–9, 8 2017. ISSN 20411723. doi: 10.1038/s41467-017-00541-4.
- [141] Zibo Chen, Ryan D. Kibler, Andrew Hunt, Florian Busch, Jocelynn Pearl, Mengxuan Jia, Zachary L. VanAerenum, Basile I.M. M Wicky, Galen Dods, Hanna Liao, Matthew S. Wilken, Christie Ciarlo, Shon Green, Hana El-samad, John Stamatoyannopoulos, Vicki H. Wysocki, Michael C. Jewett, Scott E. Boyken, and David Baker. De novo design of protein logic gates. *Science*, 84(April):78–84, 4 2020. ISSN 10959203. doi: 10.1126/science.aay2790.

- [142] Jiabing Ma, Yicheng Mo, Menglin Tang, Junjie Shen, Yanan Qi, Wenxu Zhao, Yi Huang, Yanmin Xu, and Cheng Qian. Bispecific Antibodies: From Research to Clinical Application. *Frontiers in Immunology*, 12: 1555, 5 2021. ISSN 16643224. doi: 10.3389/fimmu.2021.626616.
- [143] Longxing Cao, Inna Goreschnik, Brian Coventry, James Brett Case, Lauren Miller, Lisa Kozodoy, Rita E. Chen, Lauren Carter, Alexandra C. Walls, Young-Jun Jun Park, Eva-Maria Maria Strauch, Lance Stewart, Michael S. Diamond, David Veesler, and David Baker. De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. *Science*, 370(6515), 10 2020. ISSN 10959203. doi: 10.1126/science.abd9909.
- [144] James O'Connell, Zhixiu Li, Jack Hanson, Rhys Heffernan, James Lyons, Kuldip Paliwal, Abdollah Dehzangi, Yuedong Yang, and Yaoqi Zhou. SPIN2: Predicting sequence profiles from protein structures using deep neural networks. *Proteins: Structure, Function and Bioinformatics*, 86(6):629–633, 6 2018. ISSN 10970134. doi: 10.1002/prot.25489.
- [145] C. Nick Pace, Bret A. Shirley, Marsha McNutt, and Ketan Gajiwala. Forces contributing to the conformational stability of proteins. *The FASEB Journal*, 10(1):75–83, 1 1996. ISSN 0892-6638. doi: 10.1096/fasebj.10.1.8566551.
- [146] Mohammad M. Islam, Kei Kobayashi, Shun Ichi Kidokoro, and Yutaka Kuroda. Hydrophobic surface residues can stabilize a protein through improved water–protein interactions. *FEBS Journal*, 286(20):4122–4134, 10 2019. ISSN 17424658. doi: 10.1111/febs.14941.
- [147] Michael H. Hecht, Julian M. Sturtevant, and Robert T. Sauer. Stabilization of λ repressor against thermal denaturation by site-directed Gly \rightarrow Ala changes in α -helix 3. *Proteins: Structure, Function, and Bioinformatics*, 1(1):43–46, 1986. ISSN 10970134. doi: 10.1002/prot.340010108.
- [148] Immaculada Margarit, Susanna Camagnoli, Francesco Frigerio, Guido Grandi, Vincenzo De Filippis, and Angelo Fontana. Cumulative stabilizing effects of glycine to alanine substitutions in *Bacillus subtilis* neutral protease. *Protein Engineering, Design and Selection*, 5(6):543–550, 9 1992. ISSN 17410126. doi: 10.1093/protein/5.6.543.
- [149] Yuxing Li, Peter J. Reilly, and Clark Ford. Effect of introducing proline residues on the stability of *Aspergillus awamori*. *Protein Engineering*, 10(10):1199–1204, 1997. ISSN 02692139. doi: 10.1093/protein/10.10.1199.
- [150] Ravindra Singh Prajapati, Mili Das, Sridhar Sreeramulu, Minhajuddin Sirajuddin, Sankaranarayanan Srinivasan, Vaishnavi Krishnamurthy, Ranganathan Ranjani, C. Ramakrishnan, and Raghavan Varadarajan. Thermodynamic effects of proline introduction on protein stability. *Proteins: Structure, Function and Genetics*, 66(2):480–491, 2 2007. ISSN 08873585. doi: 10.1002/prot.21215.
- [151] Maria Zavodszky, Chao-Wei Chen, Jenq-Kuen Huang, Michal Zolkiewski, Lisa Wen, and Ramaswamy Krishnamoorthi. Disulfide bond effects on protein stability: Designed variants of *Cucurbita maxima* trypsin inhibitor-V. *Protein Science*, 10(1):149–160, 1 2001. ISSN 09618368. doi: 10.1110/ps.26801.
- [152] Bogdan S. Melnik, Tatiana V. Povarnitsyna, Anatoly S. Glukhov, Tatyana N. Melnik, and Vladimir N. Uversky. Ss-stabilizing proteins rationally: Intrinsic disorder-based design of stabilizing disulphide bridges in GFP. *Journal of Biomolecular Structure and Dynamics*, 29(4):815–824, 2012. ISSN 15380254. doi: 10.1080/07391102.2012.10507414.
- [153] Stanley C. Kwok and Robert S. Hodges. Clustering of large hydrophobes in the hydrophobic core of two-stranded α -helical coiled-coils controls protein folding and stability. *Journal of Biological Chemistry*, 278(37): 35248–35254, 9 2003. ISSN 00219258. doi: 10.1074/jbc.M305306200.
- [154] Zsuzsanna Dosztányi, András Fiser, and István Simon. Stabilization centers in proteins: Identification, characterization and predictions. *Journal of Molecular Biology*, 272(4):597–612, 10 1997. ISSN 00222836. doi: 10.1006/jmbi.1997.1242.
- [155] Raquel Godoy-Ruiz, Raul Perez-Jimenez, Beatriz Ibarra-Molero, and Jose M. Sanchez-Ruiz. A stability pattern of protein hydrophobic mutations that reflects evolutionary structural optimization. *Biophysical Journal*, 89(5):3320–3331, 11 2005. ISSN 00063495. doi: 10.1529/biophysj.105.067025.
- [156] Richard Eudes, Le Tuan Khanh, Jean Delettré, Jean Paul Mornon, and Isabelle Callebaut. A generalized analysis of hydrophobic and loop clusters within globular protein sequences. *BMC Structural Biology*, 7(1): 1–22, 1 2007. ISSN 14726807. doi: 10.1186/1472-6807-7-2.
- [157] Sujit Basak, R. Paul Nobrega, Davide Tavella, Laura M. Deveau, Nobuyasu Koga, Rie Tatsumi-Koga, David Baker, Francesca Massi, C. Robert Matthews, R. Paul Nobrega, Davide Tavella, Laura M. Deveau, Nobuyasu Koga, Rie Tatsumi-Koga, David Baker, Francesca Massi, and C. Robert Matthews. Networks of electrostatic and hydrophobic interactions modulate the complex folding free energy surface of a designed $\beta\alpha$ protein. *Proceedings of the National Academy of Sciences of the United States of America*, 116(14):6806–6811, 4 2019. ISSN 10916490. doi: 10.1073/pnas.1818744116.
- [158] J. Arunachalam and N. Gautham. Hydrophobic clusters in protein structures. *Proteins: Structure, Function and Genetics*, 71(4):2012–2025, 6 2008. ISSN 08873585. doi: 10.1002/prot.21881.

- [159] S. Selvaraj and M. Michael Gromiha. Role of hydrophobic clusters and long-range contact networks in the folding of (α/β)₈ barrel proteins. *Biophysical Journal*, 84(3):1919–1925, 3 2003. ISSN 00063495. doi: 10.1016/S0006-3495(03)75000-0.
- [160] Ying Wu, Ramakrishna Vadrevu, Sagar Kathuria, Xiaoyan Yang, and C. Robert Matthews. A Tightly Packed Hydrophobic Cluster Directs the Formation of an Off-pathway Sub-millisecond Folding Intermediate in the α Subunit of Tryptophan Synthase, a TIM Barrel Protein. *Journal of Molecular Biology*, 366(5):1624–1638, 3 2007. ISSN 00222836. doi: 10.1016/j.jmb.2006.12.005.
- [161] Rohit Jain, Khaja Muneeruddin, Jeremy Anderson, Michael J. Harms, Scott A. Shaffer, and C. Robert Matthews. A conserved folding nucleus sculpts the free energy landscape of bacterial and archaeal orthologs from a divergent TIM barrel family. *Proceedings of the National Academy of Sciences of the United States of America*, 118(17):e2019571118, 4 2021. ISSN 10916490. doi: 10.1073/pnas.2019571118.
- [162] D. J. Barlow and J. M. Thornton. Ion-pairs in proteins. *Journal of Molecular Biology*, 168(4):867–885, 8 1983. ISSN 00222836. doi: 10.1016/S0022-2836(83)80079-5.
- [163] Angela Lombardi, Christopher M. Summa, Silvano Geremia, Lucio Randaccio, Vincenzo Pavone, and William F. DeGrado. Retrostructural analysis of metalloproteins: Application to the design of a minimal model for diiron proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 97(12):6298–6305, 6 2000. ISSN 00278424. doi: 10.1073/pnas.97.12.6298.
- [164] U. Bastolla and Lloyd Demetrius. Stability constraints and protein evolution: the role of chain length, composition and disulfide bonds. *Protein Engineering, Design and Selection*, 18(9):405–415, 9 2005. ISSN 1741-0126. doi: 10.1093/PROTEIN/GZI045.
- [165] Carey D. Waldburger, Joel F. Schildbach, and Robert T. Sauer. Are buried salt bridges important for protein stability and conformational specificity? *Nature Structural Biology 1995 2:2*, 2(2):122–128, 1995. ISSN 1545-9985. doi: 10.1038/nsb0295-122.
- [166] Jason E. Donald, Daniel W. Kulp, and William F. DeGrado. Salt bridges: Geometrically specific, designable interactions. *Proteins: Structure, Function and Bioinformatics*, 79(3):898–915, 3 2011. ISSN 08873585. doi: 10.1002/prot.22927.
- [167] Sandeep Kumar and Ruth Nussinov. Close-range electrostatic interactions in proteins. *ChemBioChem*, 3(7):604–617, 1 2002. ISSN 14394227. doi: 10.1002/1439-7633(20020703)3:7<604::AID-CBIC604>3.0.CO;2-X.
- [168] Sandeep Kumar and Ruth Nussinov. Salt bridge stability in monomeric proteins. *Journal of Molecular Biology*, 293(5):1241–1255, 11 1999. ISSN 00222836. doi: 10.1006/jmbi.1999.3218.
- [169] Hans Rudolf Bosshard, Daniel N. Marti, and Ilian Jelesarov. Protein stabilization by salt bridges: Concepts, experimental approaches and clarification of some misunderstandings. *Journal of Molecular Recognition*, 17(1):1–16, 1 2004. ISSN 09523499. doi: 10.1002/jmr.657.
- [170] Sankar Basu and Debasish Mukharjee. Salt-bridge networks within globular and disordered proteins: characterizing trends for designable interactions. *Journal of Molecular Modeling 2017 23:7*, 23(7):1–17, 6 2017. ISSN 0948-5023. doi: 10.1007/S00894-017-3376-Y.
- [171] John H. Missimer, Michel O. Steinmetz, Riccardo Baron, Fritz K. Winkler, Richard A. Kammerer, Xavier Daura, and Wilfred F. van Gunsteren. Configurational entropy elucidates the role of salt-bridge networks in protein thermostability. *Protein Science*, 16(7):1349–1359, 7 2007. ISSN 1469-896X. doi: 10.1110/PS.062542907.
- [172] Inari Kursula, Sanna Partanen, Anne-Marie Marie Lambeir, and Rik K. Wierenga. The importance of the conserved Arg191–Asp227 salt bridge of triosephosphate isomerase for folding, stability, and catalysis. *FEBS Letters*, 518(1-3):39–42, 5 2002. ISSN 0014-5793. doi: 10.1016/S0014-5793(02)02639-X.
- [173] Dennis R. Livesay and David La. The evolutionary origins and catalytic importance of conserved electrostatic networks within TIM-barrel proteins. *Protein Science*, 14(5):1158–1170, 5 2005. ISSN 09618368. doi: 10.1110/ps.041221105.
- [174] Radhika P. Nagarkar, Galit Fichman, and Joel P. Schneider. Engineering and characterization of a pH-sensitive homodimeric antiparallel coiled coil. *Peptide Science*, 112(5):e24180, 9 2020. ISSN 24758817. doi: 10.1002/pep2.24180.
- [175] Bagautdin Bagautdinov and Katsuhide Yutani. Structure of indole-3-glycerol phosphate synthase from *Thermus thermophilus* HB8: Implications for thermal stability. *Acta Crystallographica Section D: Biological Crystallography*, 67(12):1054–1064, 12 2011. ISSN 09074449. doi: 10.1107/S0907444911045264.
- [176] Rie Omi, Hiroyuki Mizuguchi, Masaru Goto, Ikuko Miyahara, Hideyuki Hayashi, Hiroyuki Kagamiyama, and Ken Hirotsu. Structure of imidazole glycerol phosphate synthase from *Thermus thermophilus* HB8:

- Open-closed conformational change and ammonia tunneling. *Journal of Biochemistry*, 132(5):759–765, 11 2002. ISSN 0021924X. doi: 10.1093/oxfordjournals.jbchem.a003284.
- [177] Birte Höcker, Jörg Claren, and Reinhard Sterner. Mimicking enzyme evolution by generating new ($\beta\alpha$) 8-barrels from ($\beta\alpha$)4-half-barrels. *Proceedings of the National Academy of Sciences of the United States of America*, 101(47):16448–16453, 11 2004. ISSN 00278424. doi: 10.1073/pnas.0405832101.
- [178] Tobias Seitz, Marco Bocola, Jörg Claren, and Reinhard Sterner. Stabilisation of a ($\beta\alpha$)8-Barrel Protein Designed from Identical Half Barrels. *Journal of Molecular Biology*, 372(1):114–129, 2007. ISSN 00222836. doi: 10.1016/j.jmb.2007.06.036.
- [179] Jooyoung Park, Brinda Selvaraj, Andrew C. McShan, Scott E. Boyken, Kathy Y. Wei, Gustav Oberdorfer, William Degrado, Nikolaos G. Sgourakis, Matthew J. Cuneo, Dean A.A. Myles, and David Baker. De novo design of a homo-trimeric amantadine-binding protein. *eLife*, 8, 12 2019. ISSN 2050084X. doi: 10.7554/eLife.47839.
- [180] Molly M. Sheehan, Michael S. Magaraci, Ivan A. Kuznetsov, Joshua A. Mancini, Goutham Kodali, Christopher C. Moser, P. Leslie Dutton, and Brian Y. Chow. Rational Construction of Compact de Novo-Designed Biliverdin-Binding Proteins. *Biochemistry*, 57(49):6752–6756, 12 2018. ISSN 15204995. doi: 10.1021/ACS.BIOCHEM.8B01076/SUPPL_{_}FILE/B18B01076_{_}SI_{_}001.PDF.
- [181] Tina Fink, Jan Lonžarić, Arne Praznik, Tjaša Plaper, Estera Merljak, Katja Leben, Nina Jerala, Tina Lebar, Žiga Strmšek, Fabio Lapenta, Mojca Benčina, and Roman Jerala. Design of fast proteolysis-based signaling and logic circuits in mammalian cells. *Nature Chemical Biology*, 15(2):115–122, 2 2019. ISSN 15524469. doi: 10.1038/s41589-018-0181-6.
- [182] Andrew H. Ng, Taylor H. Nguyen, Mariana Gómez-Schiavon, Galen Dods, Robert A. Langan, Scott E. Boyken, Jennifer A. Samson, Lucas M. Waldburger, John E. Dueber, David Baker, and Hana El-Samad. Modular and tunable biological feedback control using a de novo protein switch. *Nature*, 572(7768):265–269, 7 2019. ISSN 14764687. doi: 10.1038/s41586-019-1425-7.
- [183] Robert A. Langan, Scott E. Boyken, Andrew H. Ng, Jennifer A. Samson, Galen Dods, Alexandra M. Westbrook, Taylor H. Nguyen, Marc J. Lajoie, Zibo Chen, Stephanie Berger, Vikram Khipple Mulligan, John E. Dueber, Walter R. P. Novak, Hana El-Samad, and David Baker. De novo design of bioactive protein switches. *Nature*, page 1, 7 2019. ISSN 0028-0836. doi: 10.1038/s41586-019-1432-8.
- [184] Antony J. Burton, Andrew R. Thomson, William M. Dawson, R. Leo Brady, and Derek N. Woolfson. Installing hydrolytic activity into a completely de novo protein framework. *Nature Chemistry*, 8(9):837–844, 7 2016. ISSN 17554349. doi: 10.1038/nchem.2555.
- [185] Michael S. Wang and Michael H. Hecht. A Completely de Novo ATPase from Combinatorial Protein Design. *Journal of the American Chemical Society*, 142(36):15230–15234, 9 2020. ISSN 15205126. doi: 10.1021/jacs.0c02954.
- [186] Emily G. Baker, Gail J. Bartlett, Kathryn L. Porter Goff, and Derek N. Woolfson. Miniprotein Design: Past, Present, and Prospects. *Accounts of Chemical Research*, 50(9):2085–2092, 9 2017. ISSN 0001-4842. doi: 10.1021/acs.accounts.7b00186.
- [187] Aaron Chevalier, Daniel Adriano Silva, Gabriel J. Rocklin, Derrick R. Hicks, Renan Vergara, Patience Murapa, Steffen M. Bernard, Lu Zhang, Kwok Ho Lam, Guorui Yao, Christopher D. Bahl, Shin Ichiro Miyashita, Inna Goreschnik, James T. Fuller, Merika T. Koday, Cody M. Jenkins, Tom Colvin, Lauren Carter, Alan Bohn, Cassie M. Bryan, D. Alejandro Fernández-Velasco, Lance Stewart, Min Dong, Xuhui Huang, Rongsheng Jin, Ian A. Wilson, Deborah H. Fuller, and David Baker. Massively parallel de novo protein design for targeted therapeutics. *Nature*, 550(7674):74–79, 10 2017. ISSN 14764687. doi: 10.1038/nature23912.
- [188] Enrique Marcos, Tamuka M. Chidyausiku, Andrew C. McShan, Thomas Evangelidis, Santrupti Nerli, Lauren Carter, Lucas G. Nivón, Audrey Davis, Gustav Oberdorfer, Konstantinos Tripsianes, Nikolaos G. Sgourakis, and David Baker. De novo design of a non-local β -sheet protein with high stability and accuracy. *Nature Structural and Molecular Biology*, 25(11):1028–1034, 11 2018. ISSN 15459985. doi: 10.1038/s41594-018-0141-6.
- [189] Che Yang, Fabian Sesterhenn, Jaume Bonet, Eva A. van Aalen, Leo Scheller, Luciano A. Abriata, Johannes T. Cramer, Xiaolin Wen, Stéphane Rosset, Sandrine Georgeon, Theodore Jardetzky, Thomas Krey, Martin Fussenegger, Maarten Merckx, and Bruno E. Correia. Bottom-up de novo design of functional proteins with complex structural features. *Nature Chemical Biology*, 17(4):492–500, 1 2021. ISSN 15524469. doi: 10.1038/s41589-020-00699-x.
- [190] Tina Izard and Jurgen Sygusch. Induced fit movements and metal cofactor selectivity of class II aldolases: Structure of *Thermus aquaticus* fructose-1,6-bisphosphate aldolase. *Journal of Biological Chemistry*, 279(12):11825–11833, 3 2004. ISSN 00219258. doi: 10.1074/jbc.M311375200.

- [191] Jonas Gregor Wiese, Sooruban Shanmugaratnam, and Birte Höcker. Extension of a de novo TIM barrel with a rationally designed secondary structure element. *Protein Science*, 30(5):982–989, 5 2021. ISSN 1469896X. doi: 10.1002/pro.4064.
- [192] Shane J. Caldwell, Ian C. Haydon, Nikoletta Piperidou, Po Ssu Huang, Matthew J. Bick, H. Sebastian Sjöström, Donald Hilvert, David Baker, and Cathleen Zeymer. Tight and specific lanthanide binding in a de novo TIM barrel with a large internal cavity designed by symmetric domain fusion. *Proceedings of the National Academy of Sciences of the United States of America*, 117(48):30362–30369, 12 2020. ISSN 10916490. doi: 10.1073/pnas.2008535117.
- [193] Sina Kordes, Sergio Romero-Romero, Leonie Lutz, and Birte Höcker. A newly introduced salt bridge cluster improves structural and biophysical properties of de novo TIM barrels. *Protein Science*, 12 2021. ISSN 0961-8368. doi: 10.1002/pro.4249.
- [194] Noelia Ferruz, Francisco Lobos, Dominik Lemm, Saacnicteh Toledo-Patino, José Arcadio Farías-Rico, Steffen Schmidt, and Birte Höcker. Identification and Analysis of Natural Building Blocks for Evolution-Guided Fragment-Based Protein Design. *Journal of Molecular Biology*, 432(13):3898–3914, 6 2020. ISSN 10898638. doi: 10.1016/j.jmb.2020.04.013.
- [195] Sergey Nepomnyachiy, Nir Ben-Tal, and Rachel Kolodny. Complex evolutionary footprints revealed in an analysis of reused protein segments of diverse lengths. *Proceedings of the National Academy of Sciences of the United States of America*, 114(44):11703–11708, 2017. ISSN 10916490. doi: 10.1073/pnas.1707642114.
- [196] Tobias Sikosek and Hue Sun Chan. Biophysics of protein evolution and evolutionary protein biophysics. *Journal of the Royal Society Interface*, 11(100):20140419, 2014. ISSN 17425662. doi: 10.1098/rsif.2014.0419.
- [197] Michael J. Harms and Joseph W. Thornton. Evolutionary biochemistry: Revealing the historical and physical causes of protein properties. *Nature Reviews Genetics*, 14(8):559–571, 8 2013. ISSN 14710056. doi: 10.1038/nrg3540.
- [198] Kevin T Halloran, Yanming Wang, Karunesh Arora, Srinivas Chakravarthy, Thomas C Irving, Osman Bilsel, Charles L Brooks, C Robert Matthews, and C. Robert Matthews. Frustration and folding of a TIM barrel protein. *Proceedings of the National Academy of Sciences of the United States of America*, 116(33):16378–16383, 7 2019. ISSN 10916490. doi: 10.1073/pnas.1900880116.
- [199] Yvonne H. Chan, Sergey V. Venev, Konstantin B. Zeldovich, and C. Robert Matthews. Correlation of fitness landscapes from three orthologous TIM barrels originates from sequence and structure constraints. *Nature Communications*, 8(1):1–12, 3 2017. ISSN 20411723. doi: 10.1038/ncomms14614.

Manuscripts

Sergio Romoero-Romero, Migule Costas, Daniel-Adriano Silva Manzano, **Sina Kordes**, Erendira Rojas-Ortega, Cinthya Tapia, Yasel Guerra, Sooruban Shanmugaratnam, Adel Rodriguez-Romero, David Baker, Birte Höcker, D. Alejandro Fernandez-Velasco
The Stability Landscape of *de novo* TIM Barrels Explored by a Modular Design Approach
Journal of Molecular Biology, 2021, 433(18)



The Stability Landscape of *de novo* TIM Barrels Explored by a Modular Design Approach

Sergio Romero-Romero^{1,2}, Miguel Costas³, Daniel-Adriano Silva Manzano^{4,5†}, Sina Kordes², Erendira Rojas-Ortega¹, Cinthya Tapia¹, Yasel Guerra^{1‡}, Sooruban Shanmugaratnam², Adela Rodríguez-Romero⁶, David Baker^{4,5*}, Birte Höcker^{2*} and D. Alejandro Fernández-Velasco^{1*}

1 - Laboratorio de Fisicoquímica e Ingeniería de Proteínas, Departamento de Bioquímica, Facultad de Medicina, Universidad Nacional Autónoma de México, 04510 Mexico City, Mexico

2 - Department of Biochemistry, University of Bayreuth, 95447 Bayreuth, Germany

3 - Laboratorio de Biofisicoquímica, Departamento de Fisicoquímica, Facultad de Química, Universidad Nacional Autónoma de México, 04510 Mexico City, Mexico

4 - Department of Biochemistry, University of Washington, 98195 Seattle, USA

5 - Institute for Protein Design, University of Washington, 98195 Seattle, USA

6 - Instituto de Química, Universidad Nacional Autónoma de México, 04510 Mexico City, Mexico

Correspondence to David Baker, Birte Höcker and D. Alejandro Fernández-Velasco: Institute for Protein Design, University of Washington, 98195 Seattle, USA (D. Baker), Department of Biochemistry, University of Bayreuth, 95447 Bayreuth, Germany (B. Höcker), Departamento de Bioquímica, Facultad de Medicina, Universidad Nacional Autónoma de México, 04510 Mexico City, Mexico (D. A. Fernández-Velasco). dabaker@uw.edu (D. Baker), birte.hoecker@uni-bayreuth.de (B. Höcker), fdaniel@unam.mx (D.A. Fernández-Velasco)
<https://doi.org/10.1016/j.jmb.2021.167153>

Edited by Amy Keating

Abstract

The ability to design stable proteins with custom-made functions is a major goal in biochemistry with practical relevance for our environment and society. Understanding and manipulating protein stability provide crucial information on the molecular determinants that modulate structure and stability, and expand the applications of *de novo* proteins. Since the $(\beta/\alpha)_8$ -barrel or TIM-barrel fold is one of the most common functional scaffolds, in this work we designed a collection of stable *de novo* TIM barrels (DeNovoTIMs), using a computational fixed-backbone and modular approach based on improved hydrophobic packing of sTIM11, the first validated *de novo* TIM barrel, and subjected them to a thorough folding analysis. DeNovoTIMs navigate a region of the stability landscape previously uncharted by natural TIM barrels, with variations spanning 60 degrees in melting temperature and 22 kcal per mol in conformational stability throughout the designs. Significant non-additive or epistatic effects were observed when stabilizing mutations from different regions of the barrel were combined. The molecular basis of epistasis in DeNovoTIMs appears to be related to the extension of the hydrophobic cores. This study is an important step towards the fine-tuned modulation of protein stability by design.

© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Proteins are essential macromolecules capable of performing diverse and exquisite biological functions such as immune protection, cellular communication, or enzymatic reactions. To guarantee such activities, the functional states must act under specific environmental conditions in a relevant time scale, that is, proteins must be “stable”. Protein stability is required to maintain functional structures and it enhances the ability of proteins to evolve new properties.^{1,2} The central role of proteins in the chemistry of life, as well as their increasing application in basic and applied research, implies that the understanding and manipulation of protein stability are highly relevant.

There are two main indicators of protein conformational stability at equilibrium. One is the difference of free energy between the native and unfolded states at a given temperature (ΔG), which is often obtained by chemical unfolding experiments carried out at 25 °C. In addition, stability is also assessed in the context of thermal unfolding, where the unfolding temperature (T_m), the temperature at the midpoint of the transition from native to the unfolded state, is the most common parameter employed to quantify stability. Both the ΔG and T_m parameters, usually determined as criteria for a “stable” protein, are related with the enthalpy (ΔH) and heat capacity (ΔC_p) changes through the Gibbs-Helmholtz equation, which describes the variation of ΔG with temperature, the so-called “stability curve” of proteins.³ Different mechanisms have been proposed to modify the stability curve of proteins,⁴ and numerous studies on natural proteins and their site-directed mutants have been used to rationalize the stability of thermophilic proteins and moreover to engineer thermostability.⁵

Historically, the design of stable proteins has been one of the main objectives of computational protein design.⁶ Several strategies, such as increasing the hydrophobic area in internal cores, improvement of water-protein interactions, the introduction of disulfide bridges as well as the addition of salt bridges, have been proposed.^{7–18} The design of *de novo* proteins can further enhance our understanding of the physicochemical properties that modulate stability. For example, although folding behavior has been only addressed for very few cases, the kinetic analysis of the folding mechanism of two *de novo* $\beta\alpha$ proteins has revealed complex free energy surfaces.^{19,20} The fine-tuning of conformational stability, that is, the manipulation of the protein stability curve, is an open challenge for protein design and engineering. Such a goal requires a comprehensive characterization of *de novo* proteins, describing the combination of thermodynamic parameters that can be reached in a particular fold.

Within the different topologies that a protein can adopt, the TIM-barrel or $(\beta/\alpha)_8$ -barrel fold is one of

the most abundant superfolds in nature.²¹ Based on proteomic analysis, the TIM-barrel domain is also close to the average size of proteins present in *Escherichia coli*.²² Besides, the TIM-barrel fold is one of the most successful topologies used in nature to host catalytic activities. Due to its large variety of functions and its ubiquity in different types of enzymes, the TIM barrel represents a suitable scaffold for protein design and engineering.²³ For these reasons, its construction has been an important objective over the years.^{24–28} Recently, the successful design of a *de novo* four-fold symmetric TIM barrel was described: the sTIM11 protein.²⁹ Considering that the sTIM11 sequence is significantly different from the ones found in naturally occurring TIM barrels, the potential of this scaffold to explore new thermodynamic properties and functions is highly interesting. sTIM11 shows a high melting temperature ($T_m = 80$ °C) but low conformational stability ($\Delta G_{25^\circ\text{C}} = \sim 4$ kcal mol⁻¹) when compared to natural TIM barrels.^{29–32} Since low conformational stability often results in high sensitivity to mutations and changes in the environment, this can limit the design of novel proteins with new functions.⁸ Thus, fine-tuning the stability of the sTIM11 scaffold is a prerequisite to functionalize and generate tailor-made barrels for applications in biochemistry, biotechnology, and medicine. In this work, a fixed-backbone design with a modular approach was used to generate a collection of *de novo* TIM barrels. Their thermodynamic and structural properties were characterized in detail, increasing our knowledge on how stability can be fine-tuned by design.

Results and discussion

DeNovoTIM collection designed by modular repacking of a *de novo* TIM barrel

The *de novo* protein sTIM11 is an idealized four-fold symmetric TIM barrel of 184 residues, which was designed to include two cysteines that, however, did not form the intended disulfide bond (Figure 1). To avoid reactive free thiols, both residues were reverted to the residues in the original four-fold design (C8Q and C181V), resulting in sTIM11noCys. The base design DeNovoTIM0, which is the starting point for all further constructs in this work, additionally contains the changes W34V and A38G in all symmetry-related quarters. These residues are situated in every second α/β -loop, and in sTIM11, these tryptophan residues are the most highly solvent exposed. While different strategies have been explored to increase protein stability,^{8,18} here we focused on hydrophobic repacking. The structural analysis suggested three regions to be amenable to improvements in sTIM11, one in the internal and two in the peripheral hydrophobic cores. The inner face of the circular sheet forms the internal core, whereas the outer face of the strands and

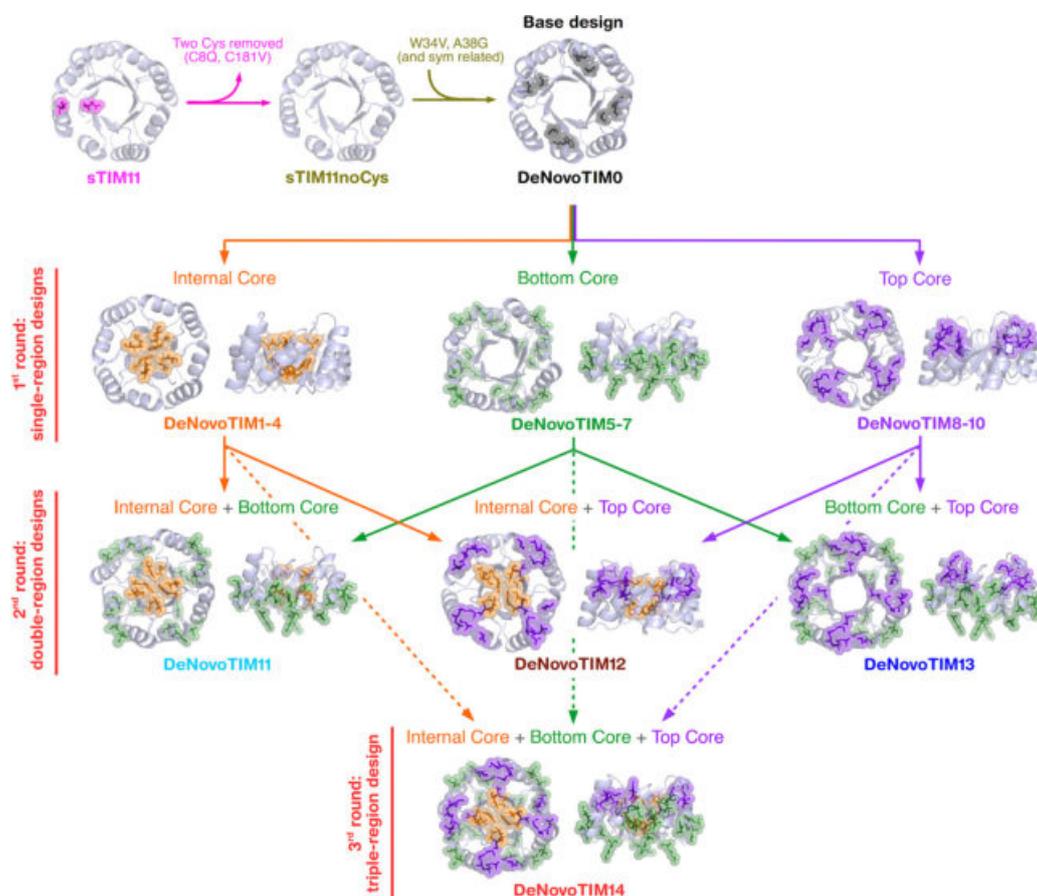


Figure 1. Modular design approach to obtain the DeNovoTIM collection. Cartoon representation of the regions and the corresponding residues modified in each design round. The two cysteine residues present in sTIM11 that were reverted to the corresponding symmetry-related residues in sTIM11noCys are shown in magenta (C8Q and C181V). Mutations W34V and A38G (as well as their 4-fold-symmetry related residues) introduced in DeNovoTIM0 are shown in black. The internal core, formed by the β -barrel residues A21, R23, I40, I42, A67, R69, I86, and I88 (as well as their 2 fold-symmetry related residues) is shown in orange. The bottom core, formed by the N-terminal region of even β -strands and the C-region of the flanking α -helices, that is, residues Q11, E15, T18, K31, and V34 (as well as their 4-fold-symmetry related residues) is colored green. The top core situated at the C-terminal region of the odd β -strands and the N-terminal region of the flanking α -helices formed by residues K2, A5, W6, Y22, S24, and D29 (as well as their 4-fold-symmetry related residues) is shown in purple. All the sequences analyzed in this work are reported in Figure S2 and Tables S1 and S2.

the internal face of the helices constitute the peripheral core. In this latter, we identified two regions, henceforth named bottom and top cores as shown in Figure 1. The residues lining the three regions were subjected to fixed-backbone Rosetta design according to the flow diagram in Figure S1.

Ten designs were selected for characterization in the first round: four with modifications in the internal core (DeNovoTIM1-4) as well as three designs each for the bottom (DeNovoTIM5-7) and the top core (DeNovoTIM8-10) (Figure S2). For the internal core, no improved designs could be identified when four-fold symmetry was preserved. Therefore, in DeNovoTIM1-4 only a two-fold symmetry was enforced. An exploratory

characterization by circular dichroism (CD) and differential scanning calorimetry (DSC) of DeNovoTIMs 0–10 showed that designs 1, 6, and 8 were the best for each region (Figure S3 and supporting text).

To test for additivity effects on stability and structure, mutations contained in the best design of each group were combined to generate the following double-region designs: DeNovoTIM11-13 as shown in Figure 1. Finally, in the third round the mutations of all three regions were combined resulting in DeNovoTIM14. All these proteins as well as sTIM11, sTIM11noCys, and DeNovoTIM0 were characterized in detail. Information on sequences, and mutations in each design are

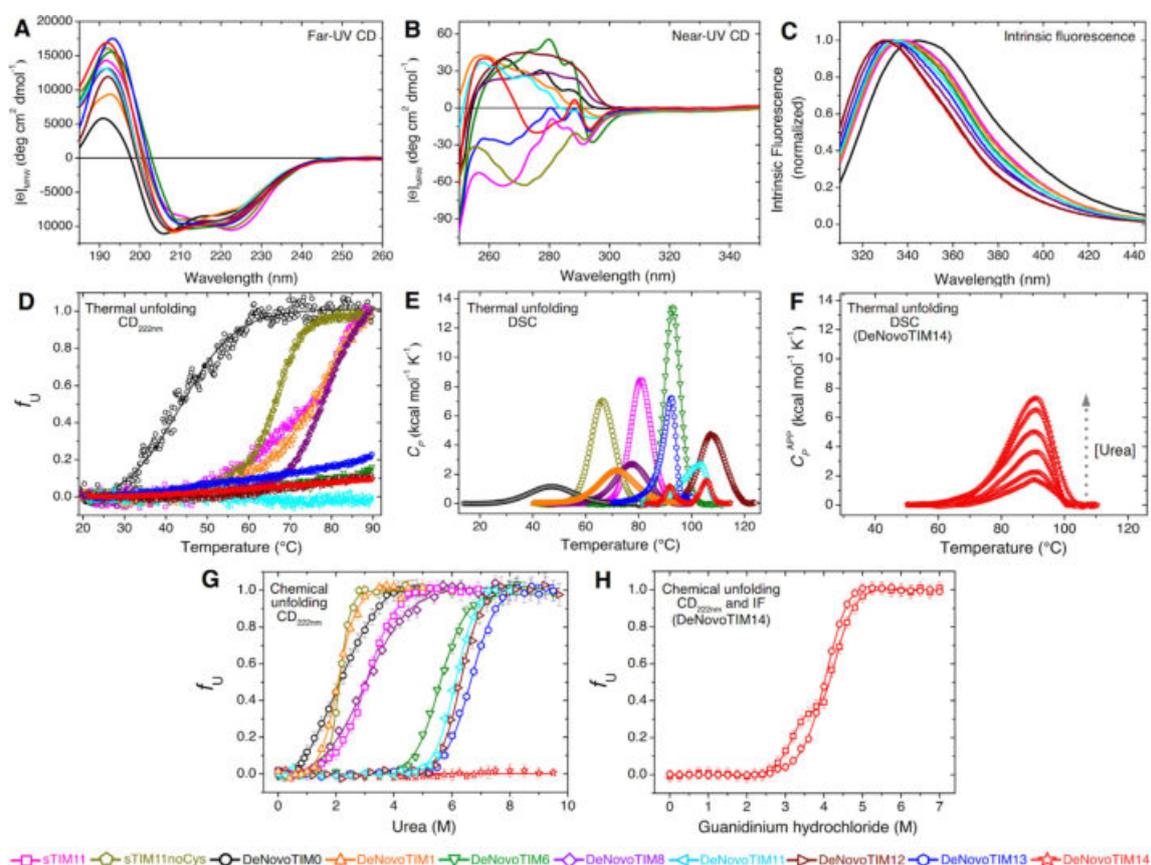


Figure 2. Conformational properties and equilibrium unfolding of DeNovoTIMs. (A) Far-UV CD spectra. (B) Near-UV CD spectra. (C) Intrinsic Fluorescence (IF) spectra ($\lambda_{\text{exc}} = 295 \text{ nm}$). (D) Thermal unfolding followed by CD_{222nm} (scan rate: 1.5 K min^{-1}). (E) Differential Scanning Calorimetry (DSC) endotherms (scan rate: 1.5 K min^{-1} ; for easy comparison, the physical and chemical baselines have been subtracted). (F) DSC endotherms of DeNovoTIM14 in the presence of increasing concentrations of urea (2.0 to 6.0 M) from bottom to top (scan rate: 1.5 K min^{-1}). For clarity, in panels E and F only a small part of the pre- and post-transition baselines are shown. (G) Chemical unfolding using urea and followed by CD (notice that DeNovoTIM14 does not unfold with urea). (H) Chemical unfolding induced by guanidinium hydrochloride for DeNovoTIM14 (squares: CD, circles: IF).

reported in the [supporting information](#) (Figure S2 and Tables S1–S4).

Folding thermodynamics of DeNovoTIMs

All variants presented the characteristic far-UV CD spectra observed for α/β proteins (Figure 2 (A) and Figure S4). The near-UV CD and intrinsic fluorescence (IF) spectra showed that the aromatic residues are buried from solvent and structured in the folded state (Figure 2(B)–(C) and Figures S5–S6; see supporting text for details). All DeNovoTIMs are monomeric and compact as revealed by analytical size exclusion chromatography (Table S5).³³

Thermal unfolding was then studied by CD and DSC (Figure 2(D)–(E)). All DeNovoTIMs showed cooperative transitions with a remarkably broad range of T_m values, from $47 \text{ }^\circ\text{C}$ (DeNovoTIM0) to $109 \text{ }^\circ\text{C}$ (DeNovoTIM12) (Table 1); indeed at $90 \text{ }^\circ\text{C}$

many of the proteins still showed secondary and tertiary structure (Figure S4(B) and Figure S5(B)). All DeNovoTIMs, except 13 and 14, showed thermal unfolding reversibility (Figure S7) and were well fitted to the two-state model ($N \rightleftharpoons U$) (Figure S8 and Table 1). This is remarkable because the temperature-induced unfolding of natural proteins of this size, particularly TIM barrels, is usually not reversible.^{22,30} DeNovoTIM14 showed two endotherms, suggesting the presence of an unfolding intermediate (Figure S8(I)). For DeNovoTIM 13 and 14, endotherms were well-fitted to an irreversible two-state mechanism ($N \rightarrow F$) giving activation energies (E_{act}) of 120 and 37 kcal mol^{-1} (Table 1), respectively, resulting in very different kinetic stabilities (Figures S9–S10 and supporting text).

For DeNovoTIMs with a reversible thermal unfolding, the observed unfolding ΔH and ΔC_p also vary greatly (Table 1). For some

Table 1 Thermodynamic properties of DeNovoTIMs.

de novo TIM barrel	Thermal unfolding (by DSC) ^a			Global thermodynamic stability (kcal K mol ⁻¹) ^c			Chemical unfolding (by CD and IF) ^e		
	T _m (°C)	ΔH (kcal mol ⁻¹)	ΔH _{85°C} (kcal mol ⁻¹) ^b	ΔC _P (kcal mol ⁻¹ K ⁻¹)	ΔH _{VH} / ΔH	Global stability (kcal K mol ⁻¹) ^c	ΔG _{85°C} (kcal mol ⁻¹)	m (kcal mol ⁻¹ M ⁻¹)	D _{T(1/2)} (M)
sTIM11	80.0 ± 0.2	93 ± 1	104 ± 2	2.19 ± 0.09	0.99 ± 0.05	280	4.8 ± 0.3	1.34 ± 0.06	3.1
sTIM11noCys	65.6 ± 0.1	82 ± 1	128 ± 2	2.36 ± 0.08	0.99 ± 0.03	176	3.2 ± 0.2	2.03 ± 0.10	1.9
DeNovoTIM0	47.0 ± 0.2	25 ± 1	41 ± 3	0.44 ± 0.06	1.09 ± 0.08	61	1.5 ± 0.1	0.76 ± 0.02	2.1
DeNovoTIM1	71.0 ± 0.4	48 ± 1	58 ± 2	0.69 ± 0.05	0.98 ± 0.05	226	3.8 ± 0.1	1.87 ± 0.10	2.0
DeNovoTIM6	92.3 ± 0.1	125 ± 2	108 ± 1	2.38 ± 0.06	1.03 ± 0.02	542	7.9 ± 0.2	1.51 ± 0.08	5.6
DeNovoTIM8	77.3 ± 0.3	52 ± 1	61 ± 1	1.19 ± 0.07	0.95 ± 0.09	165	2.5 ± 0.2	0.85 ± 0.09	2.9
DeNovoTIM11	103.5 ± 0.2	62 ± 2	49 ± 1	0.67 ± 0.14	1.02 ± 0.09	523	8.4 ± 0.4	1.75 ± 0.08	6.0
DeNovoTIM12	108.8 ± 0.3	79 ± 2	62 ± 1	0.72 ± 0.08	1.01 ± 0.08	792	10.9 ± 0.2	1.77 ± 0.03	6.2
DeNovoTIM13	92.8 ± 0.4	47 ± 5	ND E _{act} : 120 ± 3 kcal mol ⁻¹ d			ND	9.5 ± 0.2	1.54 ± 0.03	6.6
DeNovoTIM14	91.5 ± 0.1	5.4 ± 0.2	ND E _{act} : 37 ± 1 kcal mol ⁻¹ d			ND	Tot: 23.6 ± 2.0 N-I: 4.3 ± 0.9 I-U: 2.5 ± 0.1 I-U: 10.9 ± 0.6		
	105.3 ± 0.1	8.6 ± 0.4							

5

ND: Not determined due to irreversibility in the thermal unfolding. Instead, activation energy (E_{act}) was calculated from an irreversible two-state mechanism.

^aThe thermodynamic parameters reported are the average of ten experiments carried out at different protein concentrations (0.25–2.5 mg mL⁻¹; Figure S8 and Figure S11), ± indicate the standard deviation calculated from these 10 experiments.

^bΔH at 85 °C (ΔH_{85°C}) was calculated using the experimental ΔH and ΔC_P values as indicated in material and methods.

^cThe global thermodynamic stability was calculated from the area of the stability curve evaluated between 0 °C and T_m (Figure 3(B)).

^dE_{act} value and ± are the average and standard deviation, respectively, from three different calculation methods (Figure S9 and Figure S10).

^e± indicates the standard error from global fitting (Figure S13).

DeNovoTIMs these values are smaller than expected for a protein of 184 residues ($\Delta H = 128 \pm 4 \text{ kcal mol}^{-1}$ and $\Delta C_p = 2.6 \pm 0.04 \text{ kcal mol}^{-1} \text{ K}^{-1}$, according to parametric equations reported in,³⁴ while the ΔH values observed for the first and second design rounds (0.24 to $0.64 \text{ kcal mol}^{-1} \text{ residue}^{-1}$) are similar to those reported for natural monomeric TIM barrels (0.25 to $0.67 \text{ kcal mol}^{-1} \text{ residue}^{-1}$). Obtained ΔC_p values were independent of protein concentration and showed a small standard deviation (Figure S8, Figure S11, and Table 1). A decrease in ΔC_p has been shown to result from residual structure in the unfolded state.³⁵ This is observed in the far-UV CD spectra of those DeNovoTIMs that are unfolded at 90°C . In addition, the low ΔH of DeNovoTIM14 increases in the presence of urea (Figure 2(F) and Figure S10). These results suggest that for some DeNovoTIMs, the reason for the low ΔH and ΔC_p is likely the high content of residual structure in the unfolded state (supporting text).

Stability at 25°C was studied by chemical unfolding with urea or GdnHCl. Except for DeNovoTIM14, all designs were completely unfolded in 9.0 M urea (Figures S4–S6). For all designs, except for DeNovoTIM14, CD and IF curves were monophasic, cooperative, coincident, and globally-fitted well to a two-state $N \rightleftharpoons U$ model, indicating the absence of populated intermediates (Figure 2(G) and Figures S12–S13). DeNovoTIM14 showed no changes in CD or IF signal up to 9.0 M of urea, even after incubation for 5 days (Figures S4–S6). In contrast, in the presence of GdnHCl DeNovoTIM14 showed a three-state unfolding process with a populated intermediate: $N \rightleftharpoons I \rightleftharpoons U$ (Figure 2(H) and Figure S13).

All selected first- and second-round designs had a $\Delta G_{25^\circ \text{C}}$ higher than DeNovoTIM0, whereas the triple-design, DeNovoTIM14, showed a pronounced increase in stability ($\Delta G_{\text{Tot}} = 23.6 \text{ kcal mol}^{-1}$; Table 1). The stability change related to the loss of the native state in DeNovoTIM14 ($\Delta G_{N-I} = 12.7 \text{ kcal mol}^{-1}$), the so called “relevant stability”,³⁶ is higher than the ΔG of the second-round designs; whereas the stability of the intermediate, also referred as the “residual stability”, is similar ($\Delta G_{I-U} = 10.9 \text{ kcal mol}^{-1}$). For all DeNovoTIMs, the m value, a parameter proportional to the surface area exposed to the solvent upon unfolding (ΔASA),³⁷ is similar to those observed for natural proteins with the same size, except for DeNovoTIM 0 and 8 where m decreases significantly (Table 1). Low m values have been related to low packing in the native state, residual structure in the unfolded state, or the presence of folding intermediates.³⁸ Residual structure in the unfolded state is not clearly observed in the CD spectra of DeNovoTIMs at 9.0 M urea (Figures S4–S5), therefore, other techniques and kinetic studies should help to detect the persistence of native-like structure in the

unfolded state and/or the presence of intermediates.^{38,39}

The modular approach used in this work improved both ΔG and T_m substantially and hence produced significantly more stable proteins. In this context, it is worth mentioning that over the years the combination of stabilizing mutations has been considered an effective strategy to enhance the stability of small proteins.^{36,40–43} Previous work on small globular proteins with optimized hydrophobic cores and interactions on the surface exhibited increased thermal stability by up to 30°C degrees.^{9,11,15} Extending these strategies from point mutants to regions appears to be useful for bigger folds such as the TIM barrel. In what follows, using the thermal and chemical unfolding data described above, the thermodynamic properties underlying the stability of DeNovoTIMs are analyzed.

Global thermodynamic stability and non-additive effects of DeNovoTIMs

As observed in natural proteins, the m values obtained from the chemical unfolding of sTIM11, sTIM11noCys, DeNovoTIM 0, 6, and 8 correlate with their ΔC_p values (Figure 3(A)), likely because both depend on the ΔASA upon unfolding. In contrast, ΔC_p values obtained for DeNovoTIM 1, 11, and 12 are much lower than expected (Figure 3(A)). According to the Rosetta models and the native state structures (see below), these differences are not exclusively due to properties of the native state since the calculated ΔASA is close to the expected value for the size of DeNovoTIMs (17135 \AA^2).³⁷ This suggests that the unfolded state reached at high temperatures is more structured than the one obtained by chemical unfolding.

Thermal unfolding reversibility allowed the assessment of DeNovoTIM stability curves (Figure 3(B)). The $\Delta G_{25^\circ \text{C}}$ values are in excellent agreement with those obtained from chemical unfolding experiments. According to the Gibbs-Helmholtz equation, conformational stability is modulated by changes in T_m , ΔH , and ΔC_p . For natural TIM barrels, it has been observed that changes in the stability curve are influenced mainly by modifying one or two of those parameters.^{30,31} In contrast, the DeNovoTIMs differ in all three parameters. Increasing ΔH is the most commonly found mechanism for stabilization of thermophilic proteins⁵ and is also the most often exploited mechanism for engineering protein stability.^{7,40} In DeNovoTIMs, enthalpy-driven stabilization is found in all proteins but is especially important in DeNovoTIM6 (Figure 3(B)). ΔC_p determines the magnitude of the curvature of the stability curve so that changes in this parameter trigger a more or less flattened curve. A decrease in ΔC_p has been postulated as a mechanism for thermostabilization.^{35,44}

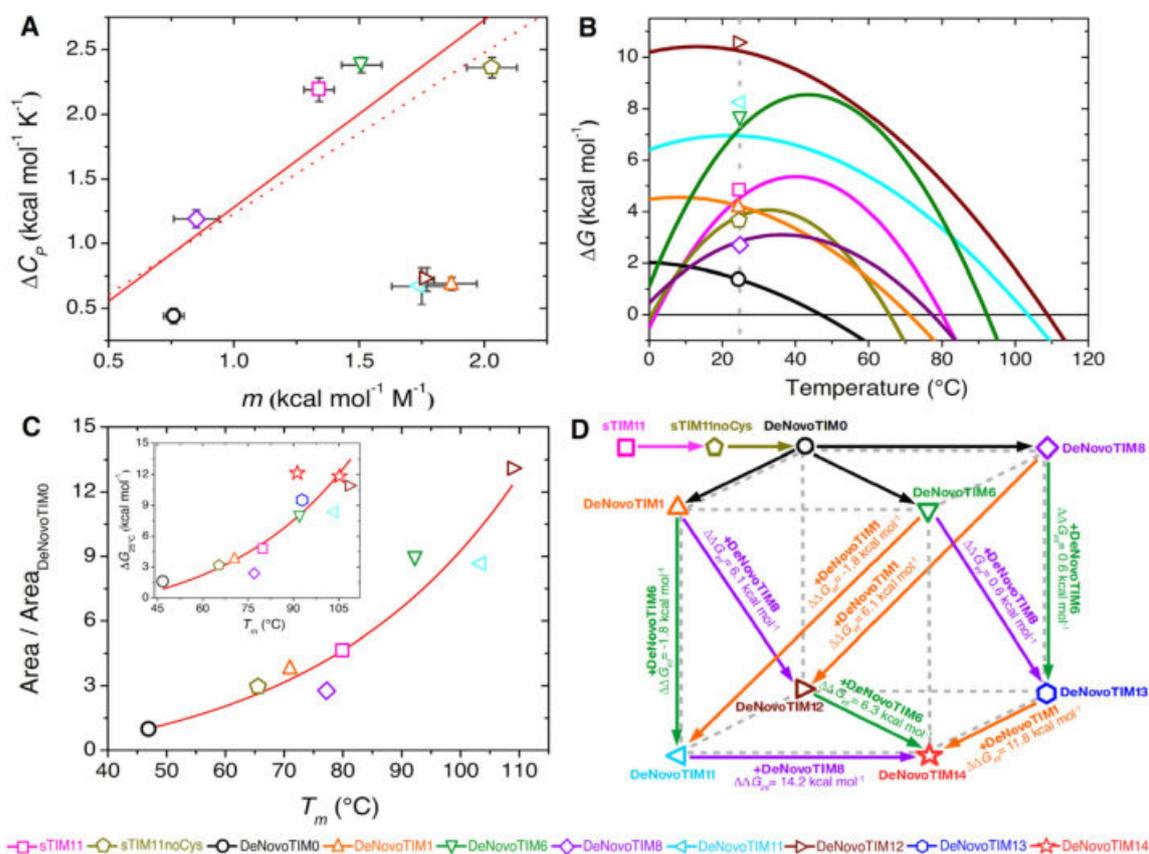


Figure 3. Stability and energetic coupling in DeNovoTIMs. (A) Correlation between two parameters which are proportional to the exposed surface area: m value from chemical unfolding and ΔC_p from temperature-induced unfolding (solid line: linear regression excluding DeNovoTIM1, DeNovoTIM11, and DeNovoTIM12 data; R^2 : 0.76. Dotted line: correlation reported by 37). (B) Stability curves calculated from DSC data (lines) using the Gibbs-Helmholtz equation (open symbols show ΔG values determined by chemical unfolding at 25 °C). Grey dashed line indicates 25 °C. (C) Correlation between the relative global thermodynamic stability ($\text{Area}/\text{Area}_{\text{DeNovoTIM0}}$) and thermostability (T_m) (R^2 : 0.93). Inset: correlation between ΔG at 25 °C determined by chemical unfolding and T_m (R^2 : 0.87). For DeNovoTIM14, where two transitions were found, it was assumed that the one observed at lower [GdnHCl] corresponds to the lower T_m . (D) Thermodynamic cube showing the coupling energy ($\Delta\Delta G_{int}$) between different regions of DeNovoTIMs. $\Delta\Delta G_{int}$ values were calculated from the double-mutant cycles shown in Figure S14. $\Delta\Delta G_{int}$ values between single-region mutants are depicted as colored arrows from the top face to the bottom face. $\Delta\Delta G_{int}$ values calculated for the addition of a single-region design to a double-region design are shown as colored arrows in the bottom face.

For DeNovoTIMs, the reduction in ΔC_p combined with an increase in ΔH is the reason for the increase in both T_m and $\Delta G_{25^\circ\text{C}}$. The results presented here indicate that, as observed for natural proteins, the unfolded ensemble plays an important role in shaping the stability curve and should be considered in protein design.

DeNovoTIMs show a non-linear correlation between $\Delta G_{25^\circ\text{C}}$ and T_m (inset in Figure 3(C)). A similar trend between ΔG at the temperature where it is a maximum ($\Delta G_{T_{\text{max}}}$) and T_m has also been reported for natural and engineered proteins with different sizes and topologies.^{34,45,46} Additionally; the global thermodynamic stability can be con-

veniently described by the area (from 0 °C to T_m) under the stability curve (A). Instead of using a single reference temperature, A integrates the conformational stability in a temperature range.⁴⁷ The relative global stability of DeNovoTIMs ($A/A_{\text{DeNovoTIM0}}$) is also correlated with T_m (Figure 3(C)). Notably, for DeNovoTIM 6, 11, and 12, $A/A_{\text{DeNovoTIM0}}$ is nearly ten-fold higher than the initial design (Figure 3(C) and Table 1).

The modular design strategy allowed us to calculate the contribution of each region to global stability, and to evaluate the presence of non-additive effects between regions of the barrel. Non-additive effects were evaluated as $\Delta\Delta G_{int}$

through an approach based on thermodynamic double mutant cycles (see material and methods). $\Delta\Delta G_{int}$ is also referred to as coupling energy, non-additive effect, interaction energy, and more recently epistatic effect.⁴⁸

Thermodynamic cycles showed that the stabilization is non-additive and depends on the structural context (Figure S14). All the $\Delta\Delta G_{int}$ values calculated in Figure S14 are summarized in the single cube shown in Figure 3(D). $\Delta\Delta G_{int}$ for double designs are much smaller than those involving the triple-region design. The regions that are most energetically coupled in double-region designs ($\Delta\Delta G_{int} = 6.1 \text{ kcal mol}^{-1}$) are the internal core (DeNovoTIM1) and the top core (DeNovoTIM8). Coupling increases considerably when a third region is incorporated on top of two already mutated regions ($\Delta\Delta G_{int} > 6 \text{ kcal mol}^{-1}$). The largest $\Delta\Delta G_{int}$ was observed when the DeNovoTIM8 mutations were added to DeNovoTIM11 ($\Delta\Delta G_{int} = 14.2 \text{ kcal mol}^{-1}$) (Figure 3(D) and Figure S14). Clearly, mutations in one place affected other regions of the barrel. The latter indicates that the TIM-barrel fold is suitable for studying modularity and, in general, cooperative effects of proteins. Also, the results presented here suggest that the modular design strategy could be used in the future for the rational stability improvement in other protein topologies.

Structural features of DeNovoTIMs

The structural properties of DeNovoTIMs were examined by X-ray crystallography (Table S6). High-resolution data were collected for sTIM11noCys and DeNovoTIM13, whereas a low-resolution structure was obtained for DeNovoTIM6. All showed the designed globular compact TIM-barrel topology (Figure 4). Structural comparison with the Rosetta models showed the lowest RMSD located in the second quarter of the barrel. As previously observed in sTIM11,²⁹ the main structural differences are found in the α -helices located at the N- and C-terminal ends. In agreement, for all the barrel structures, the RMSD among quarters of the barrel is higher in the first and fourth ones (plot in Figure 4(A)). Since the TIM barrel is a closed-repeat protein, contacts between the first and last helices depend on the precise curvature generated by each α/β unit, therefore geometrical strain may interfere with the proper closure of the barrel.

A comparison of the sTIM11noCys and sTIM11 structures showed that removal of the two cysteines causes some structural changes mainly localized in the first and last quarters; the most significant deviations are observed at the N-terminal region where the first two helices are not well-formed. So even without forming the disulfide bridge, both cysteines in sTIM11 increase the stability and promote a proper closure of the barrel (Figure 4(B) and Table 1); nevertheless,

sTIM11noCys maintains the general expected TIM-barrel architecture.

The thermodynamic properties of DeNovoTIM6 are very similar to those expected for a natural protein (Table 1). Unfortunately, due to the low quality of the crystals and therefore the low resolution obtained (2.9 Å), details such as side-chain conformations are not well resolved in its structure. Nevertheless, it could be verified that the protein is a well folded TIM-barrel (Figure 4(C)). In DeNovoTIM6, almost all α/β loops of the barrel are well defined and correspond to the model. However, for some residues within 5 of the 7 β/α loops no electron density was observed. In general, the DeNovoTIM6 structure has high B factors which may reflect higher disorder in the protein crystal or increased flexibility, similar to observations in some regions of sTIM11, namely the amino- and carboxyl-terminal α -helices.

In the DeNovoTIM13 structure, the second and third quarters display only minor differences to the Rosetta model, with the secondary structure elements and side chains superposing very well (Figure 4(D)). In going from sTIM11 to DeNovoTIM13, a 60 % increase in the total area of hydrophobic clusters was found (3765 vs. 6148 Å²); most of this change comes from a three-fold increase in the area of the major hydrophobic cluster (1116 vs. 4351 Å², Table S7).

As a consequence of the DeNovoTIMs design protocol, polar interactions were replaced by hydrophobic ones, therefore, it is not surprising that the number of H-bonds and salt bridges is lower in DeNovoTIMs than in sTIM11 and sTIM11noCys (Table S7). Some of the designs that contained the highest number of polar stabilizing interactions (such as DeNovoTIM 1 and 8) were not the most stable ones, whereas some of the most stable designs (such as DeNovoTIM 6 and 12–14) showed a reduction in this type of interaction (Table S7). In agreement with the design strategy, the stability of DeNovoTIMs increases with the number of hydrophobic interactions. The total area, as well as the number of residues and contacts in hydrophobic clusters, are substantially increased in the best first-round designs along with the more stable second- and third-round designs (Figure S15 and Table S7). This suggests that the strategy of increasing hydrophobic contacts was successful in the stabilization of DeNovoTIMs.

Epistasis on the stability landscape of de novo TIM barrels

To correlate the most common and informative parameters obtained from both temperature and chemical unfolding, T_m , ΔH , and ΔG were plotted in a two-dimensional bubble plot thereby representing the thermodynamic combinations of the designed proteins (Figure 5). The T_m values found in DeNovoTIMs are widely distributed

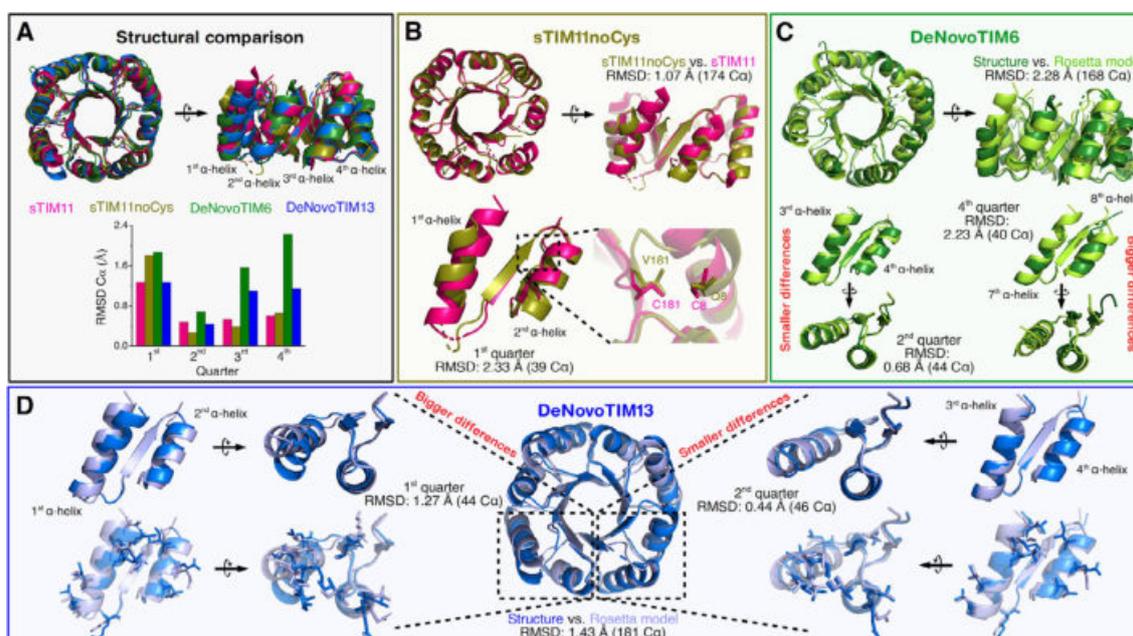


Figure 4. Three-dimensional structures of DeNovoTIMs. (A) Structural alignment of X-ray structures of sTIM11 (PDB ID: 5BVL), sTIM11noCys (PDB ID: 6YQY), DeNovoTIM6 (PDB ID: 6Z2I), and DeNovoTIM13 (PDB ID: 6YQX). The RMSD C α between the structure and the Rosetta model among the quarters in each protein is shown in the lower part of the panel. (B) Comparison of sTIM11noCys and sTIM11 structures (RMSD: 1.07 Å – 174 C α -). The mutated residues 8 and 181 in sTIM11noCys are zoomed in the bottom part. (C) Comparison of the DeNovoTIM6 structure with the Rosetta model (RMSD: 2.28 Å – 168 C α -). The quarters with the highest and lowest structural similarity are highlighted (bottom left and bottom right, respectively). (D) Comparison of the DeNovoTIM13 structure with the Rosetta model (RMSD: 1.43 Å – 181 C α -). The quarters with the highest and lowest structural similarity are highlighted (right and left, respectively). Sidechains of the mutated residues are shown in sticks.

ranging from 47 °C to 109 °C, corresponding to a 62 °C increase in thermostability, a range higher than those previously reported for engineered proteins, but in the observed range found in natural proteins (Figure 5). The stability measured from chemical-unfolding experiments ($\Delta G_{25^\circ\text{C}}$) also shows the variety found in natural TIM barrels and other natural folds (Figure 5(B) and (C)). Natural proteins populate some thermodynamic regions more than others, exploring ample space due to the diversity in size, topology, oligomeric state, function, and evolutionary history. Interestingly, Figure 5 shows that several DeNovoTIMs are located in a region of the plot corresponding to low ΔH and high $\Delta G_{25^\circ\text{C}}$ values, which has not been reported for natural proteins.

Assuming additivity, the expected change in stability calculated for DeNovoTIM14 would be the sum of the individual stabilizations provided by all the single-region designs giving a value of 11.2 kcal mol⁻¹. However, the stability of DeNovoTIM14 is 23.6 kcal mol⁻¹, indicating that more than half of the stabilization comes from positive non-additive effects. Non-additive effects or interaction energies may be referred to as epistasis, a concept traditionally used in genetics to describe the phenotype dependency of a

mutation on the genetic state at other sites.^{48–50} Previous studies have explored and analyzed the mechanisms of epistasis within proteins, especially regarding their implications for protein function, evolution, and stability.^{51–55}

Rearrangements in the TIM barrel can influence local changes in other parts of the protein, and these epistatic effects are quantified in the $\Delta\Delta G_{int}$ values whose magnitude for DeNovoTIMs is considerable. The structural analyses suggest that the epistatic effect observed in DeNovoTIMs is likely related to the extension of the hydrophobic cores, particularly to the increase of the major hydrophobic cluster located in the interface between the inner β -barrel and the outer α -helices (Figure S15 and Table S7). From the first- to the second-round designs, the highest area in hydrophobic clusters was found for DeNovoTIM12, and this corresponds to the highest positive epistatic effect in this round ($\Delta\Delta G_{int} = 6.1$ kcal mol⁻¹), whereas the decrease of the hydrophobic cluster area in DeNovoTIM11 (compared to DeNovoTIM1 and DeNovoTIM6) correlates with a negative $\Delta\Delta G_{int} = -1.8$ kcal mol⁻¹. From the second- to the third-round designs, the most notable change in hydrophobic area is observed in going from DeNovoTIM11 to

DeNovoTIM14, resulting in the highest positive epistatic effect ($\Delta\Delta G_{int} = 14.2 \text{ kcal mol}^{-1}$). The relevance and magnitude of the epistatic or non-additive effects found in DeNovoTIMs, as well as those observed in other reports, suggest that modeling such interactions can improve the success in protein design and engineering.

Conclusions

Design requires a deep understanding of the relationship between sequence, structure, and stability, and therefore, the combination of thermodynamic and structural data is fundamental to achieve this goal. Here, we designed a family of stable TIM barrels and comprehensively explored their thermodynamic and structural properties. The TIM-barrel collection reported in this work exhibits a considerable range in thermostability (more than 60 degrees in T_m) and conformational stability at 25 °C (more than 22 kcal mol⁻¹ in ΔG). These data can now be used to accelerate the development of future custom design protein stability curves which, in turn, will expand the biomedical and biotechnological applications of *de novo* proteins. For example, by fusion to another *de novo* protein, one of the stabilized scaffolds reported here (DeNovoTIM13) has been successfully used to create a reaction chamber on the top of the barrel,⁵⁸ confirming the convenience of working with robust and stable TIM barrels in the path towards functional *de novo* proteins.

In the same way that one explores the sequence space by studying homologous proteins from different organisms, *de novo* design with a fixed backbone follows a similar strategy generating new sequences within the same topology. It is well known that highly stable proteins can be generated by computational design. However, one of the unexpected findings resulting from the thermodynamic characterization of this family of DeNovoTIMs is that very stable proteins presenting unexplored combinations of thermodynamic parameters can be designed. The stability of DeNovoTIMs is severely influenced by epistatic effects that appear to arise from the design strategy involving an increase in hydrophobic clusters. The design and characterization of stable *de novo* proteins, such as those described in this work, is an essential step on the route to the next generation of new protein functions charting novel sequence space.

Material and methods

Enzymes and chemicals

All reagents were of analytical grade from Merck KGaA[®]. Genes were ordered from GenScript Biotech. Water was distilled and deionized.

Design protocol

De novo TIM barrels were designed using the Rosetta software suite v.3.2^{59,60} (<https://www.rosetta-commons.org/>). All DeNovoTIMs were designed using DeNovoTIM0 as a template. The script used for the DeNovoTIM collection follows and executes the steps indicated by the algorithm as indicated in Figure S1. In general, the algorithm first selects the symmetry with which it designs the proteins. Once the two-fold or four-fold symmetry was chosen, it selects the number of residues that mutate (depending on whether it is a quarter or half of the protein). Then, an energy minimization step is performed by simulated annealing considering and evaluating the packing and RMSD. Subsequently, it performs a Monte Carlo (MC) fast layer design to improve the packing of the protein's hydrophobic cores (in one or several cavities selected according to the regions described in Figure 1), minimizes the constraints of the main and side chains, and compares each of them with the starting design. For each step, it verifies the RMSD value between both proteins (in the case of DeNovoTIMs, the design was done with a fixed backbone and a cut-off pointed out < 0.7 Å). Then, the algorithm filters the results to keep those designs that, with the suggested mutations, were able to increase the packing and preserve the reference topology (ScoreRes: ≤ -1.9 , Talaris: ≤ -3.5 BetaNov, Sspred: ≥ 0.85 , Packstat: ≥ 0.65). To evaluate if the suggested protein folds as expected, selected designs were computationally validated by a later step of forward folding to predict the *ab initio* three-dimensional structure. The selection was done by an energy score, choosing the designs with the lowest energy value and smallest possible RMSD (located at the bottom left when the energy score against RMSD is plotted). In all selected DeNovoTIMs, funnel plots were observed. Finally, the designs were analyzed and the candidates for experimental characterization were selected based on energy criteria, a fewer number of mutations, and physicochemical properties of the suggested mutations.

Cloning, overexpression, and protein purification

The nucleotide sequence of all DeNovoTIMs was optimized for expression in *Escherichia coli*. The coding genes were synthesized and cloned into the pET29b(+) vector by GeneScript (New Jersey, USA), except sTIM11noCys, which was cloned into pET21b(+). Proteins were overexpressed in *E. coli* strain BL21(DE3) (Invitrogen[®]) in 1 L of Terrific Broth (TB) medium supplemented with 30 $\mu\text{g mL}^{-1}$ kanamycin or 100 $\mu\text{g mL}^{-1}$ ampicillin, inoculated with 5 mL preculture and incubated at 37 °C and 200 rpm. After an OD₆₀₀ of 0.6–0.8 was reached, overexpression was induced by adding 1 mM isopropyl-D-1-thiogalactopyranoside (IPTG); growth was continued for 16 h at 30 °C. After

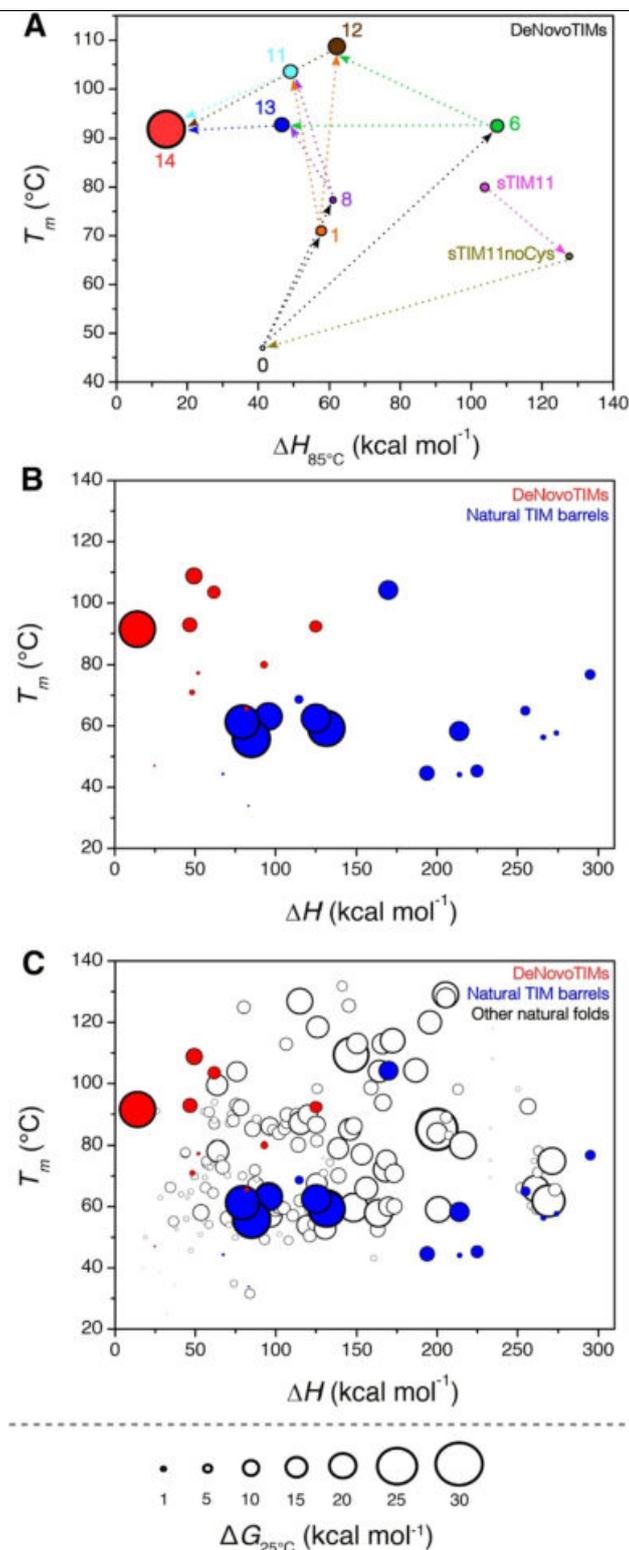


Figure 5. Thermodynamic 2D bubble plots of *de novo* TIM barrels in comparison with natural proteins. (A) 2D plot of $\Delta H_{85^\circ\text{C}}$ versus T_m of all DeNovoTIMs. Colored arrows indicate the design flow. **(B)** 2D plot of ΔH versus T_m of natural TIM barrels (in blue) and DeNovoTIMs (in red). **(C)** 2D plot as shown in B including data for other natural folded proteins (open circles). The diameter in the bubbles correspond to the $\Delta G_{25^\circ\text{C}}$ magnitude. T_m and ΔH data were obtained from thermal unfolding, whereas ΔG values derive from chemical unfolding. Data for all non-redundant proteins presented here were obtained from ProThermDB database.^{56,57}

incubation, cells were harvested by centrifugation (Thermo/SLA-3000[®], 15 min, 8000 rpm, 4 °C), pellets resuspended in buffer A: 35 mM sodium phosphate, 300 mM NaCl and 35 mM Imidazole pH 8 (supplemented with 0.2 mM of protease inhibitor Phenylmethylsulfonyl fluoride), lysed by sonication (Cole Parmer Ultrasonic Processor[®], 10 cycles in 45 s intervals, 30% pulse, 4 °C), and centrifuged again (Sorvall/SS-34[®], 40 min, 16000 rpm, 4 °C). In some cases, to increase the efficiency of lysis, the resuspended cells were incubated with lysozyme (250 µg mL⁻¹) at 37 °C for 1 h before sonication. The purification was performed loading the supernatant onto a HisTrap HP column (5 mL; GE Healthcare Life Sciences[®]) coupled to an ÄKTA system (GE Healthcare Life Sciences[®]). The unbound fraction was washed out with 20 column volumes (CV) of buffer A. Bound protein was eluted with a linear gradient of 35–500 mM Imidazole using buffer B: 35 mM sodium phosphate, 300 mM NaCl and 500 mM Imidazole pH 8. The pooled fractions were loaded onto a HiLoad 16/600 Superdex 75 preparative grade column (GE Healthcare Life Sciences[®]). The proteins were purified using isocratic elution with 1.5 CV of buffer C: 150 mM NaCl, 35 mM sodium phosphate pH 8. The fractions corresponding to the monomeric population were pooled and stored at 4 °C for use in subsequent experiments. It should be noted that at the end of protein purification, all designs contain a polyhistidine-tag in the carboxyl-terminal region. For DeNovoTIM11 and DeNovoTIM14, the following purification variables were modified to increase the yield: 0.1 mM of IPTG for induction at OD₆₀₀ of 0.2–0.3, 30 °C and 6 h for overexpression, buffer A and B containing 1 M NaCl, and buffer C with 300 mM NaCl. At each step of the purification process, aliquots were taken to quantify the amount of protein and to calculate the corresponding purification tables. The final yields are indicated in Table S5.

Far- and near-UV circular dichroism

Circular Dichroism (CD) spectra were collected in buffer D: 10 mM sodium phosphate pH 8 in a Chirascan Spectropolarimeter using a Peltier device to control the temperature (Applied Photophysics[®]). For Far-UV spectra, 0.4 mg mL⁻¹ of DeNovoTIM was used for all measurements (1 nm bandwidth, 185–260 nm wavelength range, 1 mm cuvette). For Near-UV spectra, 1 mg mL⁻¹ of DeNovoTIM was used for all measurements (1 nm bandwidth, 250–350 nm wavelength range, 10 mm cuvette). The spectra for thermally-unfolded states were collected at 90 °C. Spectra for chemically-unfolded states were collected at 9 M urea for all DeNovoTIMs, except for DeNovoTIM14, which was collected at 7 M GdnHCl. Raw data were converted to mean residue molar ellipticity ([θ]) using: $[\theta] = \theta / (l C N_r)$,

where θ is ellipticity collected in millidegrees, l is the cell path length in mm, C is the DeNovoTIM molar concentration, and N_r the number of residues per protein. Far-UV spectra were deconvoluted with CDNN.⁶¹

Intrinsic fluorescence

Intrinsic Fluorescence (IF) spectra were collected on a PC1 ISS Spectrofluorometer (Champaign IL-USA[®]) equipped with a Peltier device controlling the temperature. In all measurements, protein concentration was 0.4 mg mL⁻¹ in buffer D: 10 mM sodium phosphate pH 8 (1 nm bandwidth slits, 295 nm excitation wavelength, 310–450 nm emission wavelength range). Spectra for chemically-unfolded states were collected at 9 M urea for all DeNovoTIMs, except for DeNovoTIM14, which was collected at 7 M GdnHCl. Fluorescence spectral center of mass (SCM) was calculated from intensity data (I_λ) obtained at different wavelengths (λ): $SCM = \frac{\sum \lambda I_\lambda}{\sum I_\lambda}$.

Three-dimensional structure determination

DeNovoTIMs were concentrated with Amicon Ultra centrifugal filter units (Millipore[®]) and dialyzed in buffer C: 10 mM sodium phosphate pH 8, 150 mM NaCl. Sitting-drop vapor-diffusion method, and JCSG Core I-IV, JCSG +, Classics I-II, PACT, PEGs I-II, and AmSO₄ screening suites (Qiagen[®]) were used to screen crystallization conditions in 96 well Intelli plates (Art Robbins Instruments[®]) stored at 20 °C in the hotel-based Rock Imager RI 182 (Formulatrix[®]). 0.8 µL drops were prepared in a 1:1 ratio with mother liquid using a nanodispensing crystallization robot Phoenix (Art Robbins Instruments[®]) and then optimized by multiple crystallization rounds using a sitting-drop vapor-diffusion method. To improve the diffraction quality of DeNovoTIM crystals, different pre- and post-crystallization methods were used: reductive methylation (JBS Methylation Kit, Jena Biosciences[®]), seeding, additive screening, controlled dehydration, cryoprotection screening, crystal annealing, and room-temperature diffraction. In total, more than 300 different crystals in various conditions were tested.

Suitable crystals for X-ray diffraction were found in the following conditions: sTIM11noCys: 0.2 M Ammonium Sulfate, 0.1 M Trisodium Citrate pH 5.6, 25% w/v Polyethylene glycol (PEG) 4000, with a protein concentration of 15 mg mL⁻¹; DeNovoTIM6: 0.095 M Sodium Citrate pH 5.0, 19% v/v Isopropanol, 25% w/v PEG 4000, 5% v/v Glycerol, with a protein concentration of 8.6 mg mL⁻¹; DeNovoTIM13: 0.17 M Sodium Acetate, 0.085 M Tris pH 8.5, 25.5% w/v PEG 4000, 15% v/v Glycerol, with a protein concentration of 10 mg mL⁻¹.

For sTIM11noCys and DeNovoTIM13, diffraction data were collected at 100 K at the Swiss Light Source at the Paul Scherrer Institute in Villigen (Switzerland) (X10SA-PXII beamline for sTIM11noCys and X06DA-PXIII beamline for DeNovoTIM13) using a wavelength of 1 Å and a PILATUS 6M detector for sTIM11noCys and a PILATUS 2M-F detector for DeNovoTIM13.^{62,63} For DeNovoTIM6, diffraction data were collected at 100 K at the Berlin Electron Storage Ring Society for Synchrotron Radiation beamline 14.2 (BESSY BL14.2) operated by Helmholtz-Zentrum Berlin using a wavelength of 0.91 Å and a PILATUS3S 2 M detector.⁶⁴

Diffraction data were processed with the X-ray Detector Software (XDS) using XDSAPP v.2.0^{65,66} for sTIM11noCys and DeNovoTIM13; and DIALS⁶⁷ for DeNovoTIM6. For data reduction; criteria used to cut off the data were the resolution shell with a mean/sigma(I) between 1–2 and the best CC1/2 according to redundancy and completeness. The structures were solved by molecular replacement with PHASER in the PHENIX software suite v.1.17⁶⁸ using sTIM11 (PDB ID: 5BVL) as a starting model for sTIM11noCys and the own Rosetta model for DeNovoTIM6 and DeNovoTIM13. Refinement was done with phenix.refine.⁶⁸ The model was improved by map inspection and iteratively manual rebuilding performed in COOT v.0.9.⁶⁹ The final coordinates were validated with PDB_REDO,⁷⁰ MolProbity v.4.2,⁷¹ and the Protein Data Bank validation service⁷²; in all servers; the 3D-structure satisfied all quality criteria. The coordinates and structure factors were deposited in the PDB with accession codes: 6YQY (sTIM11noCys), 6ZZI (DeNovoTIM6), and 6YQX (DeNovoTIM13). The figures were created using PyMOL Molecular Graphics System v.4.5.0 (Schrodinger, LLC).

Analytical size exclusion chromatography

Hydrodynamic measurements were performed on a Superdex 75 10/300 GL analytical column coupled to an ÄKTA System (GE Healthcare Life Sciences®). All experiments were performed in buffer C: 10 mM sodium phosphate pH 8, 150 mM NaCl at 25 °C and a protein concentration range from 0.01 to 2.0 mg mL⁻¹. Experimental molecular weight, Stokes-radii, and oligomeric state were calculated from elution volumes and a calibration curve derived from 7 different known proteins.

Thermal unfolding followed by circular dichroism

Temperature-induced unfolding was monitored by CD at 222 nm as a function of temperature using 0.4 mg mL⁻¹ in buffer D: 10 mM sodium phosphate pH 8, a heating rate of 1.0 and 1.5 K min⁻¹, and a 1 mm path-length cell. The changes in the CD signal were normalized to the fraction of unfolded molecules (f_U) by:

$$f_U = \frac{y_{obs} - (y_N + m_N T)}{(y_U + m_U T) - (y_N + m_N T)} \quad (1)$$

where y_{obs} is the experimentally observed CD signal at a given temperature, and $(y_N + m_N T)$ and $(y_U + m_U T)$ are the linear fitting equations corresponding to the native and unfolded regions, respectively. T_m values were estimated from normalized data fitted with a Boltzmann-type function:

$$f_U = \frac{-1}{\left(1 + e^{\frac{T-T_m}{a}}\right)} + 1 \quad (2)$$

where a is related to the slope of the transition.

Thermal unfolding followed by differential scanning calorimetry

Differential Scanning Calorimetry (DSC) scans were carried out in a VP-Capillary DSC system (MicroCal®, Malvern Analytical). Samples were prepared by exhaustive dialysis in buffer D: 10 mM sodium phosphate pH 8 and then degassed at room temperature. To ascertain proper instrument equilibration, two buffer–buffer scans were performed before each protein–buffer scan (Figure S7). Corresponding buffer–buffer traces were subtracted from each endotherm. For all proteins a reheating scan was performed to determine the reversibility or irreversibility of the process (Figure S7). Reversibility percentage was calculated by comparing the calorimetric ΔH (area under the curve) recovered in the second scan and that obtained in the first one ($\Delta H_{\text{secondscan}} / \Delta H_{\text{firstscan}} * 100$). To verify that irreversibility was not the result of a too high final scanning temperature, the first scans were also performed heating near the T_m . For DeNovoTIMs with a reversible thermal unfolding, protein concentration varied from 0.25 to 2.5 mg mL⁻¹ and scan rates from 1 to 3 K min⁻¹, except for DeNovoTIM0 where protein concentration varied from 1 to 5 mg mL⁻¹. For DeNovoTIMs with an irreversible thermal unfolding, protein concentration was 1 mg mL⁻¹ and scan rates from 1 to 3 K min⁻¹. For DeNovoTIM14 in native conditions, protein concentration was increased to 2.5 and 4.5 mg mL⁻¹ to accurately determine the transition. For DeNovoTIM14 in the presence of urea, all the scans were done at 1 mg mL⁻¹ from 2.0 to 6.0 M urea with samples incubated for 6 h at 10 °C. Origin v.9.0 (OriginLab Corporation, Northampton, MA, USA.) with MicroCal software was used for data analysis.

Thermodynamic parameters from reversible DSC transitions

DSC endotherms were fitted to equilibrium two-state model ($N \rightleftharpoons U$):

$$C_P(T) = B_0 + B_1 T + f(T) \Delta C_P + \frac{\Delta H(T)}{RT_m^2} \left[\frac{1 - f(T)}{1 - n + \frac{n}{f(T)}} \right] \quad (3)$$

where B_0 and B_1 define the slope and intercept of the low-temperature baseline segment, n is the number of subunits in the native protein (1 for all DeNovoTIMs) and $f(T)$ is the protein fraction in the folded monomeric state, producing ΔH (at T_m), ΔC_P , and T_m . The thermodynamic parameters reported are the average of ten experiments carried out in the 0.25 to 2.5 mg mL⁻¹ range. The van't Hoff enthalpy (ΔH_{vH}) was evaluated by⁷³:

$$\Delta H_{vH} = \frac{4RT_m^2 C_{P,T_m}}{\Delta H} \quad (4)$$

where R is the universal gas constant, T_m is the temperature at which C_P is maximal, C_{P,T_m} is the heat capacity value at T_m , and ΔH is the total calorimetric enthalpy of the endotherm.

Thermodynamic parameters from irreversible DSC transitions

Calorimetric transitions were adequately described by the two-state irreversible model (N → F) where N is the native protein and F is the final state.^{74,75} The kinetic conversion from N to F is described by a first-order rate constant (k) changing with temperature according to the Arrhenius equation:

$$k = \exp\left[\frac{-E_{act}}{R}\left(\frac{1}{T} - \frac{1}{T''}\right)\right] \quad (5)$$

where T'' is the temperature at which the $k = 1 \text{ min}^{-1}$ and E_{act} is the activation energy between the native and the transition states that describes the unfolding process. The apparent heat capacity is given by:

$$C_P^{APP} = \frac{\Delta H E_{act}}{RT_m^2} \exp(x) \exp[-\exp(x)]; x = \frac{E_{act}}{RT_m} (T - T_m) \quad (6)$$

where T is the temperature and ΔH is the unfolding enthalpy. The E_{act} was also obtained following these two procedures: from the slope of Arrhenius plots, i.e. $\ln k$ vs. $1/T$; and derived from a data consistency test, evaluating the effect of scanning rate (ν) on T_m .⁷⁶

Chemical-induced unfolding

All experiments were carried out at a protein concentration of 0.1 mg mL⁻¹ in buffer D: 10 mM sodium phosphate pH 8 at 25 °C. To determine whether urea induced unfolding was reversible, unfolding and refolding experiments were assayed. For unfolding experiments, native DeNovoTIM was the initial state, whereas for refolding, the starting state was the unfolded DeNovoTIM incubated overnight in 9.0 M urea. Thereafter samples were incubated at different concentrations of urea (0–9.0 M), either increasing or decreasing the initial concentration (for unfolding and refolding experiments, respectively). Intrinsic fluorescence of both, unfolding and refolding samples, was measured at different times to determine the equilibrium time. Unfolding and refolding transitions are coincident and the

signal does not change after incubation for 12 h, i.e. chemical unfolding is reversible and at equilibrium under the experimental conditions. Once the equilibrium time was found, unfolding experiments with samples incubated for 12 h and followed by CD and IF were performed as aforementioned. IF data at fixed emission wavelength and CD data at 222 nm were both collected over 2 minutes at each urea concentration. The changes in IF and CD were normalized to the fraction of unfolded molecules (f_U) by:

$$f_U = \frac{y_{obs} - (y_N + m_N[\text{urea}])}{(y_U + m_U[\text{urea}]) - (y_N + m_N[\text{urea}])} \quad (7)$$

where y_{obs} is the experimentally observed IF and CD signal at a given temperature, and $(y_N + m_N[\text{urea}])$ and $(y_U + m_U[\text{urea}])$ are the linear fitting equations corresponding to the native and unfolded regions, respectively. All two-state transitions were fitted to Santoro and Bolen equation⁷⁷ which assumes a two-state model (N ⇌ D):

$$f_U = \frac{(y_N + m_N[\text{urea}]) + (y_U + m_U[\text{urea}]) e^{-\frac{\Delta G^{H2O} - m[\text{urea}]}{RT}}}{1 + e^{-\frac{\Delta G^{H2O} - m[\text{urea}]}{RT}}} \quad (8)$$

where ΔG^{H2O} is the unfolding free energy in absence of denaturant, m is $\Delta G/[\text{urea}]$, T is the temperature of the experiment (25 °C), and $(y_N + m_N[\text{urea}])$ and $(y_U + m_U[\text{urea}])$ are the linear fitting equations for the pre- and post-transition states. The chemical unfolding transitions for DeNovoTIM14 in GdnHCl were fitted to a three-state model with an intermediate:

$$f_U = \frac{(y_U + m_U[\text{GdnHCl}])K_1K_2 + (y_N + m_N[\text{GdnHCl}]) + (y_I + m_I[\text{GdnHCl}])K_1}{1 + K_1 + K_1K_2} \quad (9)$$

where $K_1 = e^{-\frac{\Delta G_{NtoI} - m_{NtoI}[\text{GdnHCl}]}{RT}}$, $K_2 = e^{-\frac{\Delta G_{ItoU} - m_{ItoU}[\text{GdnHCl}]}{RT}}$, ΔG_{NtoI} and ΔG_{ItoU} is the unfolding free energy from native state to intermediate and from intermediate to unfolded state, m_{NtoI} and m_{ItoU} is $\Delta G/[\text{GdnHCl}]$ of each step, T is the temperature of the experiment (25 °C), and $(y_N + m_N[\text{GdnHCl}])$, $(y_I + m_I[\text{GdnHCl}])$, and $(y_U + m_U[\text{GdnHCl}])$ are the linear fitting equations for native, intermediate, and unfolded states, respectively. Similar ΔG values were obtained when experimental protein concentration was increased five fold, ruling out the possibility of a bimolecular association/folding step.

Stability curve and global thermodynamic stability

Global stability curves, $\Delta G(T)$, were calculated using the thermodynamic parameters obtained from DSC experiments and the Gibbs-Helmholtz equation⁷⁸:

$$\Delta G(T) = \Delta H \left(1 - \frac{T}{T_m}\right) - \Delta C_P \left(T_m - T + T \ln\left(\frac{T}{T_m}\right)\right) \quad (10)$$

The area under the stability curve is a measure of the global stability of the protein.⁴⁷ It was calculated integrating Eq. (10) from the lowest temperature at

which the protein is in the liquid state i.e. 0 °C (273.15 K) to T_m :

$$\begin{aligned} \text{Area} = & ((\Delta H - T_m \Delta C_P)(T_m - T)) \\ & - \left(\frac{\Delta H}{2T_m} - \frac{\Delta C_P}{2} \right) (T_m^2 - T^2) + \left(\frac{\Delta C_P}{4} T_m^2 \right) \\ & + \frac{\Delta C_P}{2} \left(T^2 \ln \frac{T}{T_m} - \frac{T^2}{2} \right) \end{aligned} \quad (11)$$

Thermodynamic 2D bubble plots

The 2D bubble plots were constructed by plotting T_m and $\Delta H_{85^\circ\text{C}}$ obtained from thermal unfolding experiments, and $\Delta G_{25^\circ\text{C}}$ obtained from chemical unfolding data. Since DeNovoTIMs have different T_m values, experimental ΔH from DSC experiments cannot be directly compared. To put the thermodynamic parameters on a similar ground for comparison in Figure 5(A), ΔH at 85 °C ($\Delta H_{85^\circ\text{C}}$), which is the average T_m of the DeNovoTIM collection, was calculated using ΔH and ΔC_P from DSC experiments as follows⁷⁹:

$$\Delta H_{85^\circ\text{C}} = \Delta H + \Delta C_P(85^\circ\text{C} - T_m) \quad (12)$$

where T_m , ΔH and ΔC_P are the experimental values obtained from DSC experiments for each protein, and 85 °C is the reference temperature. $\Delta H_{85^\circ\text{C}}$ was not calculated for DeNovoTIM13 and DeNovoTIM14, because their irreversible thermal unfolding hampered the determination of ΔC_P . The plots were constructed with Origin v.9.0 (OriginLab Corporation, Northampton, MA, USA.) and the diameter in the bubbles correspond to the $\Delta G_{25^\circ\text{C}}$ magnitude.

Thermodynamic double-mutant cycles

To calculate non-additive effects between different DeNovoTIM barrel regions, an approximation based on double mutant cycles was used.^{48,80,81} The thermodynamic cycles were constructed using the experimental $\Delta G_{25^\circ\text{C}}$ values obtained from chemical unfolding experiments and linking single-region/double-region designs and double-region/triple-region designs as indicated in Figure S14.

Each corner of the square represents a different DeNovoTIM where the mutations are located in a specific region of the barrel or in a combination of them. For double-region cycles (upper panel), from the first to the second design round, ΔG_1 and ΔG_2 are the changes in stability produced when a single region of the barrel was mutated, ΔG_3 and ΔG_4 are the changes in stability generated when the same mutations are evaluated in the background of another first-round design. In the triple-region cycles (lower panel), from the second to the third design round, ΔG_1 and ΔG_3 are the changes in stability produced when the mutations of a single region are introduced in the background of DeNovoTIM0 or in a double-region design, whereas ΔG_2 and ΔG_4 are the changes in

stability generated when a double-region design was incorporated in the background of DeNovoTIM0 or in a single region design, respectively.

Considering that ΔG is a state property, if two regions of the barrel are energetically independent, their effects will be additive and not coupled. Therefore, stability changes linked to a particular region will result in the same values on parallel sides of the square, i.e., $\Delta G_1 = \Delta G_3$ and $\Delta G_2 = \Delta G_4$. Any difference the values on the parallel sides of the squares indicates a deviation from additivity and measures the coupling energy between different regions of the barrel, given by $\Delta\Delta G_{int} = \Delta G_4 - \Delta G_2 = \Delta G_3 - \Delta G_1$, where $\Delta\Delta G_{int}$ values have been referred as coupling energy, non-additive effects, interaction energies, and more recently epistatic effects.⁴⁸ A positive ΔG_{int} indicates that the introduction of favorable interactions has a higher stabilizing effect when a nearby region is already mutated.

Sequence and structural analysis

Sequence alignment was performed with MAFFT v.7.450⁸² using the secondary structure information from the sTIM11 structure (PDB ID: 5BVL). Sequence identity was calculated with the SIAS server (Universidad Complutense de Madrid, 2013). Structural alignments and RMSD calculations were performed using PyMOL Molecular Graphics System v.4.5.0 (Schrodinger, LLC). Cavity volumes were calculated with MOLE v.2.5⁸³ using a standard probe radius of 5 Å and an interior threshold of 1.1 Å with a non-directed exploration path. The accessible surface area (ASA) was calculated with VADAR v.1.8.⁸⁴ In these analyses; changes in ASA for the unfolded state were calculated with an extended Gly-X-Gly peptide. Hydrogen bonds, as well as salt bridges, were calculated using HBPLUS v.3.06⁸⁵ and ESBRI⁸⁶ with default parameters for distances and angles. A salt bridge was assigned when two atoms of opposite charge were observed within 4 Å. Hydrophobic clusters (formed by ILV residues) were calculated following an algorithm previously reported by Sobolev⁸⁷ and available in the *ProteinTools* toolkit developed by Dr. Noelia Ferruz-Capapey from the Höcker Lab.⁸⁸ The analysis considers ILE; VAL; and LEU residues and then recaptures the coordinates of their neighboring atoms. Then, the buried solvent-accessible hydrophobic surface area is calculated, and the cluster's total area is computed by the sum of the individual residue areas that comprise it.

Accession numbers

Coordinates and structure files have been deposited to the Protein Data Bank (PDB) with

accession numbers: 6YQY (sTIM11noCys), 6Z2I (DeNovoTIM6), and 6YQX (DeNovoTIM13).

Data and materials availability

All data to support the conclusions of this manuscript are included in the main text and [supporting information](#).

CRedit authorship contribution statement

Sergio Romero-Romero: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization, Supervision. **Miguel Costas:** Conceptualization, Formal analysis, Investigation, Resources, Writing - original draft, Writing - review & editing, Funding acquisition. **Daniel-Adriano Silva Manzano:** Software, Investigation, Writing - review & editing. **Sina Kordes:** Methodology, Investigation, Writing - review & editing. **Erendira Rojas-Ortega:** Methodology, Investigation, Writing - review & editing. **Cinthya Tapia:** Methodology, Investigation, Writing - review & editing. **Yasel Guerra:** Methodology, Investigation, Writing - review & editing. **Sooruban Shanmugaratnam:** Methodology, Investigation, Writing - review & editing. **Adela Rodríguez-Romero:** Methodology, Investigation, Resources, Writing - review & editing, Supervision, Funding acquisition. **David Baker:** Conceptualization, Resources, Writing - review & editing, Supervision, Project administration, Funding acquisition. **Birte Höcker:** Conceptualization, Resources, Writing - original draft, Writing - review & editing, Supervision, Project administration, Funding acquisition. **D. Alejandro Fernández-Velasco:** Conceptualization, Formal analysis, Investigation, Resources, Writing - original draft, Writing - review & editing, Supervision, Project administration, Funding acquisition.

the Fernandez-Velasco, Höcker, and Baker Labs for their constructive suggestions to improve the research.

Funding

This work was supported by scholarships from CONACYT (749489 to C.T., 387653, 291062, 14401, and 27897 to S.R.R.), UNAM-DGAPA-PAPIIT (IN220516 to S.R.R.), UNAM-DGAPA (postdoctoral fellowship to Y.G.), CONACYT and UNAM-DGAPA (sabbatical stay fellowship to D.A. F.V.), and Foundations Alexander von Humboldt and Bayer Science & Education (Humboldt-Bayer Research Fellowship for Postdoctoral Researchers to S.R.R.). This research was also financed by grants from CONACYT (221169 to A. R.R., 254514 to D.A.F.V.), UNAM-DGAPA-PAPIIT (IN220519 to M.C., IN208418 to A.R.R., IN219519 and IN220516 to D.A.F.V.), and Programa de Apoyo a la Investigación y el Posgrado FQ-UNAM (5000-9018 to M.C.). B.H. gratefully acknowledges financial support by the European Research Council (ERC Consolidator Grant 647548 'Protein Lego') and by HZB to visit the beamlines at BESSY.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

This article contains supporting information that includes: Supporting text, Supporting Figures S1 to S15, and Supporting Tables S1 to S7. Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jmb.2021.167153>.

Received 29 March 2021;

Accepted 6 July 2021;

Available online 14 July 2021

Acknowledgments

We acknowledge financial support and allocation of beamtime by PSI and HZB. We thank the beamline staff at the SLS and at BESSY for assistance, and LANEM-IQ-UNAM for the support in crystal characterization. We thank María Isabel Velázquez López, Laura Iliana Alvarez Añorve, Alma Jessica Díaz Salazar, and Georgina Espinosa Pérez for their competent technical support, Gregor Wiese for generating and crystallizing sTIM11noCys, Noelia Ferruz-Capapey for her help in the structural analyses, as well as Po-Ssu Huang for his comments on the manuscript. We kindly thank all the members of

Keywords:

de novo protein design;
protein folding and stability;
stability landscape;
non-additive effects;
(β/α)₈-barrel

† Current address: Neoleukin Therapeutics, 98102 Seattle, USA.

Abbreviations:

A, area under the stability curve; T_m , midpoint of thermal unfolding; CD, Circular Dichroism; DSC, Differential Scanning Calorimetry; E_{act} , activation energy; GdnHCl,

guanidinium chloride; IF, Intrinsic Fluorescence; $\Delta\Delta G_{int}$, Gibbs free energy of coupling; $\Delta H_{85^\circ C}$, change in enthalpy extrapolated at 85 °C

References

- Tokuriki, N., Tawfik, D.S., (2009). Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.*, **19** (5), 596–604. <https://doi.org/10.1016/j.sbi.2009.08.003>.
- Bloom, J.D., Labthavikul, S.T., Otey, C.R., Arnold, F.H., (2006). Protein stability promotes evolvability. *Proc. Natl. Acad. Sci. U.S.A.*, **103** (15), 5869–5874. <https://doi.org/10.1073/pnas.0510098103>.
- Becktel, W.J., Schellman, J.A., (1987). Protein stability curves. *Biopolymers*, **26** (11), 1859–1877. <https://doi.org/10.1002/bip.360261104>.
- Nojima, H., Ikai, A., Oshima, T., Noda, H., (1977). Reversible thermal unfolding of thermostable phosphoglycerate kinase. Thermostability associated with mean zero enthalpy change. *J. Mol. Biol.*, **116** (3), 429–442. [https://doi.org/10.1016/0022-2836\(77\)90078-X](https://doi.org/10.1016/0022-2836(77)90078-X).
- Razvi, A., Scholtz, J.M., (2006). Lessons in stability from thermophilic proteins. *Protein Sci.*, **15** (7), 1569–1578. <https://doi.org/10.1110/ps.062130306>.
- Huang, P.S., Boyken, S.E., Baker, D., (2016). The coming of age of de novo protein design. *Nature*, **537** (7620), 320–327. <https://doi.org/10.1038/nature19946>.
- Islam, M.M., Kobayashi, K., Kidokoro, S.I., Kuroda, Y., (2019). Hydrophobic surface residues can stabilize a protein through improved water-protein interactions. *FEBS J.*, **286** (20), 4122–4134. <https://doi.org/10.1111/febs.14941>.
- Goldenzweig, A., Fleishman, S.J., (2018). Principles of protein stability and their application in computational design. *Annu. Rev. Biochem.*, **87** (1), 105–129. <https://doi.org/10.1146/annurev-biochem-062917-012102>.
- Kim, D.N., Jacobs, T.M., Kuhlman, B., (2016). Boosting protein stability with the computational design of β -sheet surfaces. *Protein Sci.*, **25** (3), 702–710. <https://doi.org/10.1002/pro.2869>.
- Huang, P.S., Oberdorfer, G., Xu, C., Pei, X.Y., Nannenga, B.L., Rogers, J.M., DiMaio, F., Gonen, T., Luisi, B., Baker, D., (2014). High thermodynamic stability of parametrically designed helical bundles. *Science*, **346** (6208), 481–485. <https://doi.org/10.1126/science.1257481>.
- Borgo, B., Havranek, J.J., (2012). Automated selection of stabilizing mutations in designed and natural proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **109** (5), 1494–1499. <https://doi.org/10.1073/pnas.1115172109>.
- Kuhlman, B., Baker, D., (2004). Exploring folding free energy landscapes using computational protein design. *Curr. Opin. Struct. Biol.*, **14** (1), 89–95. <https://doi.org/10.1016/j.sbi.2004.01.002>.
- Kwok, S.C., Hodges, R.S., (2003). Clustering of large hydrophobes in the hydrophobic core of two-stranded α -helical coiled-coils controls protein folding and stability. *J. Biol. Chem.*, **278** (37), 35248–35254. <https://doi.org/10.1074/jbc.M305306200>.
- Selvaraj, S., Gromiha, M.M., (2003). Role of hydrophobic clusters and long-range contact networks in the folding of (α/β)₈ barrel proteins. *Biophys. J.*, **84** (3), 1919–1925. [https://doi.org/10.1016/S0006-3495\(03\)75000-0](https://doi.org/10.1016/S0006-3495(03)75000-0).
- Malakauskas, S.M., Mayo, S.L., (1998). Design, structure and stability of a hyperthermophilic protein variant. *Nat. Struct. Biol.*, **5** (6), 470–475. <https://doi.org/10.1038/nsb0698-470>.
- Dahiyat, B.I., Mayo, S.L., (1997). Probing the role of packing specificity in protein design. *Proc. Natl. Acad. Sci. U.S.A.*, **94** (19), 10172–10177. <https://doi.org/10.1073/pnas.94.19.10172>.
- Colón, W., Elöve, G.A., Wakem, L.P., Sherman, F., Roder, H., (1996). Side chain packing of the N- and C-terminal helices plays a critical role in the kinetics of cytochrome c folding. *Biochemistry*, **35** (17), 5538–5549. <https://doi.org/10.1021/bi960052u>.
- Fersht, A.R., Serrano, L., (1993). Principles of protein stability derived from protein engineering experiments. *Curr. Opin. Struct. Biol.*, **3** (1), 75–83. [https://doi.org/10.1016/0959-440X\(93\)90205-Y](https://doi.org/10.1016/0959-440X(93)90205-Y).
- Basak, S., Nobrega, P.R., Tavella, D., Deveau, L.M., Koga, N., Tatsumi-Koga, R., Baker, D., Massi, F., Matthews, C. R., (2019). Networks of electrostatic and hydrophobic interactions modulate the complex folding free energy surface of a designed B α protein. *Proc. Natl. Acad. Sci. U. S.A.*, **116** (14), 6806–6811. <https://doi.org/10.1073/pnas.1818744116>.
- Watters, A.L., Deka, P., Corrent, C., Callender, D., Varani, G., Sosnick, T., Baker, D., (2007). The highly cooperative folding of small naturally occurring proteins is likely the result of natural selection. *Cell*, **128** (3), 613–624. <https://doi.org/10.1016/j.cell.2006.12.042>.
- Stern, R., Höcker, B., (2005). Catalytic versatility, stability, and evolution of the (beta α)₈-barrel enzyme fold. *Chem. Rev.*, **105** (11), 4038–4055. <https://doi.org/10.1021/cr030191z>.
- Braselmann, E., Chaney, J.L., Clark, P.L., (2013). Folding the proteome. *Trends Biochem. Sci.*, **38** (7), 337–344. <https://doi.org/10.1016/j.tibs.2013.05.001>.
- Romero-Romero, S., Kordes, S., Michel, F., Höcker, B., (2021). Evolution, folding, and design of TIM barrels and related proteins. *Curr. Opin. Struct. Biol.*, **68**, 94–104. <https://doi.org/10.1016/j.sbi.2020.12.007>.
- Nagarajan, D., Deka, G., Rao, M., (2015). Design of symmetric TIM barrel proteins from first principles. *BMC Biochem.*, **16** (1), 1–22. <https://doi.org/10.1186/s12858-015-0047-4>.
- Figuroa, M., Oliveira, N., Lejeune, A., Kaufmann, K.W., Dorr, B.M., Matagne, A., Martial, J.A., Meiler, J., Van de Weerd, C., (2013). Octarellin VI: using rosetta to design a putative artificial (β/α)₈ protein. *PLoS ONE*, **8**, (8) <https://doi.org/10.1371/journal.pone.0071858> e71858.
- Höcker, B., Lochner, A., Seitz, T., Claren, J., Stern, R., (2009). High-resolution crystal structure of an artificial ($\beta\alpha$)₈-barrel protein designed from identical half-barrels. *Biochemistry*, **48** (6), 1145–1147. <https://doi.org/10.1021/bi802125b>.
- Höcker, B., Claren, J., Stern, R., (2004). Mimicking enzyme evolution by generating new (beta α)₈-barrels from (beta α)₄-half-barrels. *Proc. Natl. Acad. Sci. U.S.A.*, **101** (47), 16448–16453. <https://doi.org/10.1073/pnas.0405832101>.
- Offredi, F., Dubail, F., Kischel, P., Sarinski, K., Stern, a.S., van de Weerd, C., Hoch, J.C., Prospero, C., Francois, J.M., Mayo, S.L., Martial, J.A., (2004). De novo backbone and sequence design of an idealized alpha/ protein: evidence of

- stable tertiary structure. *J. Mol. Biol.*, **325** (1), 163–174. [https://doi.org/10.1061/S022-2836\(02\)01206-8](https://doi.org/10.1061/S022-2836(02)01206-8).
29. Huang, P.S., Feldmeier, K., Parmeggiani, F., Fernandez-Velasco, D.A., Hocker, B., Baker, D., (2016). De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nat. Chem. Biol.*, **12** (1), 29–34. <https://doi.org/10.1038/nchembio.1966>.
 30. Romero-Romero, S., Costas, M., Rodríguez-Romero, A., Fernández-Velasco, D.A., (2015). Reversibility and two state behaviour in the thermal unfolding of oligomeric TIM barrel proteins. *Phys. Chem. Chem. Phys.*, **17** (32), 20699–20714. <https://doi.org/10.1039/c5cp01599e>.
 31. Cipolla, A., Delbrassine, F., Da Lage, J.L., Feller, G., (2012). Temperature adaptations in psychrophilic, mesophilic and thermophilic chloride-dependent alpha-amylases. *Biochimie*, **94** (9), 1943–1950. <https://doi.org/10.1016/j.biochi.2012.05.013>.
 32. Rudolph, R., Siebendritt, R., Kiefhaber, T., (1992). Reversible unfolding and refolding behavior of a monomeric aldolase from staphylococcus aureus. *Protein Sci.*, **1** (5), 654–666. <https://doi.org/10.1002/pro.5560010511>.
 33. Uversky, V.N., (1993). Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. *Biochemistry*, **32** (48), 13288–13298. <https://doi.org/10.1021/bi00211a042>.
 34. Robertson, A.D., Murphy, K.P., (1997). Protein structure and the energetics of protein stability. *Chem. Rev.*, **97** (5), 1251–1267. <https://doi.org/10.1021/cr960383c>.
 35. Robic, S., Guzman-Casado, M., Sanchez-Ruiz, J.M., Marqusee, S., (2003). Role of residual structure in the unfolded state of a thermophilic protein. *Proc. Natl. Acad. Sci. U.S.A.*, **100** (20), 11345–11349. <https://doi.org/10.1073/pnas.1635051100>.
 36. Lamazares, E., Clemente, I., Bueno, M., Velázquez-Campoy, A., Sancho, J., (2015). Rational stabilization of complex proteins: a divide and combine approach. *Sci. Rep.*, **5** (9129), 1–11. <https://doi.org/10.1038/srep09129>.
 37. Myers, J.K., Nick Pace, C., Martin Scholtz, J., (1995). Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.*, **4** (10), 2138–2148. <https://doi.org/10.1002/pro.5560041020>.
 38. Spudich, G., Marqusee, S., (2000). A change in the apparent m value reveals a populated intermediate under equilibrium conditions in Escherichia coli ribonuclease HI. *Biochemistry*, **39** (38), 11677–11683. <https://doi.org/10.1021/bi000466u>.
 39. Shortle, D., Ackerman, M.S., (2001). Persistence of native-like topology in a denatured protein in 8 M urea. *Science*, **293** (5529), 487–489. <https://doi.org/10.1126/science.1060438>.
 40. Haruki, M., Tanaka, M., Motegi, T., Tadokoro, T., Koga, Y., Takano, K., Kanaya, S., (2007). Structural and thermodynamic analyses of Escherichia Coli RNase HI variant with quintuple thermostabilizing mutations. *FEBS J.*, **274** (22), 5815–5825. <https://doi.org/10.1111/j.1742-4658.2007.06104.x>.
 41. Serrano, L., Day, A.G., Fersht, A.R., (1993). Step-wise mutation of barnase to binase. A procedure for engineering increased stability of proteins and an experimental analysis of the evolution of protein stability. *J. Mol. Biol.*, **233**, 305–312. <https://doi.org/10.1006/jmbi.1993.1508>.
 42. Stearman, R.S., Frankel, A.D., Freire, E., Liu, B., Pabo, C. O., (1988). Combining thermostable mutations increases the stability of λ repressor. *Biochemistry*, **27** (19), 7571–7574. <https://doi.org/10.1021/bi00419a059>.
 43. Matsumura, M., Signor, G., Matthews, B.W., (1989). Substantial increase of protein stability by multiple disulphide bonds. *Nature*, **342** (6247), 291–293. <https://doi.org/10.1038/342291a0>.
 44. Hart, K.M., Harms, M.J., Schmidt, B.H., Elya, C., Thornton, J.W., Marqusee, S., (2014). Thermodynamic system drift in protein evolution. *PLoS Biol.*, **12**, (11) <https://doi.org/10.1371/journal.pbio.1001994> e1001994.
 45. Howell, S.C., Inampudi, K.K., Bean, D.P., Wilson, C.J., (2014). Understanding thermal adaptation of enzymes through the multistate rational design and stability prediction of 100 adenylate kinases. *Structure*, **22** (2), 218–229. <https://doi.org/10.1016/j.str.2013.10.019>.
 46. Rees, D.C., Robertson, A.D., (2001). Some thermodynamic implications for the thermostability of proteins. *Protein Sci.*, **10** (6), 1187–1194. <https://doi.org/10.1110/ps.180101>.
 47. Alfano, C., Sanfelice, D., Martin, S.R., Pastore, A., Temussi, P.A., (2017). An optimized strategy to measure protein stability highlights differences between cold and hot unfolded states. *Nat. Commun.*, **8** (May), 1–9. <https://doi.org/10.1038/ncomms15428>.
 48. Horovitz, A., Fleisher, R.C., Mondal, T., (2019). Double-mutant cycles: new directions and applications. *Curr. Opin. Struct. Biol.*, **58** (Oct), 10–17. <https://doi.org/10.1016/j.sbi.2019.03.025>.
 49. Miton, C.M., Tokuriki, N., (2016). How mutational epistasis impairs predictability in protein evolution and design. *Protein Sci.*, **25** (7), 1260–1272. <https://doi.org/10.1002/pro.2876>.
 50. Starr, T.N., Thornton, J.W., (2016). Epistasis in protein evolution. *Protein Sci.*, **25** (7), 1204–1218. <https://doi.org/10.1002/pro.2897>.
 51. Risso, V.A., Manssour-Triedo, F., Delgado-Delgado, A., Arco, R., Barroso-delJesus, A., Ingles-Prieto, A., Godoy-Ruiz, R., Gavira, J.A., Gaucher, E.A., Ibarra-Molero, B., Sanchez-Ruiz, J.M., (2015). Mutational studies on resurrected ancestral proteins reveal conservation of site-specific amino acid preferences throughout evolutionary history. *Mol. Biol. Evol.*, **32** (2), 440–455. <https://doi.org/10.1093/molbev/msu312>.
 52. Ashenberg, O., Gong, L.I., Bloom, J.D., (2013). Mutational Effects on stability are largely conserved during protein evolution. *Proc. Natl. Acad. Sci. U.S.A.*, **110** (52), 21071–21076. <https://doi.org/10.1073/pnas.1314781111>.
 53. Natarajan, C., Inoguchi, N., Weber, R.E., Fago, A., Moriyama, H., Storz, J.F., (2013). Epistasis among adaptive mutations in deer mouse hemoglobin. *Science*, **340** (6138), 1324–1327. <https://doi.org/10.1126/science.1236862>.
 54. Lynch, V.J., May, G., Wagner, G.P., (2011). Regulatory evolution through divergence of a phosphoswitch in the transcription factor CEBPB. *Nature*, **480** (7377), 383–386. <https://doi.org/10.1038/nature10595>.
 55. Ortlund, E.A., Bridgman, J.T., Redinbo, M.R., Thornton, J. W., (2007). Crystal structure of an ancient protein: evolution by conformational epistasis. *Science*, **317** (5844), 1544–1548. <https://doi.org/10.1126/science.1142819>.

56. Kumar, M.D.S., (2006). ProTherm and ProNIT: thermodynamic databases for proteins and protein-nucleic acid interactions. *Nucleic Acids Res.*, **34** (90001), D204–D206. <https://doi.org/10.1093/nar/gkj103>.
57. Nikam, R., Kulandaisamy, A., Harini, K., Sharma, D., Michael Gromiha, M., (2021). ProThermDB: thermodynamic database for proteins and mutants revisited after 15 years. *Nucleic Acids Res.*, **49** (D1), D420–D424. <https://doi.org/10.1093/nar/gkaa1035>.
58. Caldwell, S.J., Haydon, I.C., Piperidou, N., Huang, P.S., Bick, M.J., Sebastian Sjöström, H., Hilvert, D., Baker, D., Zeymer, C., (2020). Tight and specific lanthanide binding in a de novo TIM barrel with a large internal cavity designed by symmetric domain fusion. *Proc. Natl. Acad. Sci. U.S.A.*, **117** (48), 30362–30369. <https://doi.org/10.1073/pnas.2008535117>.
59. Koga, N., Tatsumi-Koga, R., Liu, G., Xiao, R., Acton, T.B., Montelione, G.T., Baker, D., (2012). Principles for designing ideal protein structures. *Nature*, **491** (7423), 222–227. <https://doi.org/10.1038/nature11600>.
60. Rohl, C.A., Strauss, C.E.M., Misura, K.M.S., Baker, D., (2004). Protein structure prediction using rosetta. *Methods Enzymol.*, **383**, 66–93. https://doi.org/10.1007/978-0-387-92738-1_11.
61. Böhm, G., Muhr, R., Jaenicke, R., (1992). Quantitative analysis of protein far UV circular dichroism spectra by neural networks. *Protein Eng. Des. Sel.*, **5** (3), 191–195. <https://doi.org/10.1093/protein/5.3.191>.
62. Fuchs, M.R., Pradervand, C., Thominet, V., Schneider, R., Panepucci, E., Grunder, M., Gabadinho, J., Dworkowski, F. S.N., Tomizaki, T., Schneider, J., Mayer, A., Curtin, A., Olieric, V., Frommherz, U., Kotrlé, G., Welte, J., Wang, X., Maag, S., Schulze-Briese, C., Wang, M., (2014). D3, the new diffractometer for the macromolecular crystallography beamlines of the swiss light source. *J. Synchrotron Radiat.*, **21** (2), 340–351. <https://doi.org/10.1107/S160057751400006X>.
63. Bingel-Erlenmeyer, R., Olieric, V., Grimshaw, J.P.A., Gabadinho, J., Wang, X., Ebner, S.G., Isenegger, A., Schneider, R., Schneider, J., Gletting, W., Pradervand, C., Panepucci, E.H., Tomizaki, T., Wang, M., Schulze-Briese, C., (2011). SLS crystallization platform at beamline X06DA-A fully automated pipeline enabling in situ x-ray diffraction screening. *Cryst. Growth Des.*, **11** (4), 916–923. <https://doi.org/10.1021/cg101375j>.
64. Helmholtz-Zentrum Berlin für Materialien und Energie. (2016). The MX Beamlines BL14 .1-3 at BESSY II. *J. Large-Scale Res. Facil.* **47** (2), 1–6. <http://dx.doi.org/10.17815/jlsrf-2-64>
65. Kabsch, W., (2010). XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **66** (2), 125–132. <https://doi.org/10.1107/S0907444909047337>.
66. Sparta, K.M., Krug, M., Heinemann, U., Mueller, U., Weiss, M.S., (2016). XDSAPP2.0. *J. Appl. Crystallogr.*, **49**, 1085–1092. <https://doi.org/10.1107/S1600576716004416>.
67. Winter, G., Waterman, D.G., Parkhurst, J.M., Brewster, A. S., Gildea, R.J., Gerstel, M., Fuentes-Montero, L., Vollmar, M., Michels-Clark, T., Young, I.D., Sauter, N.K., Evans, G., (2018). DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr. Sect. D Struct. Biol.*, **74**, 85–97. <https://doi.org/10.1107/S2059798317017235>.
68. Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G. J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J. S., Terwilliger, T.C., Zwart, P.H., (2010). PHENIX: A comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **66** (2), 213–221. <https://doi.org/10.1107/S0907444909052925>.
69. Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., (2010). Features and development of coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **66** (4), 486–501. <https://doi.org/10.1107/S0907444910007493>.
70. Joosten, R.P., Long, F., Murshudov, G.N., Perrakis, A., (2014). The PDB-REDO server for macromolecular structure model optimization. *IUCrJ*, **1**, 213–220. <https://doi.org/10.1107/S2052252514009324>.
71. Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., Richardson, D.C., (2010). MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **66** (1), 12–21. <https://doi.org/10.1107/S0907444909042073>.
72. Berman, H.M., Battistuz, T., Bhat, T.N., Bluhm, W.F., Bourne, P.E., Burkhardt, K., Feng, Z., Gilliland, G.L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J.D., Zardecki, C., (2002). The protein data bank. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **58** (6 I), 899–907. <https://doi.org/10.1107/S0907444902003451>.
73. Freire, E., (1994). Statistical thermodynamic analysis of differential scanning calorimetry data: structural deconvolution of heat capacity function of proteins. *Methods Enzymol.*, **240** (C), 502–530. [https://doi.org/10.1016/S0076-6879\(94\)40062-8](https://doi.org/10.1016/S0076-6879(94)40062-8).
74. Lyubarev, A.E., Kurganov, B.I., (1998). Modeling of irreversible thermal protein denaturation at varying temperature. I. The model involving two consecutive irreversible steps. *Biochem.*, **63** (4), 434–440.
75. Sanchez-Ruiz, J.M., (1992). Theoretical analysis of lumryeyring models in differential scanning calorimetry. *Biophys. J.*, **61** (4), 921–935. [https://doi.org/10.1016/S0006-3495\(92\)81899-4](https://doi.org/10.1016/S0006-3495(92)81899-4).
76. Costas, M., Rodríguez-Larrea, D., De Maria, L., Borchert, T.V., Gómez-Puyou, A., Sanchez-Ruiz, J.M., (2009). Between-species variation in the kinetic stability of TIM proteins linked to solvation-barrier free energies. *J. Mol. Biol.*, **385** (3), 924–937. <https://doi.org/10.1016/j.jmb.2008.10.056>.
77. Santoro, M.M., Bolen, D.W., (1988). Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl a-chymotrypsin using different denaturants. *Biochemistry*, **27** (21), 8063–8068. <https://doi.org/10.1021/bi00421a014>.
78. Kumar, S., Tsai, C.J., Nussinov, R., (2002). Maximal stabilities of reversible two-state proteins. *Biochemistry*, **41** (17), 5359–5374. <https://doi.org/10.1021/bi012154c>.
79. Privalov, P.L., Gill, S.J., (1988). Stability of protein structure and hydrophobic interaction. *Adv. Protein Chem.*, **39** (C), 191–234. [https://doi.org/10.1016/S0065-3233\(08\)60377-0](https://doi.org/10.1016/S0065-3233(08)60377-0).
80. Horovitz, A., Fersht, A.R., (1990). Strategy for analysing the co-operativity of intramolecular interactions in peptides and proteins. *J. Mol. Biol.*, **214** (3), 613–617. [https://doi.org/10.1016/0022-2836\(90\)90275-Q](https://doi.org/10.1016/0022-2836(90)90275-Q).
81. Carter, P.J., Winter, G., Wilkinson, A.J., Fersht, A.R., (1984). The use of double mutants to detect structural changes in the active site of the tyrosyl-TRNA synthetase

- (*Bacillus Stearotherophilus*). *Cell*, **38** (3), 835–840. [https://doi.org/10.1016/0092-8674\(84\)90278-2](https://doi.org/10.1016/0092-8674(84)90278-2).
82. Katoh, K., Standley, D.M., (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.*, **30** (4), 772–780. <https://doi.org/10.1093/molbev/mst010>.
83. Pravda, L., Sehnal, D., Toušek, D., Navrátilová, V., Bazgier, V., Berka, K., Vařeková, R.S., Koča, J., Otyepka, M., (2018). MOLEonline: A web-based tool for analyzing channels, tunnels and pores (2018 update). *Nucleic Acids Res.*, **46** (W1), W368–W373. <https://doi.org/10.1093/nar/gky309>.
84. Willard, L., Ranjan, A., Zhang, H., Monzavi, H., Boyko, R. F., Sykes, B.D., Wishart, D.S., (2003). VADAR: A web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res.*, **31** (13), 3316–3319. <https://doi.org/10.1093/nar/gkg565>.
85. McDonald, I.K., Thornton, J.M., (1994). Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.*, **238** (5), 777–793. <https://doi.org/10.1006/jmbi.1994.1334>.
86. Costantini, S., Colonna, G., Facchiano, A.M., (2008). ESBRI: A web server for evaluating salt bridges in proteins. *Bioinformation*, **3** (3), 137–138. <https://doi.org/10.6026/97320630003137>.
87. Sobolev, V., Sorokine, A., Prilusky, J., Abola, E.E., Edelman, M., (1999). Automated analysis of interatomic contacts in proteins. *Bioinformatics*, **15** (4), 327–332. <https://doi.org/10.1093/bioinformatics/15.4.327>.
88. Ferruz, N., Schmidt, S., Höcker, B., (2021). ProteinTools: a toolkit to analyze protein structures. *Nucleic Acids Res.*, **gkab375** <https://doi.org/10.1093/nar/gkab375>.
89. Woody, R.W., (1978). Aromatic side-chain contributions to the far ultraviolet circular dichroism of peptides and proteins. *Biopolymers*, **17** (6), 1451–1467. <https://doi.org/10.1002/bip.1978.360170606>.
90. Hilser, V.J., Gómez, J., Freire, E., (1996). The enthalpy change in protein folding and binding: refinement of parameters for structure-based calculations. *Proteins Struct. Funct. Genet.*, **26** (2), 123–133. [https://doi.org/10.1002/\(SICI\)1097-0134\(199610\)26:2<123::AID-PROT2>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1097-0134(199610)26:2<123::AID-PROT2>3.0.CO;2-H).
91. Makhatadze, G., Privalov, P.L., (1992). Protein interactions with urea and guanidinium chloride: a calorimetric study. *J. Mol. Biol.*, **226** (2), 491–505. [https://doi.org/10.1016/0022-2836\(92\)90963-K](https://doi.org/10.1016/0022-2836(92)90963-K).

Supporting Information for

The stability landscape of *de novo* TIM barrels explored by a modular design approach

Sergio Romero-Romero,^{1,2} Miguel Costas,³ Daniel-Adriano Silva Manzano,^{4,5}† Sina Kordes,² Erendira Rojas-Ortega,¹ Cinthya Tapia,¹ Yasel Guerra,¹‡ Sooruban Shanmugaratnam,² Adela Rodríguez-Romero,⁶ David Baker,^{4,5,*} Birte Höcker,^{2,*} D. Alejandro Fernández-Velasco.^{1,*}

¹ Laboratorio de Físicoquímica e Ingeniería de Proteínas, Departamento de Bioquímica, Facultad de Medicina, Universidad Nacional Autónoma de México, 04510 Mexico City, Mexico.

² Department of Biochemistry, University of Bayreuth, 95447 Bayreuth, Germany.

³ Laboratorio de Biofísicoquímica, Departamento de Físicoquímica, Facultad de Química, Universidad Nacional Autónoma de México, 04510 Mexico City, Mexico.

⁴ Department of Biochemistry, University of Washington, 98195 Seattle, USA.

⁵ Institute for Protein Design, University of Washington, 98195 Seattle, USA.

⁶ Instituto de Química, Universidad Nacional Autónoma de México, 04510 Mexico City, Mexico.

* Corresponding authors. D. A. Fernández-Velasco: fdaniel@unam.mx, B. Höcker: birte.hoecker@uni-bayreuth.de, D. Baker: dabaker@uw.edu

This file includes:

- Supporting Text.
- Supporting Figures S1 to S16.
- Supporting Tables S1 to S7.

Supporting text

Exploratory characterization of first-round designs

The proteins from the first design round (DeNovoTIMs 1-10) as well as DeNovoTIM0, sTIM11, and sTIM11noCys were characterized by circular dichroism (CD) and differential scanning calorimetry (DSC) (Fig. S3). All variants presented far-UV CD spectra with a high content of regular secondary structural elements compatible with α/β proteins. For sTIM11 and sTIM11noCys, the most pronounced minimum of the spectra was observed at 222 nm whereas, for DeNovoTIM0 and DeNovoTIM1-10, it was found at 206 and 208 nm, respectively (Fig. S3 and Fig. S4A). According to the Gibbs Helmholtz equation, high T_m and ΔH values are reflected in larger areas under the stability curve, therefore, selection of the best designs was based on their thermal unfolding parameters.

Analysis of the internal core designs showed that DeNovoTIM2 and DeNovoTIM3 unfold irreversibly and uncooperatively, whereas DeNovoTIM4 unfolds cooperatively with a T_m value which is 15 degrees higher than DeNovoTIM0 (Fig. S3B). However, DeNovoTIM1 unfolds cooperatively with a further increased T_m that is 24 degrees higher than DeNovoTIM0 (Fig. S3B) and its unfolding ΔH determined by the DSC endotherm is almost twice than that of DeNovoTIM0 (Fig. S3C). Thus, DeNovoTIM1 was chosen as the best design of this group. Amongst the bottom core designs, DeNovoTIM7 did not overexpress soluble, and no significant efforts were made to solubilize it after unfolding/refolding of inclusion bodies. The thermal unfolding of DeNovoTIM5 is less cooperative and the protein unfolds with a T_m slightly higher than DeNovoTIM0. In contrast, DeNovoTIM6 has a T_m that is 45 degrees higher than DeNovoTIM0 and shows a highly cooperative unfolding transition (Fig. S3E-S3F and table 1). Hence, DeNovoTIM6 is the most successful design within this group. Finally, all top core designs (DeNovoTIM8-10) unfold cooperatively; nevertheless, DeNovoTIM8 displays the highest T_m and ΔH of this group (Fig. S3H-S3I). Therefore, DeNovoTIM8 is considered the best design among the proteins belonging to this group.

Spectroscopic characterization of DeNovoTIMs

sTIM11, sTIM11noCys, and all DeNovoTIM variants presented the characteristic far-UV CD spectra observed for α/β proteins (Fig. 2A). Nevertheless, the secondary structure content deconvoluted from the spectra showed variations in the relative amount of helices and strands, particularly for DeNovoTIM0 and DeNovoTIM1, which have the lowest helical content (table S5). For those DeNovoTIMs for which a three-dimensional structure was obtained (see below), the secondary structure content calculated from the structure and the deconvolution of CD spectra correlates adequately (table S5).

The near-UV CD spectra of DeNovoTIMs (Fig. 2B) showed a peak with fine structure between 290 and 295 nm, characteristic for tryptophan residues. A peak between 275-283 nm was also observed in the designs that contain tyrosine residues (table S3). It is known that the shape and intensities of the near-UV CD spectrum depend not only on the number and identity of aromatic residues but also on their environment and three-dimensional position within the protein core (86). The spectra observed in Fig. 2B indicate that aromatic residues in the protein are structured with significant differences in the native environment. The latter is in agreement with the intrinsic fluorescence (IF) spectra, where native λ_{max} is in the 329-345 nm range (Fig. 2C and Fig. S6). Likewise, the spectral center of mass (SCM) is between 346 and 357 nm, indicating differences in the environment of aromatic residues among native DeNovoTIMs (table S5). The fluorescence properties of DeNovoTIM0 indicate partial exposure of aromatic residues. All first-round designs show a considerable blue-shift in both λ_{max} and SCM; this trend continues in the second- and third-round designs and should be attributed to changes in both the number and the solvent exposure of Trp residues (table S3 and table S5).

Residual structure in the thermal unfolding of DeNovoTIMs

All DeNovoTIMs, except DeNovoTIM13 and DeNovoTIM14, showed reversibility i.e. samples were heated up to 125 °C and their endotherms were completely recovered in a second heating scan (Fig. S7). All reversible transitions were well fitted to the two-state model ($N \rightleftharpoons U$) (Fig. S8) and the calorimetric criterion ($\Delta H_{\text{vH}}/\Delta H$) was very close to 1, in agreement with a two-state mechanism (table 1). DSC experiments performed at different protein concentrations

(0.25 to 2.5 mg mL⁻¹) exhibited the same T_m , indicating that thermal unfolding is under thermodynamic control (Fig. S8).

The observed unfolding ΔH varies greatly, showing values both lower and higher than DeNovoTIM0, sTIM11noCys, and sTIM11 (table 1). When proteins of the same size are compared, the main reasons for finding differences in their unfolding ΔH are the number of disrupted internal interactions, as well as the hydration of groups exposed upon unfolding (87). Extreme scenarios were observed in DeNovoTIMs, e.g. DeNovoTIM6 has a $\Delta H = 124.9 \pm 1.5$ kcal mol⁻¹, which is 100 kcal mol⁻¹ more than that of DeNovoTIM0, and close to the expected value for a protein of 184 residues according to previously reported parametric equations (128.4 ± 3.5 kcal mol⁻¹; 34).

Otherwise, several DeNovoTIMs exhibit small ΔH values, particularly DeNovoTIM14 ($\Delta H = 14$ kcal mol⁻¹; Fig. S10E and table 1). There are at least two main reasons for such a small ΔH , namely either the native state is not fully folded or the unfolded state is not completely unfolded. The spectroscopic, structural, and chemical unfolding properties shown in Fig. 2 indicate that the proteins are well folded in the native state, a conclusion that is supported by the crystal structures (see below). It is known that both ΔH and T_m decrease in the presence of chemical denaturants (88). When DeNovoTIM14 was unfolded by temperature in the presence of urea, the second transition disappeared and the T_m of the observed endotherm remained almost unchanged at ~90 °C (Fig. 2F, Fig. S10B, and Fig. S10D). Unexpectedly, the unfolding enthalpy increased linearly from the value observed without denaturant to a value close to the parametric one at 6.0 M urea (Fig. S10A and Fig. S10E). This atypical behavior in DeNovoTIM14 may be explained by an increased exposure of nonpolar residues upon unfolding in the presence of urea. These results suggest that the reason for the low unfolding enthalpy in some DeNovoTIMs is likely the high content of residual structure in the unfolded state.

Likewise, it has been shown that the residual structure of the unfolded state leads to decreased ΔC_p values (35). For sTIM11, sTIM11noCys, and DeNovoTIM6, the ΔC_p were close to the estimated value from parametric equations for a protein of that size (2.6 kcal mol⁻¹ K⁻¹), whereas the other DeNovoTIMs showed a much lower ΔC_p , with DeNovoTIM0 having the lowest value (0.44 kcal mol⁻¹ K⁻¹, table 1). Accuracy in thermodynamic parameters of unfolding

has been discussed over the years and is especially crucial for ΔC_p due to the baseline shift and its small magnitude. For DeNovoTIMs, the obtention of ΔC_p was addressed with multiple determinations at varying protein concentration. The standard deviation ranged from 3 to 20 %, very similar to previous estimates for uncertainties in ΔC_p (34, 3). A decrease in unfolding ΔC_p suggests a non-fully solvated random-coil conformation with residual hydrophobic clusters in the unfolded state. Even though it was not possible to obtain the CD spectra of DeNovoTIMs that display T_m values higher than 90 °C, CD spectra of the unfolded state of the low T_m variants DeNovoTIM0, DeNovoTIM1, and DeNovoTIM8, as well as sTIM11 and sTIM11noCys, clearly showed residual structure (Fig. S4B).

Irreversible thermal unfolding of DeNovoTIM13 and DeNovoTIM14

DeNovoTIM13 shows a T_m scan-rate dependent, indicating that its thermal unfolding is under kinetic control (Fig. S9). Because the irreversible thermal unfolding of DeNovoTIM13 and DeNovoTIM14 follows an irreversible two-state mechanism (N→F), well described by a first-order rate constant (72, 73), it was possible to determine the activation energy (E_{act}) between the native and the transition state. For DeNovoTIM13, the average value obtained from the fitting of each endotherm ($E_{act} = 118.2 \pm 2.7$ kcal mol⁻¹; Fig. S9A) is within the range reported for natural proteins of similar size. This value agrees with that determined from the Arrhenius plot ($E_{act} = 118.4 \pm 2.4$ kcal mol⁻¹; Fig. S9B) and with the calculation from the effect of the scan rate on T_m ($E_{act} = 124.2 \pm 1.7$ kcal mol⁻¹; Fig. S9C).

In contrast, for DeNovoTIM14 a much lower kinetic stability was observed fitting the endotherms to the irreversible two-state model ($E_{act} = 37.5 \pm 0.6$ kcal mol⁻¹; Fig. S10A), determined from the Arrhenius plot ($E_{act} = 37.5 \pm 0.3$ kcal mol⁻¹; Fig. S10C), and extrapolating to 0 M urea ($E_{act} = 37.2 \pm 0.8$ kcal mol⁻¹; Fig. S10F). As a consequence of having very different kinetic stabilities, the estimated half-life of DeNovoTIM13 at 25 °C is 1.8×10^{10} years and about 196 days for DeNovoTIM14. Clearly, it would be interesting to determine folding/unfolding rates in future kinetic studies of the *de novo* designed TIM barrels.

Supporting figures

List of supporting figures

- **Fig. S1:** Flowchart of the Rosetta design protocol used to generate the DeNovoTIM collection.
- **Fig. S2:** Sequence alignment of DeNovoTIM barrels.
- **Fig. S3:** Exploratory characterization of first-round designs.
- **Fig. S4:** CD spectra in the peptidic region for DeNovoTIMs.
- **Fig. S5:** CD spectra in the aromatic region for DeNovoTIMs.
- **Fig. S6:** Intrinsic fluorescence spectra of DeNovoTIMs.
- **Fig. S7:** DSC instrument equilibration and thermal unfolding reversibility assessment of DeNovoTIMs.
- **Fig. S8:** DSC endotherms of DeNovoTIMs.
- **Fig. S9:** Irreversible thermal unfolding of DeNovoTIM13.
- **Fig. S10:** Irreversible thermal unfolding of DeNovoTIM14 in the presence of urea.
- **Fig. S11:** ΔC_p determination of DeNovoTIMs at different protein concentrations.
- **Fig. S12:** Chemical unfolding of DeNovoTIMs followed by CD and IF (raw data).
- **Fig. S13:** Chemical unfolding of DeNovoTIMs followed by CD and IF (normalized data).
- **Fig. S14:** Thermodynamic cycles for DeNovoTIMs.
- **Fig. S15:** Hydrophobic clusters of DeNovoTIMs.
- **Fig. S16:** Stability landscape of natural proteins.

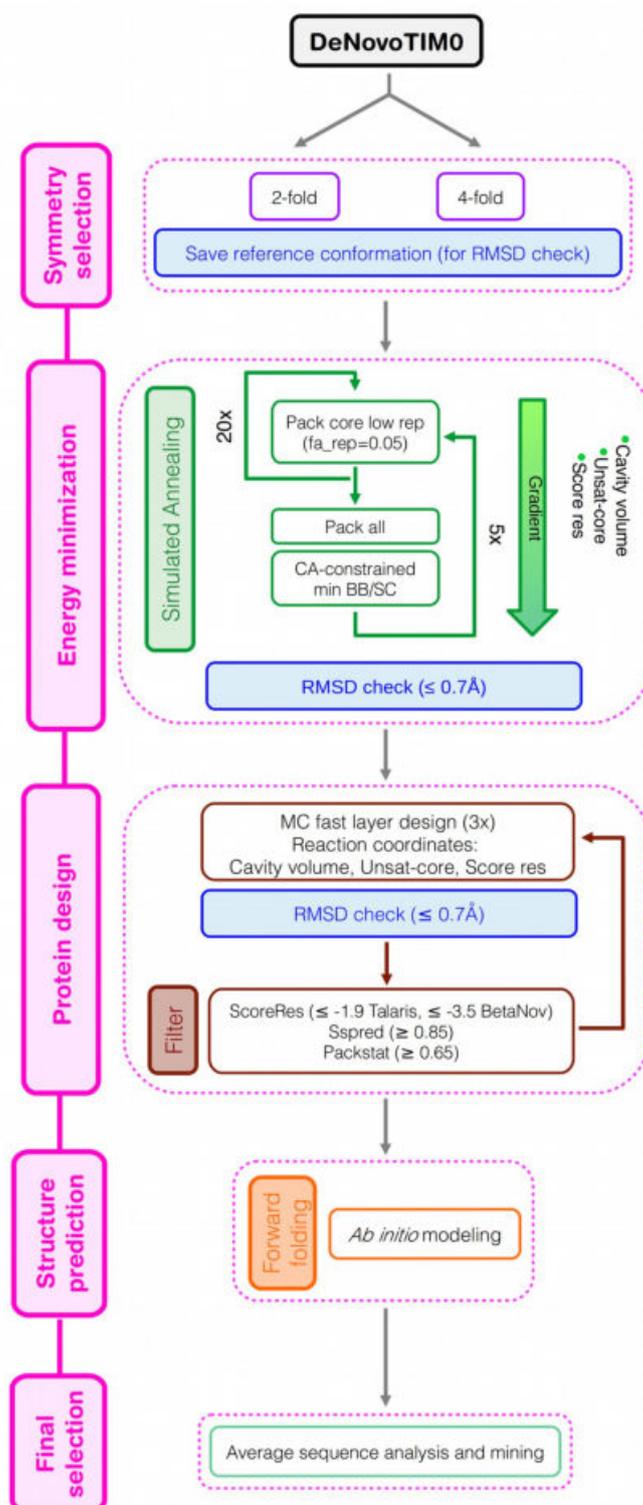


Fig. S1. Flowchart of the Rosetta design protocol used to generate the DeNovoTIM collection.

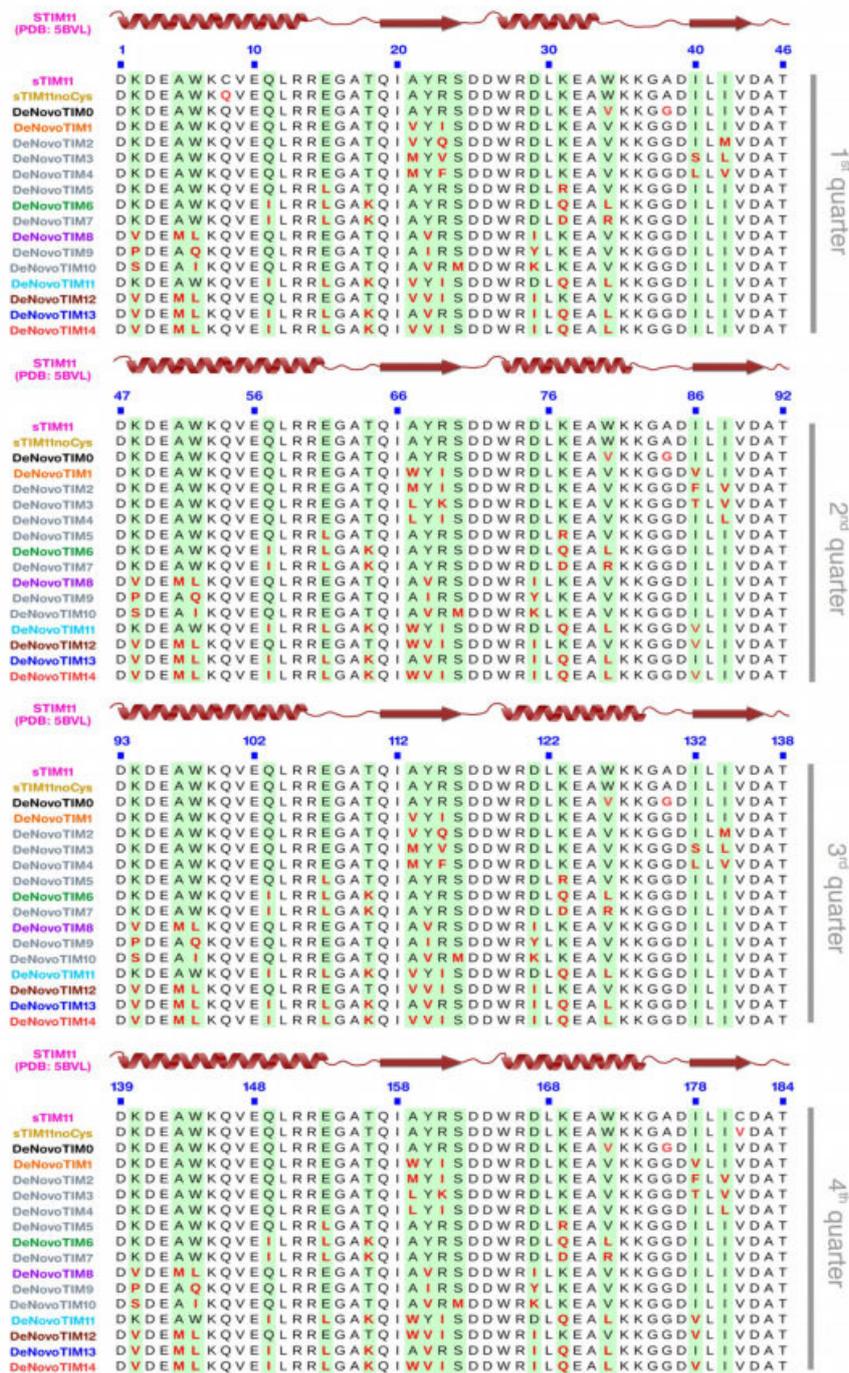


Fig. S2. Sequence alignment of DeNovoTIM barrels. On top of each TIM barrel quarter, the secondary structure of sTIM11 is shown (PDB ID: 5BVL). Mutations incorporated in each design are highlighted in red. sTIM11noCys is the symmetric version of sTIM11 removing the cysteine residues (C8Q/C181V). DeNovoTIM0 contains the additional mutations W34V and A38G, as well as their symmetry-related positions. See table S2 for a complete list of the mutations.

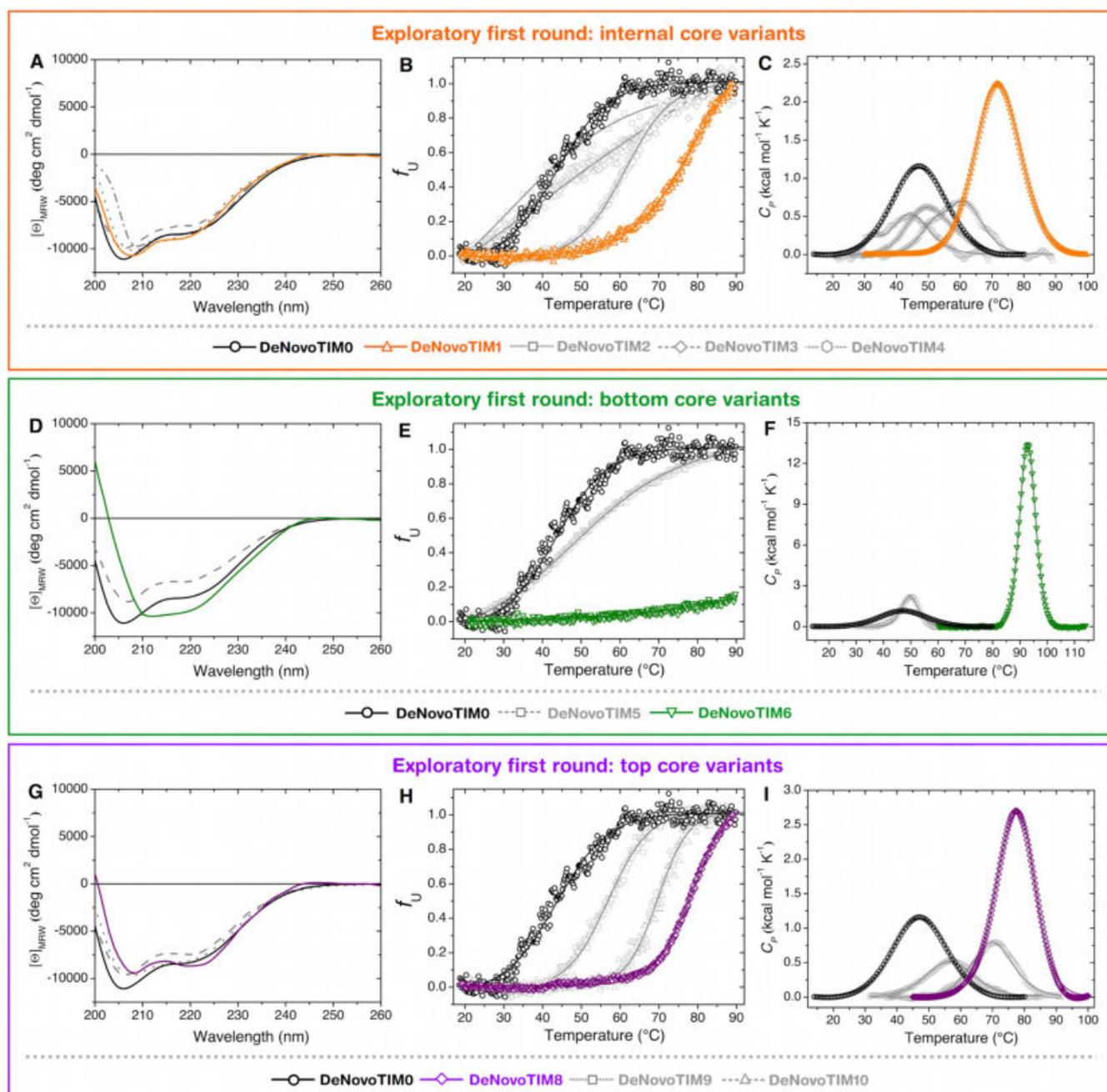


Fig. S3. Exploratory characterization of first-round designs. Far-UV CD spectra are shown in panels A, D, and G. Normalized thermal unfolding data followed by CD_{222 nm} are presented in panels B, E, and H. Thermal unfolding experiments followed by DSC are shown in panels C, F, and I. Selected variants of each design group are highlighted in orange (DeNovoTIM1), green (DeNovoTIM6), or purple (DeNovoTIM8). In all experiments protein concentration was 0.4 mg mL⁻¹ in 10 mM sodium phosphate pH 8.

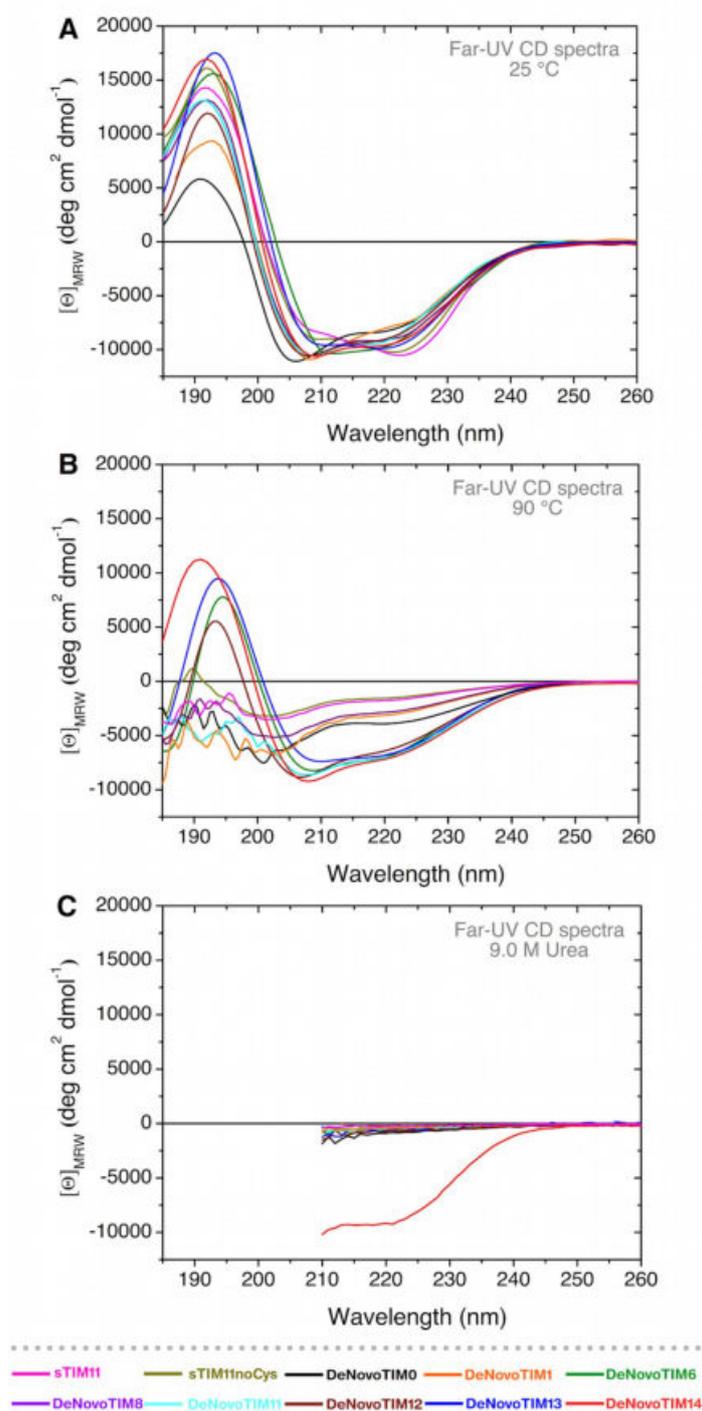


Fig. S4. CD spectra in the peptidic region for DeNovoTIMs. Far-UV CD spectra at 25 °C (A), 90 °C (B), and 9.0 M urea (C). Note that for DeNovoTIM14 at 9.0 M urea (solid red line) the spectrum is very similar to the native one (panel A). In 7.0 M GdnHCl the protein is completely unfolded and its spectrum is identical to all other DeNovoTIMs unfolded in 9.0 M urea (dotted red line). In all experiments protein concentration was 0.4 mg mL^{-1} in 10 mM sodium phosphate pH 8.

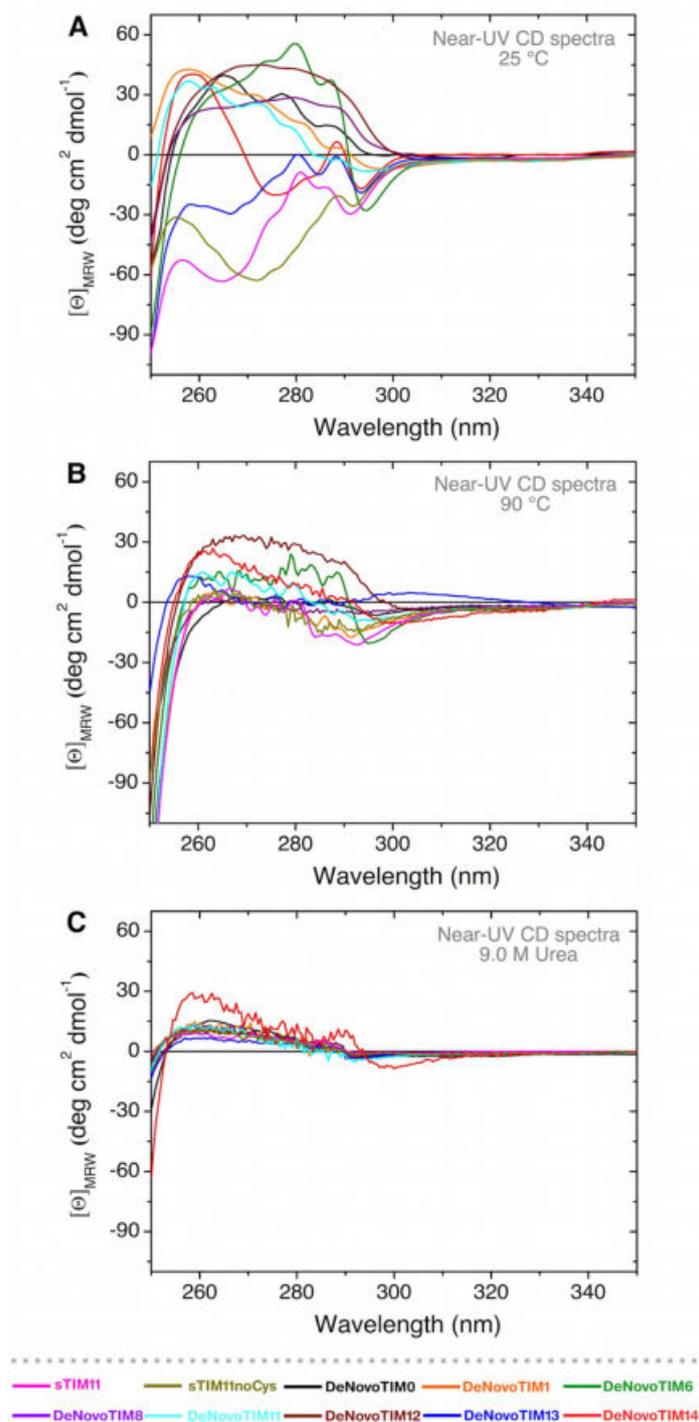


Fig. S5. CD spectra in the aromatic region for DeNovoTIMs. Near-UV CD spectra at 25 °C (A), 90 °C (B), and 9.0 M urea (C). Note that for DeNovoTIM14 at 9.0 M urea (solid red line) the spectrum is very similar to the native one (panel A). In 7.0 M GdnHCl the protein is completely unfolded and its spectrum is identical to all other DeNovoTIMs unfolded in 9.0 M urea (dotted red line). In all experiments protein concentration was 0.4 mg mL⁻¹ in 10 mM sodium phosphate pH 8.

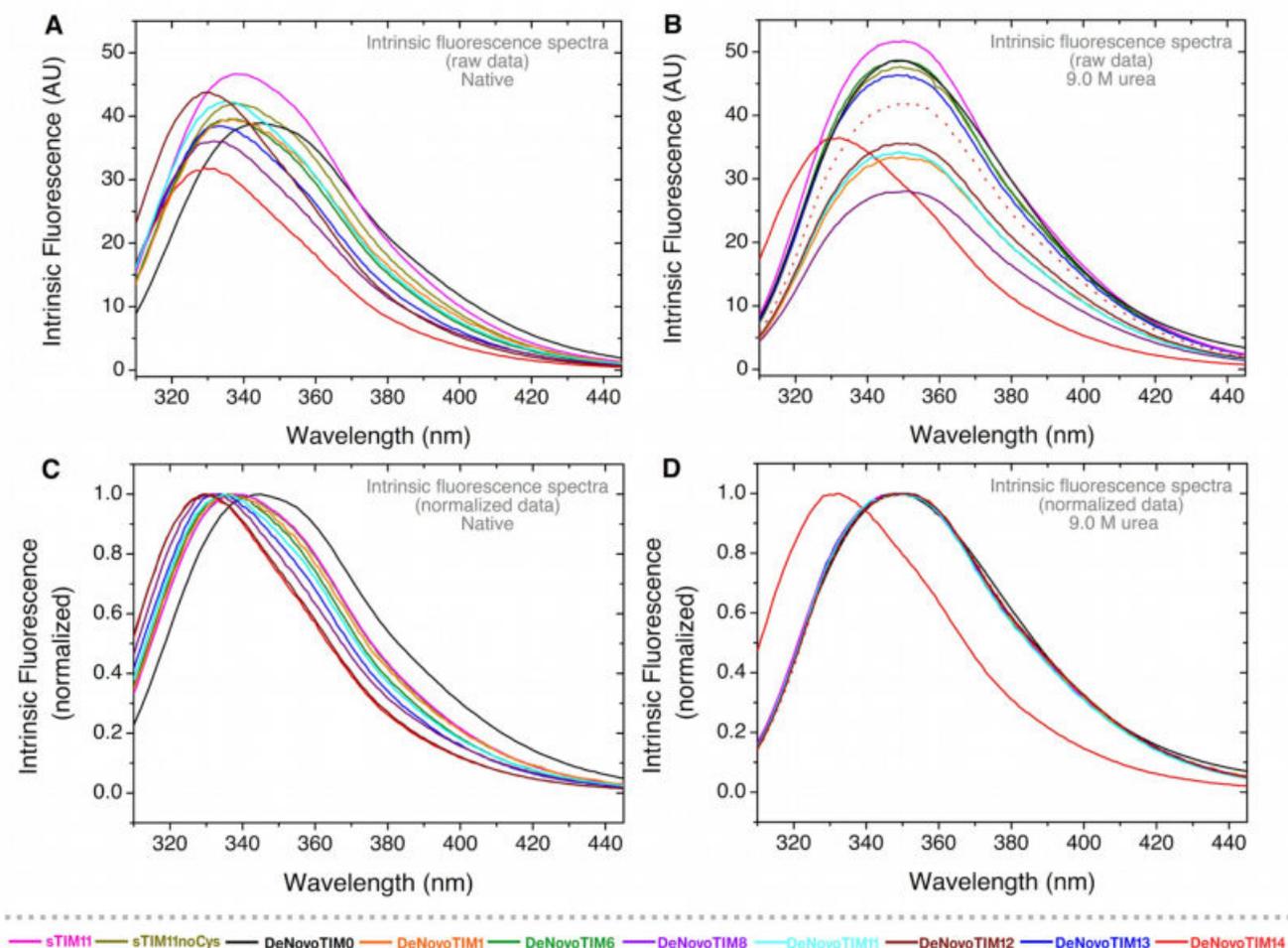


Fig. S6. Intrinsic fluorescence spectra of DeNovoTIMs. Native spectra are presented in panel **A** (raw data) and **C** (normalized data). Spectra obtained in 9.0 M urea are presented in panel **B** (raw data) and **D** (normalized data). Note that for DeNovoTIM14 in 9.0 M urea (solid red lines in panels **B** and **D**) the spectra are very similar to the native ones (solid red lines in panels **A** and **C**). In 7.0 M GdnHCl the protein is completely unfolded and its spectrum is identical to all other DeNovoTIMs unfolded in urea (dotted red line in panels **C** and **D**). In all experiments protein concentration was 0.4 mg mL^{-1} in 10 mM sodium phosphate pH 8.

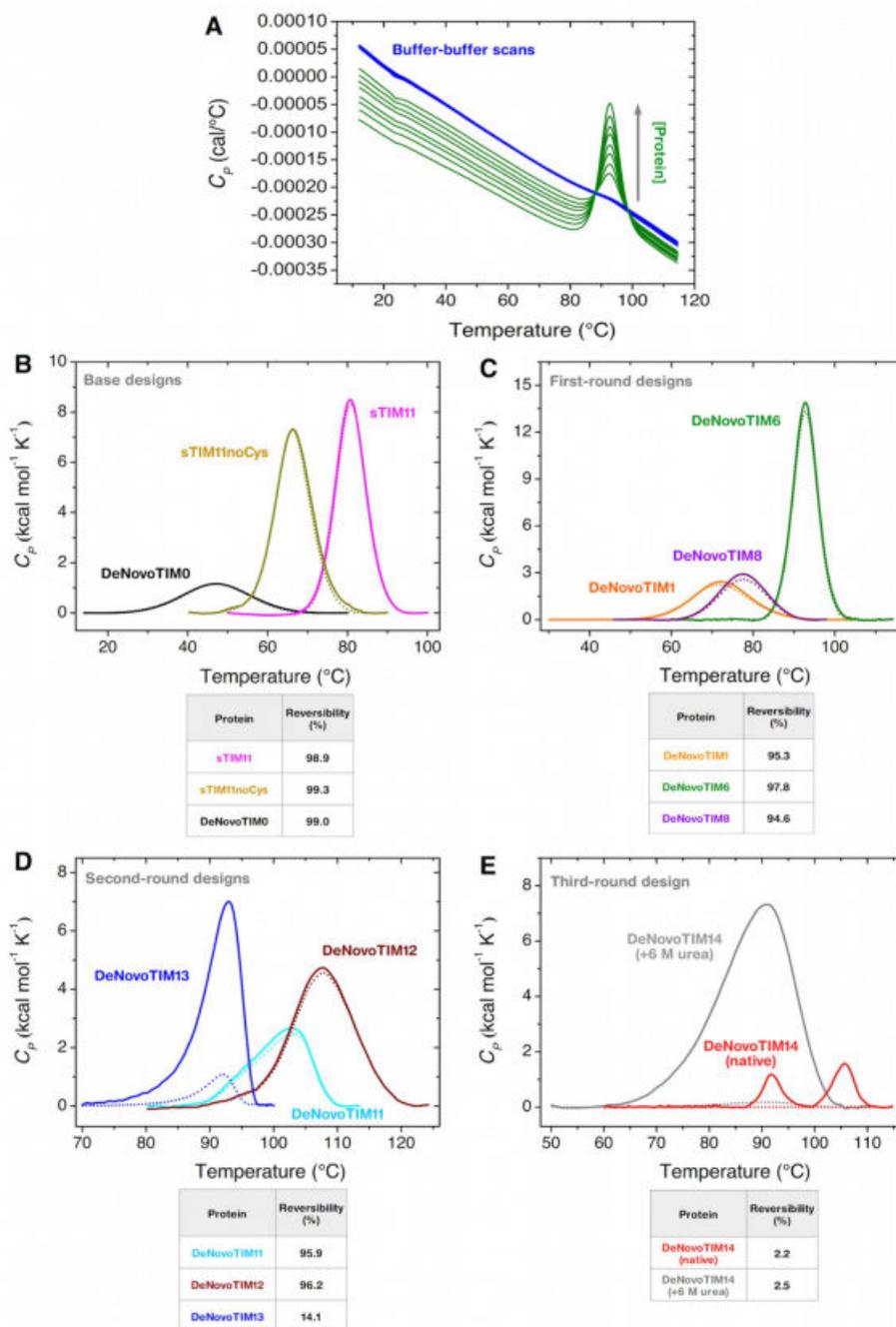


Fig. S7. DSC instrument equilibration and thermal unfolding reversibility assessment of DeNovoTIMs. A)

Proper instrument equilibration was ascertained by performing two buffer-buffer scans before each protein-buffer scan. One example at different protein concentrations (DeNovoTIM6) is shown in the panel. **B-E)** Thermal unfolding reversibility was assessed with 1.0 mg protein mL⁻¹ and 1 K min⁻¹ in 10 mM sodium phosphate pH 8. Continuous lines show the first scan and dotted lines show the second scan collected after cooling down and reheating the sample. Reversibility % was calculated as the ratio of the calorimetric ΔH (area under the curve) recovered in the second scan and that obtained in the first scan ($\Delta H_{\text{secondscan}} / \Delta H_{\text{firstscan}} \times 100$).

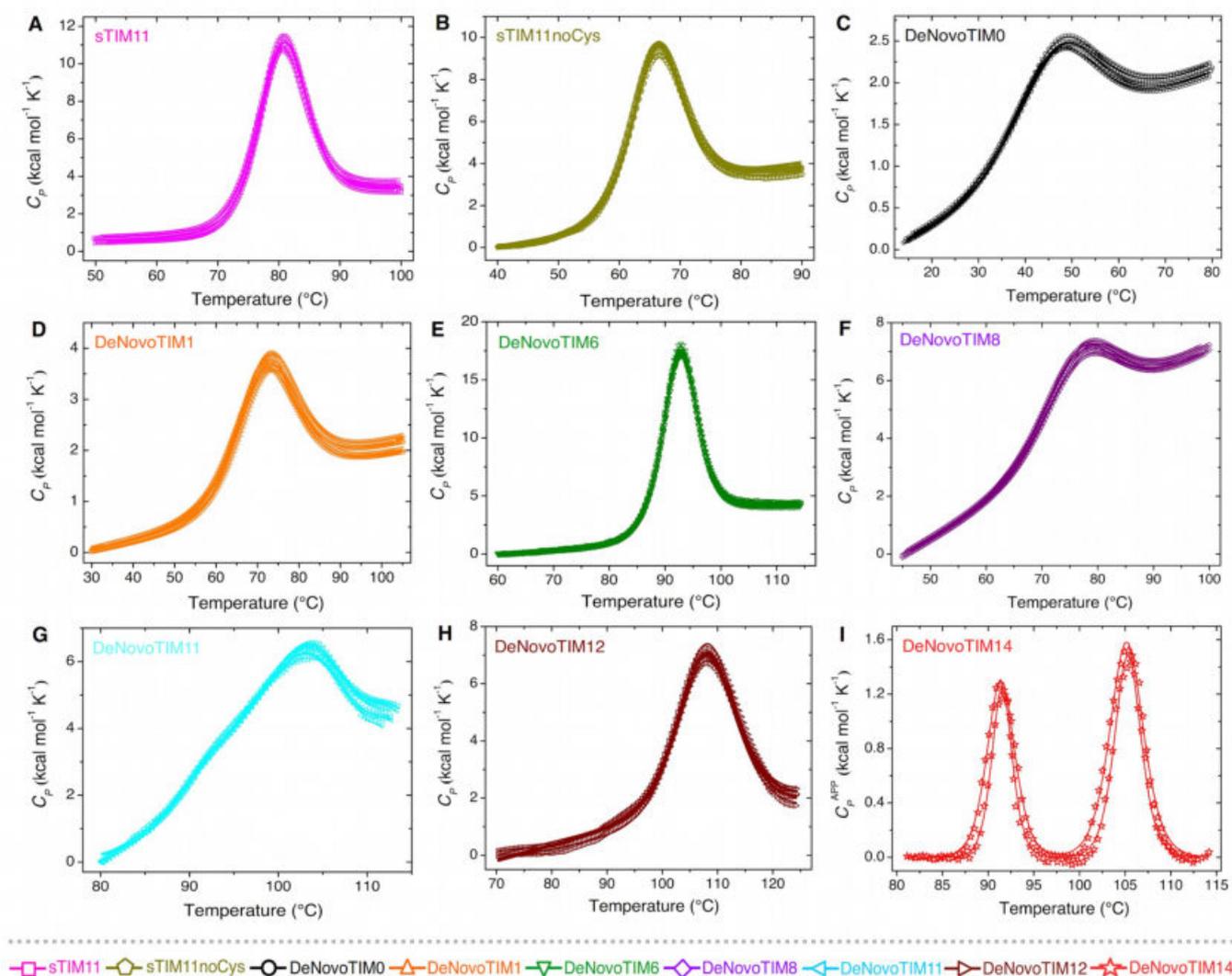


Fig. S8. DSC endotherms of DeNovoTIMs. Experiments were carried out at different protein concentrations (panels A-H: 0.25-2.5 mg mL⁻¹, panel I: 2.5 and 4.5 mg mL⁻¹). Open symbols show experimental data and solid lines are the best fits to a two-state model, except for DeNovoTIM14 where a non-two-state model with two transitions was used.

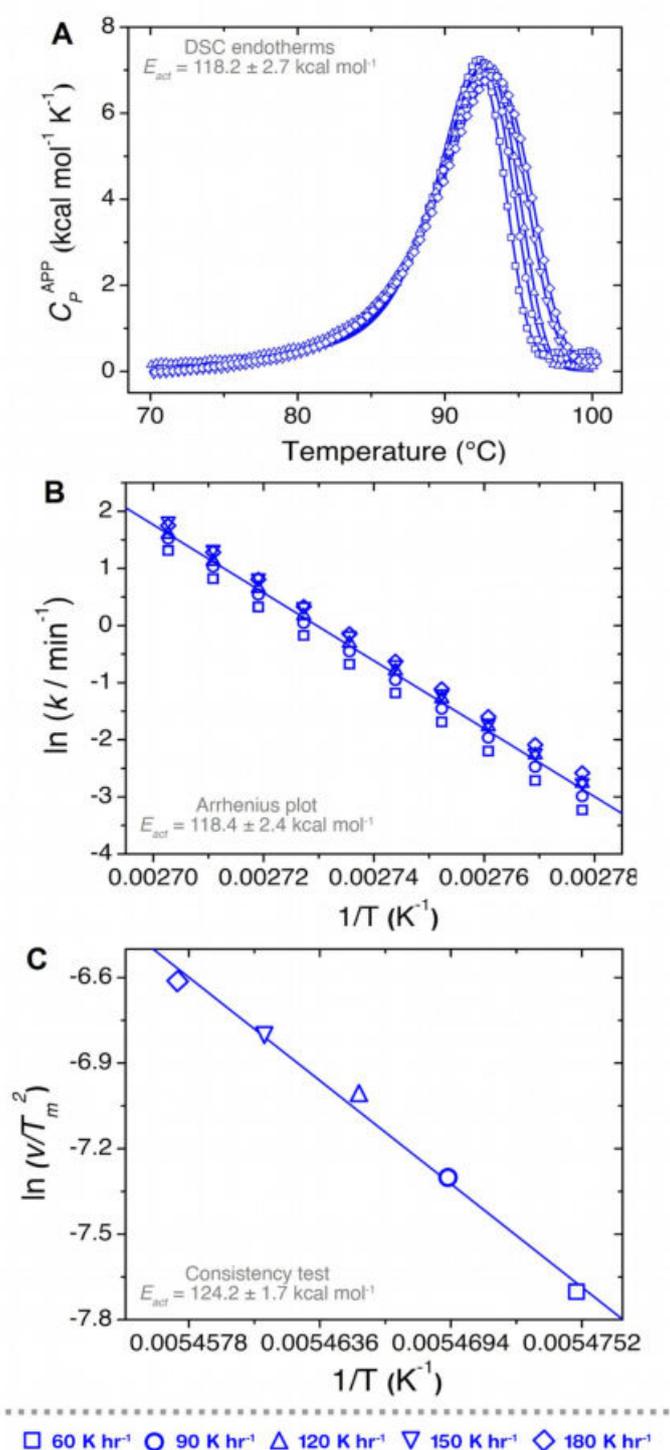


Fig. S9. Irreversible thermal unfolding of DeNovoTIM13. **A**) Endotherms at different scan rates (60 to 180 K h⁻¹). Lines represent the best fit to a two-state irreversible model. **B**) Arrhenius plot. The line shows the best fit to the Arrhenius equation (R^2 : 0.98). **C**) Effect of the scan rate on T_m (R^2 : 0.99). In all experiments protein concentration was 1.0 mg mL⁻¹ in 10 mM sodium phosphate pH 8.

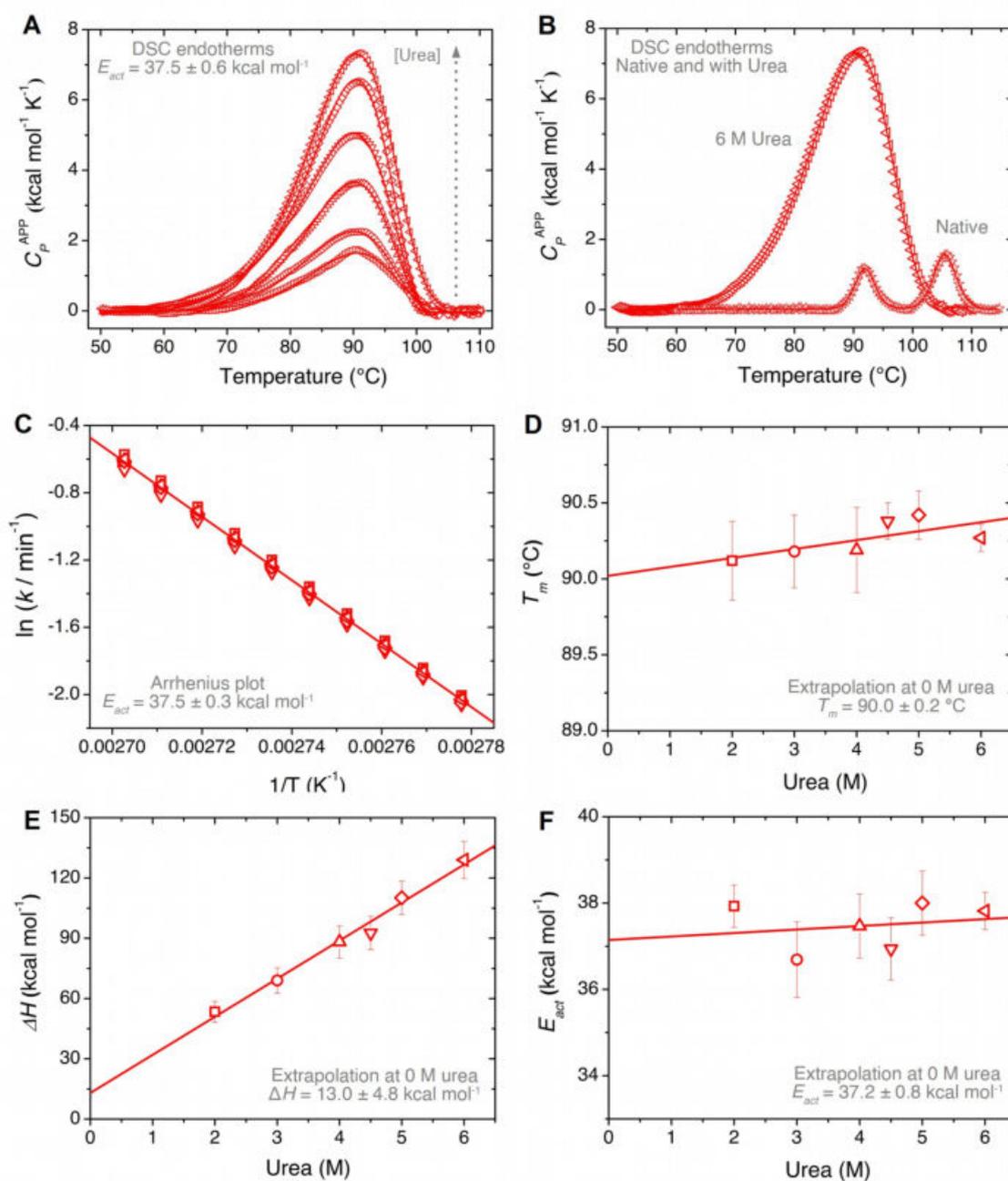


Fig. S10. Irreversible thermal unfolding of DeNovoTIM14 in the presence of urea. A) DSC endotherms at different urea concentrations (2.0 to 6.0 M). Lines represent the best fit to a two-state irreversible model. B) Comparison between DeNovoTIM14 endotherms in native conditions and at 6.0 M urea. C) Arrhenius plot. The line shows the best fit to the Arrhenius equation (R^2 : 0.99). D) T_m vs. urea concentration (R^2 : 0.70). E) ΔH vs. urea concentration (R^2 : 0.99). F) E_{act} vs. urea concentration (R^2 : 0.62). In all experiments protein concentration was 1 mg mL^{-1} in 10 mM sodium phosphate pH 8.

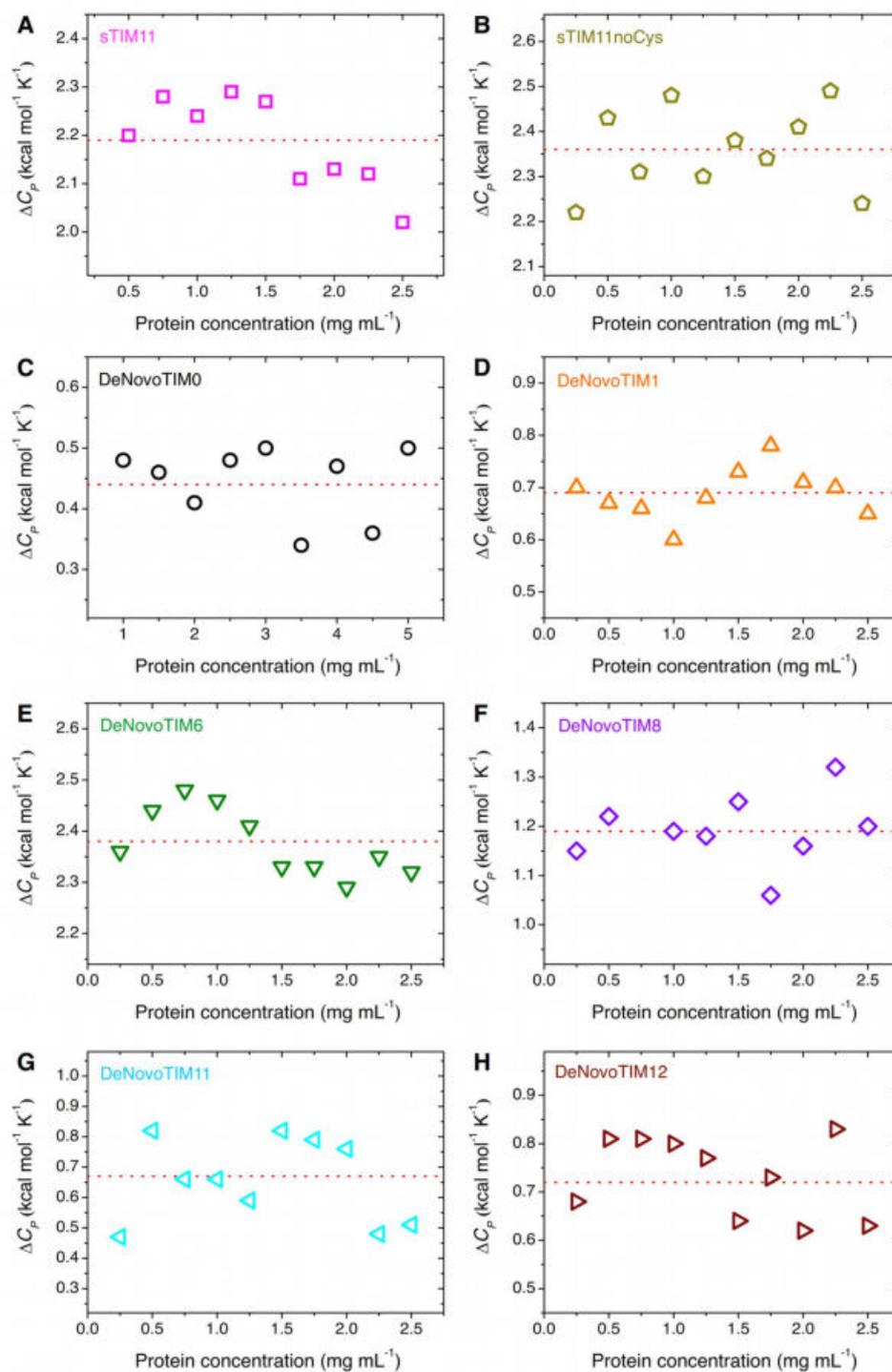


Fig. S11. ΔC_p determination of DeNovoTIMs at different protein concentrations. DSC experiments (Fig. S8) were carried out in the range of 0.25-2.5 mg mL⁻¹ (panels A, B, D, E, F, G, H) or 1.0-5.0 mg mL⁻¹ (panel C). Dotted red lines represent the average value for each protein.

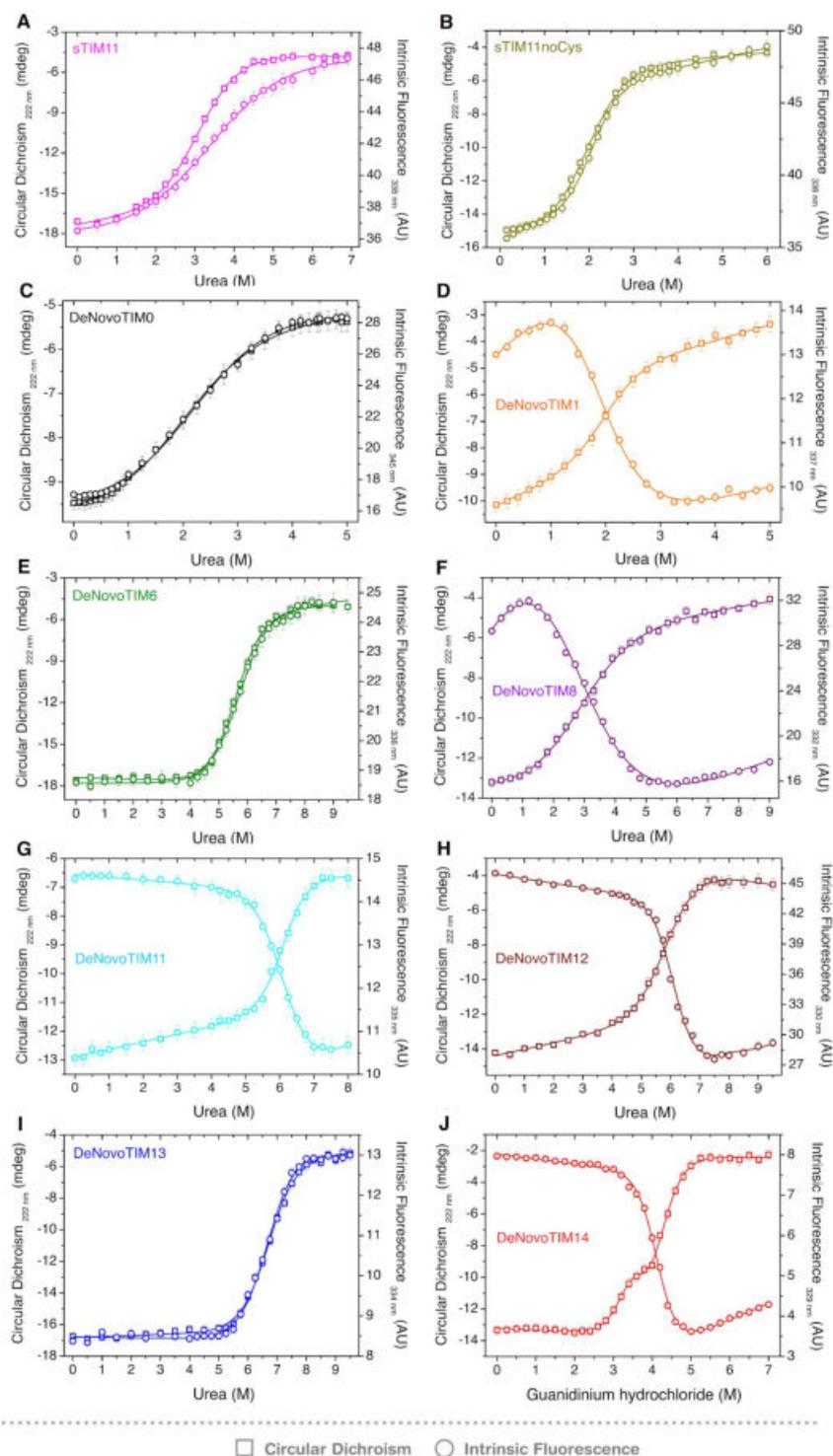


Fig. S12. Chemical unfolding of DeNovoTIMs followed by CD and IF (raw data). Lines are the best fits to a two-state model, except for DeNovoTIM14 where a three-state model was used. In all experiments protein concentration was 0.1 mg mL^{-1} in 10 mM sodium phosphate pH 8.

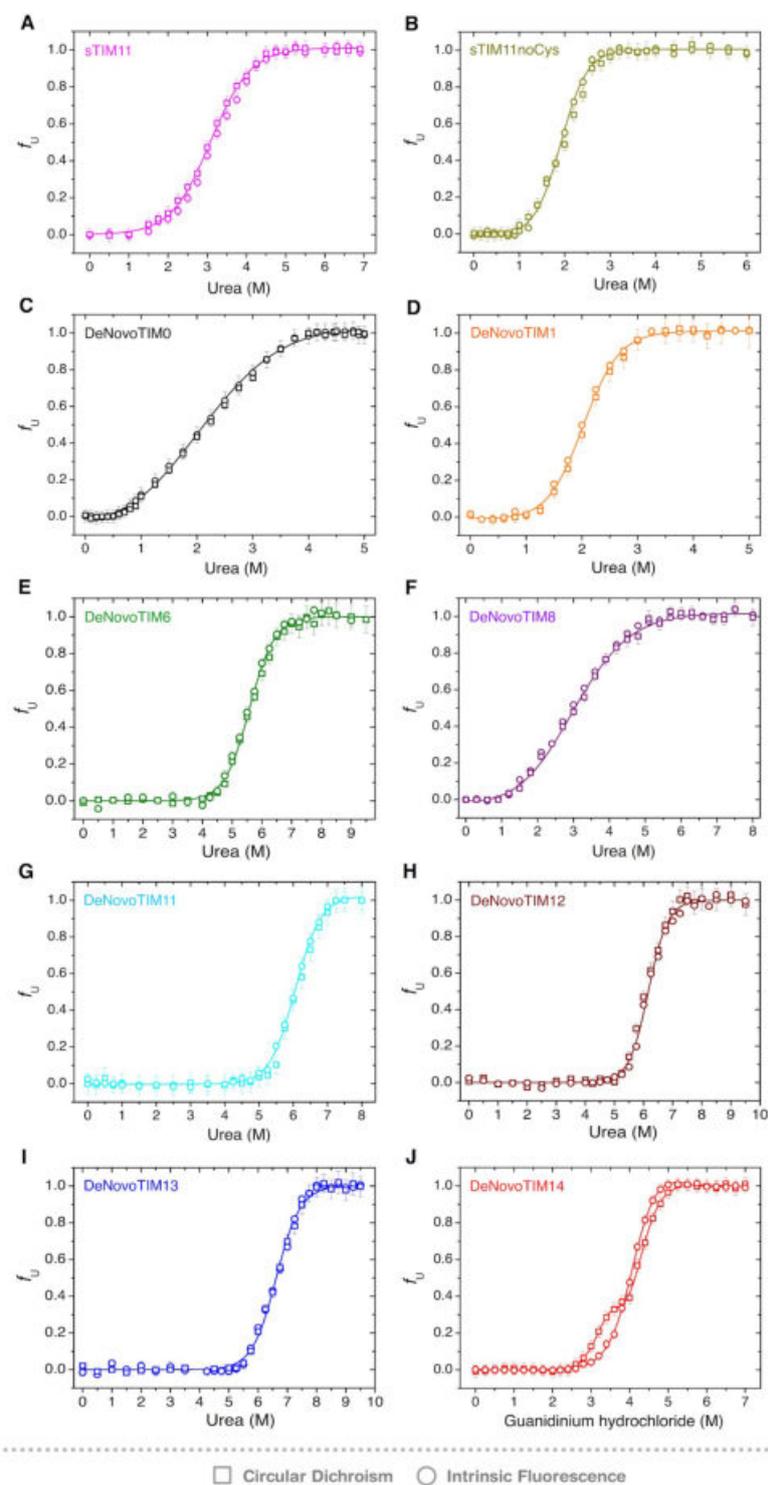


Fig. S13. Chemical unfolding of DeNovoTIMs followed by CD and IF (normalized data). Lines are the best fits to a two-state model, except for DeNovoTIM14 where a three-state model was used. In all experiments protein concentration was 0.1 mg mL^{-1} in 10 mM sodium phosphate pH 8.

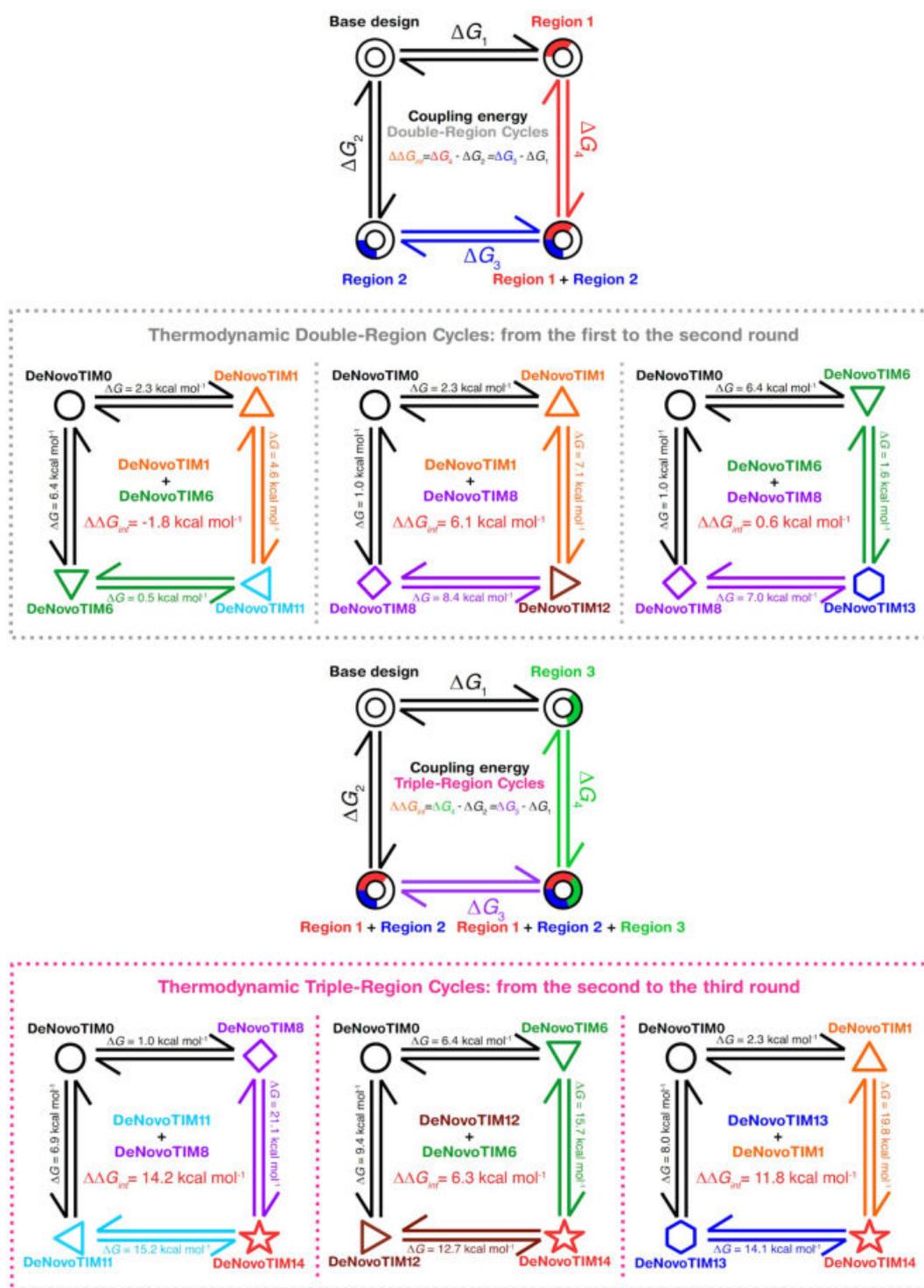


Fig. S14. Thermodynamic cycles for DeNovoTIMs. Thermodynamic cycles for double- and triple-region designs (top and bottom panels, respectively). Coupling energy ($\Delta\Delta G_{int}$) for each case is shown inside each cycle.

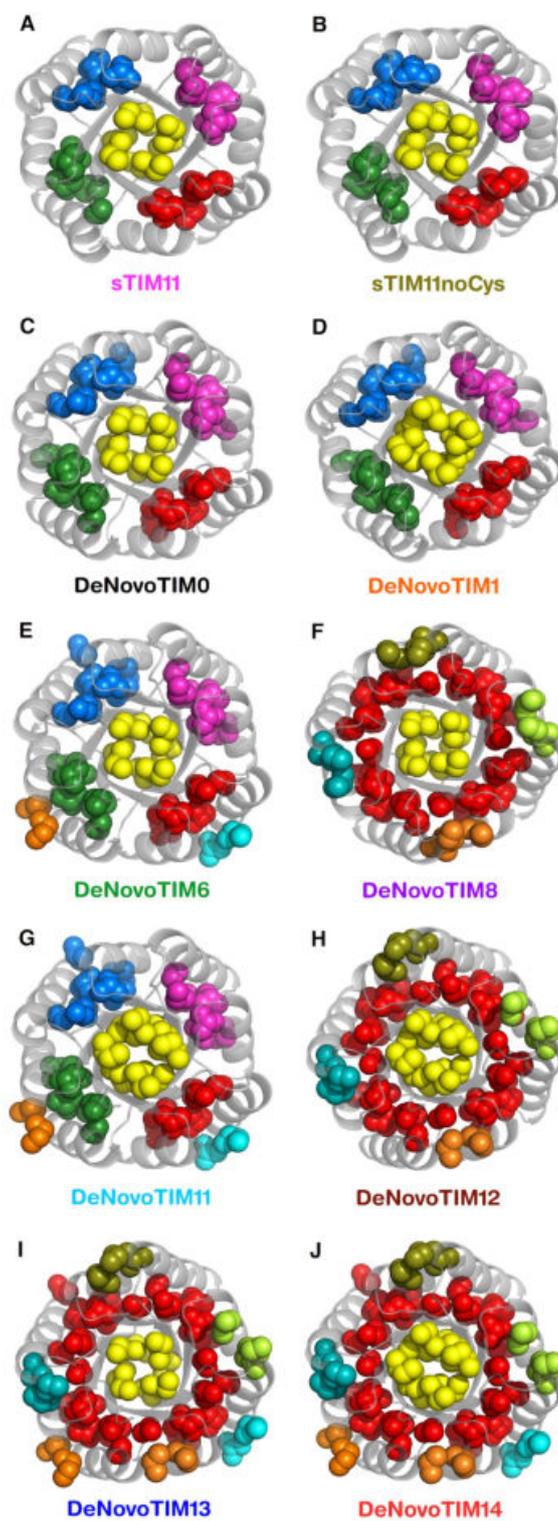


Fig. S15. Hydrophobic clusters of DeNovoTIMs. Clusters are shown in different colors. Hydrophobic clusters properties are reported in table S7.

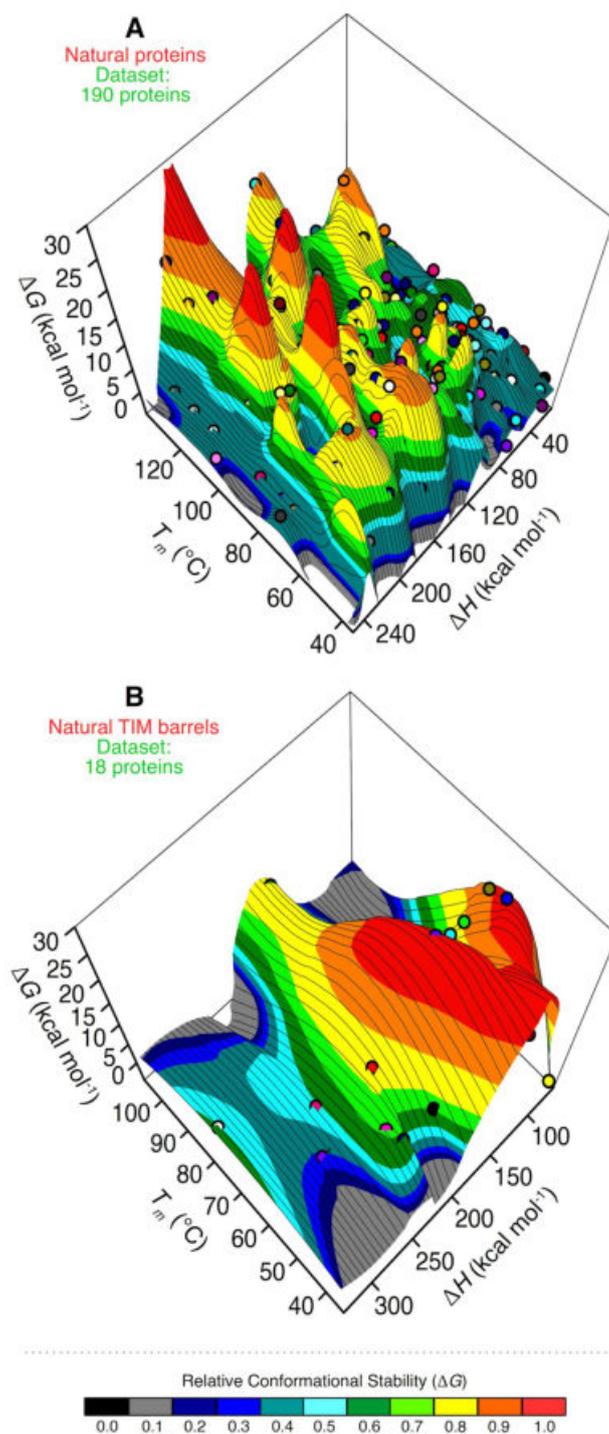


Fig. S16. Stability landscape of natural proteins. Panel **A** includes 190 natural proteins with different topologies varying in size from 42 to 572 residues (average: 173 residues). Panel **B** contains natural TIM barrels varying in size from 225 to 476 (average: 311 residues). The stability surface is colored according to normalized ΔG at 25 °C in 0.1 bins. Data for all non-redundant proteins presented here were obtained from ProThermDB database (89, 90).

Supporting tables

List of supporting tables

- **Table S1:** Amino acid sequences of DeNovoTIM collection.
- **Table S2:** List of mutations in DeNovoTIM collection.
- **Table S3:** Aromatic residues in DeNovoTIM collection.
- **Table S4:** Sequence identity matrix of DeNovoTIM collection.
- **Table S5:** Biochemical and biophysical characterization of the DeNovoTIM collection.
- **Table S6:** Crystallographic data collection and refinement statistics for sTIM11noCys and DeNovoTIM structures.
- **Table S7:** Comparison of structural features between Rosetta models and three-dimensional structures.

Table S1. Amino acid sequences of DeNovoTIM collection.

<i>de novo</i> TIM barrel	Sequence
sTIM11	DKDEAWKQVEQLRREGATQIAYRSDDWRDLKEAWKKGADILIVDATDKDEAWKQVEQLRREGATQIAYRSDDWRDLKEAWKKGADILIVDAT DKDEAWKQVEQLRREGATQIAYRSDDWRDLKEAWKKGADILIVDATDKDEAWKQVEQLRREGATQIAYRSDDWRDLKEAWKKGADILICDAT
sTIM11noCys	DKDEAWKQVEQLRREGATQIAYRSDDWRDLKEAWKKGADILIVDATDKDEAWKQVEQLRREGATQIAYRSDDWRDLKEAWKKGADILIVDAT DKDEAWKQVEQLRREGATQIAYRSDDWRDLKEAWKKGADILIVDATDKDEAWKQVEQLRREGATQIAYRSDDWRDLKEAWKKGADILIVDAT
DeNovoTIM0	DKDEAWKQVEQLRREGATQIAYRSDDWRDLKEAVKKGDDILIVDATDKDEAWKQVEQLRREGATQIAYRSDDWRDLKEAVKKGDDILIVDAT DKDEAWKQVEQLRREGATQIAYRSDDWRDLKEAVKKGDDILIVDATDKDEAWKQVEQLRREGATQIAYRSDDWRDLKEAVKKGDDILIVDAT
DeNovoTIM1	DKDEAWKQVEQLRREGATQIYISDDWRDLKEAVKKGDDILIVDATDKDEAWKQVEQLRREGATQIYISDDWRDLKEAVKKGDDILIVDAT DKDEAWKQVEQLRREGATQIYISDDWRDLKEAVKKGDDILIVDATDKDEAWKQVEQLRREGATQIYISDDWRDLKEAVKKGDDILIVDAT
DeNovoTIM2	DKDEAWKQVEQLRREGATQIVYQSDWRDLKEAVKKGDDILMVDATDKDEAWKQVEQLRREGATQIMYISDDWRDLKEAVKKGDDFLVVDAT DKDEAWKQVEQLRREGATQIVYQSDWRDLKEAVKKGDDILMVDATDKDEAWKQVEQLRREGATQIMYISDDWRDLKEAVKKGDDFLVVDAT
DeNovoTIM3	DKDEAWKQVEQLRREGATQIMYVSDDWRDLKEAVKKGDDLLVVDATDKDEAWKQVEQLRREGATQILYKSDWRDLKEAVKKGDDTLVVDAT DKDEAWKQVEQLRREGATQIMYVSDDWRDLKEAVKKGDDLLVVDATDKDEAWKQVEQLRREGATQILYKSDWRDLKEAVKKGDDTLVVDAT
DeNovoTIM4	DKDEAWKQVEQLRREGATQIMYFSDWRDLKEAVKKGDDLLVVDATDKDEAWKQVEQLRREGATQILYISDDWRDLKEAVKKGDDILVVDAT DKDEAWKQVEQLRREGATQIMYFSDWRDLKEAVKKGDDLLVVDATDKDEAWKQVEQLRREGATQILYISDDWRDLKEAVKKGDDILVVDAT
DeNovoTIM5	DKDEAWKQVEQLRRLGATQIAYRSDDWRDLREAVKKGDDILIVDATDKDEAWKQVEQLRRLGATQIAYRSDDWRDLREAVKKGDDILIVDAT DKDEAWKQVEQLRRLGATQIAYRSDDWRDLREAVKKGDDILIVDATDKDEAWKQVEQLRRLGATQIAYRSDDWRDLREAVKKGDDILIVDAT
DeNovoTIM6	DKDEAWKQVEILRRLGAKQIAYRSDDWRDLQEALKKGGDILIVDATDKDEAWKQVEILRRLGAKQIAYRSDDWRDLQEALKKGGDILIVDAT DKDEAWKQVEILRRLGAKQIAYRSDDWRDLQEALKKGGDILIVDATDKDEAWKQVEILRRLGAKQIAYRSDDWRDLQEALKKGGDILIVDAT
DeNovoTIM7	DKDEAWKQVEILRRLGAKQIAYRSDDWRDLDEARKKGGDILIVDATDKDEAWKQVEILRRLGAKQIAYRSDDWRDLDEARKKGGDILIVDAT DKDEAWKQVEILRRLGAKQIAYRSDDWRDLDEARKKGGDILIVDATDKDEAWKQVEILRRLGAKQIAYRSDDWRDLDEARKKGGDILIVDAT
DeNovoTIM8	DVDEMLKQVEQLRREGATQIAVRSDDWRILKEAVKKGDDILIVDATDVDEMLKQVEQLRREGATQIAVRSDDWRILKEAVKKGDDILIVDAT DVDEMLKQVEQLRREGATQIAVRSDDWRILKEAVKKGDDILIVDATDVDEMLKQVEQLRREGATQIAVRSDDWRILKEAVKKGDDILIVDAT
DeNovoTIM9	DPDEAQKQVEQLRREGATQIAIRSDDWRYLKEAVKKGDDILIVDATDPDEAQKQVEQLRREGATQIAIRSDDWRYLKEAVKKGDDILIVDAT DPDEAQKQVEQLRREGATQIAIRSDDWRYLKEAVKKGDDILIVDATDPDEAQKQVEQLRREGATQIAIRSDDWRYLKEAVKKGDDILIVDAT
DeNovoTIM10	DSDEAIKQVEQLRREGATQIAVRMDDWRKLKEAVKKGDDILIVDATDSDEAIKQVEQLRREGATQIAVRMDDWRKLKEAVKKGDDILIVDAT DSDEAIKQVEQLRREGATQIAVRMDDWRKLKEAVKKGDDILIVDATDSDEAIKQVEQLRREGATQIAVRMDDWRKLKEAVKKGDDILIVDAT
DeNovoTIM11	DKDEAWKQVEILRRLGAKQIYISDDWRDLQEALKKGGDILIVDATDKDEAWKQVEILRRLGAKQIYISDDWRDLQEALKKGGDVLIVDAT DKDEAWKQVEILRRLGAKQIYISDDWRDLQEALKKGGDILIVDATDKDEAWKQVEILRRLGAKQIYISDDWRDLQEALKKGGDVLIVDAT
DeNovoTIM12	DVDEMLKQVEQLRREGATQIVVISDDWRILKEAVKKGDDILIVDATDVDEMLKQVEQLRREGATQIVVISDDWRILKEAVKKGDDVLIVDAT DVDEMLKQVEQLRREGATQIVVISDDWRILKEAVKKGDDILIVDATDVDEMLKQVEQLRREGATQIVVISDDWRILKEAVKKGDDVLIVDAT
DeNovoTIM13	DVDEMLKQVEILRRLGAKQIAVRSDDWRILQEALKKGGDILIVDATDVDEMLKQVEILRRLGAKQIAVRSDDWRILQEALKKGGDILIVDAT DVDEMLKQVEILRRLGAKQIAVRSDDWRILQEALKKGGDILIVDATDVDEMLKQVEILRRLGAKQIAVRSDDWRILQEALKKGGDILIVDAT
DeNovoTIM14	DVDEMLKQVEILRRLGAKQIVVISDDWRILQEALKKGGDILIVDATDVDEMLKQVEILRRLGAKQIVVISDDWRILQEALKKGGDVLIVDAT DVDEMLKQVEILRRLGAKQIVVISDDWRILQEALKKGGDILIVDATDVDEMLKQVEILRRLGAKQIVVISDDWRILQEALKKGGDVLIVDAT

Table S2. List of mutations in DeNovoTIM collection.

<i>de novo</i> TIM barrel	Symmetry (fold)	Number of mutations in each modular unit ^a	Total number of mutations in the Sequence	Mutations added in each modular unit ^a
sTIM11	pseudo 4	-----	-----	-----
sTIM11noCys	4	2 (non-symmetric)	2	C8Q, C181V
DeNovoTIM0	4	2	8	W34V, A38G
DeNovoTIM1	2	5	10	A21V, R23I, A67W, R69I, I86V
DeNovoTIM2	2	7	14	A21V, R23Q, I42M, A67M, R69I, I86F, I88V
DeNovoTIM3	2	8	16	A21M, R23V, I40S, I42L, A67L, R69K, I86T, I88V
DeNovoTIM4	2	7	14	A21M, R23F, I40L, I42V, A67L, R69I, I88L
DeNovoTIM5	4	2	8	E15L, K31R
DeNovoTIM6	4	5	20	Q11I, E15L, T18K, K31Q, V34L
DeNovoTIM7	4	5	20	Q11I, E15L, T18K, K31D, V34R
DeNovoTIM8	4	5	20	K2V, A5M, W6L, Y22V, D29I
DeNovoTIM9	4	4	16	K2P, W6Q, Y22I, D29Y
DeNovoTIM10	4	5	20	K2S, W6I, Y22V, S24M, D29K
DeNovoTIM11	2	15	30	Q11I, E15L, T18K, A21V, R23I, K31Q, V34L, Q57I, E61L, A67W, T64K, R69I, K77Q, V80L, I86V
DeNovoTIM12	2	15	30	K2V, A5M, W6L, A21V, Y22V, R23I, D29I, K48V, A51M, W52L, A67W, Y68V, R69I, D75I, I86V
DeNovoTIM13	4	10	40	K2V, A5M, W6L, Q11I, E15L, T18K, Y22V, D29I, K31Q, V34L
DeNovoTIM14	2	25	50	K2V, A5M, W6L, Q11I, E15L, T18K, A21V, Y22V, R23I, D29I, K31Q, V34L, K48V, A51M, W52L, Q57I, E61L, T64K, A67W, Y68V, R69I, D75I, K77Q, V80L, I86V

^a Only the mutations incorporated in each design are shown. See Fig. S2 for the complete sequence alignment.

Table S3. Aromatic residues in DeNovoTIM collection.

<i>de novo</i> TIM barrel	Number of Tyr residues in each modular unit	Total number of Trp residues	Position of Trp residues in the modular unit	Number of Tyr residues in each modular unit	Total number of Tyr residues	Position of Tyr residues in the modular unit
sTIM11	3	12	W6, W27, W34	1	4	Y22
sTIM11noCys	3	12	W6, W27, W34	1	4	Y22
DeNovoTIM0	2	8	W6, W27	1	4	Y22
DeNovoTIM1	5	10	W6, W27, W52, W67, W73	2	4	Y22, Y68
DeNovoTIM2	4	8	W6, W27, W52, W73	2	4	Y22, Y68
DeNovoTIM3	4	8	W6, W27, W52, W74	2	4	Y22, Y68
DeNovoTIM4	4	8	W6, W27, W52, W75	2	4	Y22, Y68
DeNovoTIM5	2	8	W6, W27	1	4	Y22
DeNovoTIM6	2	8	W6, W27	1	4	Y22
DeNovoTIM7	2	8	W6, W27	1	4	Y22
DeNovoTIM8	1	4	W27	0	0	----
DeNovoTIM9	1	4	W27	0	0	----
DeNovoTIM10	1	4	W27	0	0	----
DeNovoTIM11	5	10	W6, W27, W52, W67, W73	2	4	Y22, Y68
DeNovoTIM12	5	10	W6, W27, W52, W67, W73	0	0	----
DeNovoTIM13	2	8	W6, W27	0	0	----
DeNovoTIM14	5	10	W6, W27, W52, W67, W73	0	0	----

Table S5. Biochemical and biophysical characterization of DeNovoTIM collection.

de novo TIM barrel	Expression/Solubility properties		Spectroscopic properties									Hydrodynamic properties ^e				
			Circular Dichroism			Intrinsic Fluorescence ^d										
	Soluble overexpression ^a	Aggregation-prone ^b	Predicted secondary structure content (%) ^c			Native λ_{max} (nm)	Unfolded λ_{max} (nm)	$\Delta\lambda_{max\ N-U}$ (nm)	Native SCM (nm)	Unfolded SCM (nm)	ΔSCM_{N-U} (nm)	Theoretical MW (Da)	Experimental MW (Da)	Exp./Theor. ratio	Oligomeric state	Stokes radius (Å)
α -helix			β -strand	Random coil												
sTIM11	+++	+	46.3 (46.0)	19.8 (20.0)	34.0 (34.0)	338	350	12	354	360	7	22854	22017 ± 465	1.0	Monomer	23.2 ± 0.5
sTIM11noCys	+++	+	47.6 (48.1)	19.2 (22.8)	33.2 (29.1)	338	349	11	353	361	7	22875	20670 ± 146	0.9	Monomer	22.6 ± 0.1
DeNovoTIM0	++	+	41.4	30.7	27.9	345	352	7	357	362	5	22471	29827 ± 309	1.3	Monomer	26.1 ± 0.3
DeNovoTIM1	++	++	42.7	29.3	27.9	337	350	13	354	360	6	22557	21017 ± 637	0.9	Monomer	22.8 ± 0.8
DeNovoTIM6	+++	+	46.5 (43.4)	18.8 (20.7)	34.8 (35.9)	336	350	14	352	361	9	22568	22000 ± 440	1.0	Monomer	23.2 ± 0.5
DeNovoTIM8	++	++	47.1	22.3	30.6	332	351	19	348	361	13	22039	20518 ± 360	0.9	Monomer	22.6 ± 0.3
DeNovoTIM11	+	+++	46.9	22.9	30.0	335	349	14	353	362	9	22598	20920 ± 572	0.9	Monomer	22.3 ± 0.6
DeNovoTIM12	+++	+	45.5	24.7	29.7	330	351	21	347	360	13	22125	20029 ± 760	0.9	Monomer	23.1 ± 0.9
DeNovoTIM13	+++	+++	48.3 (48.8)	18.1 (21.8)	33.5 (29.4)	334	351	17	351	360	8	22080	21601 ± 725	1.0	Monomer	22.7 ± 0.7
DeNovoTIM14	+	+++	50.3	18.7	31.1	329	350 (331)	21	346	362 (349)	15	22166	25357 ± 1833	1.2	Monomer	24.7 ± 1.8

^a Symbols represent the amount of protein purified from the soluble fraction: +++ >10 mg L⁻¹ culture, ++ 5-10 mg L⁻¹ culture, + 2-5 mg L⁻¹ culture.

^b Symbols represent the protein propensity to aggregate: +++ aggregation visible within the first week after purification, ++ aggregation visible between 1-2 weeks after purification, + not visible aggregation after at least 1 month after purification.

^c numbers within brackets are the calculated value from the three-dimensional structure.

^d unfolded state represents the spectra at 9.0 M urea.

^e ± indicate the standard deviation calculated from 4 different experiments at protein concentrations of 0.01, 0.5, 1.0, and 2.0 mg mL⁻¹.

Table S6. Crystallographic data collection and refinement statistics for sTIM11noCys and DeNovoTIM structures.

<i>de novo</i> TIM barrel	sTIM11noCys	DeNovoTIM6	DeNovoTIM13
Data collection^a			
PDB ID	6YQY	6Z2I	6YQX
Crystallization condition	0.2 M Ammonium sulfate 0.1 M Trisodium citrate pH: 5.6 25% w/v PEG 4000	0.095 M Sodium citrate pH: 5.0 19% v/v Isopropanol 25% w/v PEG 4000 5% v/v Glycerol	0.17 M Sodium acetate 0.085 M TRIS pH: 8.5 25.5% w/v PEG 4000 15% v/v Glycerol
Resolution range (Å)	47.15 – 1.88 (1.94 – 1.88)	30.93 – 2.90 (3.01 – 2.90)	44.79 – 1.64 (1.70 – 1.64)
Space group	P 4 ₁ 2 ₁ 2 (92)	P 4 (75)	P 3 ₂ (145)
Unit cell dimensions			
a, b, c, (Å)	50.45, 50.45, 132.41	87.48, 87.48, 44.26	51.72, 51.72, 63.97
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 120
Total reflections	342791 (32681)	56016 (3033)	79664 (7327)
Unique reflections	14703 (1410)	7620 (389)	23407 (2280)
Multiplicity	23.3 (23.2)	7.4 (7.8)	3.4 (3.2)
Completeness (%)	99.9 (99.0)	99.8 (100.0)	99.5 (97.6)
Mean I / sigma (I)	27.0 (1.9)	4.3 (1.0)	14.3 (1.5)
R-merge	0.061 (1.281)	0.180 (1.195)	0.042 (0.552)
CC1/2	1.000 (0.810)	0.940 (0.773)	0.998 (0.659)
Average mosaicity (deg)	0.109	0.179	0.116
Refinement statistics^a			
R _{work} + R _{free}	0.220 (0.287), 0.271 (0.371)	0.316 (0.368), 0.372 (0.431)	0.171 (0.274), 0.201 (0.329)
Average B-value (Å ²)	54.2	89.3	39.1
Protein	53.9	89.4	37.7
Ligand	-----	-----	61.3
Solvent	62.2	77.1	53.5
Overall B factor from Wilson Plot (Å ²)	41.0	73.8	28.4
Number of atoms			
Protein	1489	1185	1603
Ligand	0	0	7
Water	49	10	132
Protein residues	178	168	181
RMS (bonds) (Å)	0.008	0.003	0.006
RMS (angles) (°)	0.900	0.560	0.770
Ramachandran favored (%)	97.1	93.6	100.0
Ramachandran allowed (%)	2.9	6.4	0.0
Ramachandran outliers (%)	0.0	0.0	0.0
Rotamer outliers (%)	0.0	0.0	0.0
Clashscore	8.2	5.3	3.9

^a Statistics for the highest resolution shell are shown in parentheses.

Table S7. Comparison of structural features between Rosetta models and three-dimensional structures.

Design round	Base designs			First-round designs: Internal Core				First-round designs: Bottom Core			First-round designs: Top Core			Second-round designs			Third-round design				
	sTIM1		sTIMnoCys		DeNovoTIM0	DeNovoTIM1	DeNovoTIM2	DeNovoTIM3	DeNovoTIM4	DeNovoTIM5	DeNovoTIM6	DeNovoTIM7	DeNovoTIM8	DeNovoTIM9	DeNovoTIM10	DeNovoTIM11	DeNovoTIM12	DeNovoTIM13	DeNovoTIM14		
Property / <i>de novo</i> TIM barrel	model	structure	model	structure	model	model	model	model	model	model	model	model	model	model	model	model	model	model	model		
	Structural properties																				
H-bonds	236	204	194	171	198	201	197	201	201	209	176	138	189	201	184	177	169	174	167	195	163
Salt bridges	47	22	37	16	27	39	39	41	39	47	27	4	37	35	25	27	34	14	22	18	28
Total area in hydrophobic clusters (Å ²)	4088	3765	4260	4215	4793	6530	4800	4774	6042	5221	5195	5097	4533	6849	5812	5507	4485	7471	6381	6148	7730
Area of the major hydrophobic cluster (Å ²)	1171	1116	1035	1111	1066	2605	983	984	2114	1169	1180	1111	1162	4610	4611	4329	2524	4265	4416	4351	4519
% of total area present in the major hydrophobic cluster	29	30	24	26	22	40	20	21	35	22	23	22	26	67	79	79	56	57	69	71	58
Total residues in hydrophobic clusters	31	31	32	32	36	42	36	36	40	40	41	43	36	52	40	40	50	57	56	60	64
Residues in the major hydrophobic cluster	8	8	8	8	8	14	7	7	12	8	8	9	8	32	32	32	14	32	33	38	34
% of total residues present in the major hydrophobic cluster	26	26	25	25	22	33	19	19	30	20	20	21	22	62	80	80	28	56	59	63	53
Total contacts in hydrophobic clusters	82	81	89	91	98	120	84	88	112	104	103	113	94	143	122	114	129	152	133	150	166
Contacts in the major hydrophobic cluster	24	23	24	23	23	48	18	18	40	24	23	24	23	96	99	92	48	91	94	111	104
% of total contacts present in the major hydrophobic cluster	29	28	27	25	23	40	21	20	36	23	22	21	24	67	81	81	37	60	71	74	63
Total ASA of the folded protein (Å ²)	8661	8669	8784	8761	8920	8336	8366	8362	8360	8323	9126	9052	9149	8472	9187	9140	9002	9016	9070	8867	8730
Hydrophobic ASA of the folded protein (%)	48	51	46	51	47	46	45	46	46	47	52	58	50	46	48	48	50	50	53	53	51
Total ΔASA buried on protein folding (Å ²)	18893	17815	18844	17331	18199	19380	19335	19017	19429	19060	18878	17251	18141	18955	18000	17947	19469	19008	19111	18928	20048
Hydrophobic ΔASA buried on protein folding (%)	79	77	79	74	77	79	78	78	79	79	79	70	81	80	78	80	77	80	79	75	78
Physicochemical amino acid properties (%)																					
Aliphatic (AGILPV)	34		35		37	38	35	35	37	39	41	39	43	41	41	42	45	48		49	
Aromatic (FHMY)	9		9		7	8	8	7	8	7	7	7	7	2	4	2	8	3		2	3
Charged (DEHKR)	43		43		43	41	41	42	41	41	41	46	39	39	41	39	37			37	35
Hydrophobic (CFILMVW)	24		24		24	28	27	25	28	26	28	26	26	33	24	28	33			37	41
Polar (DEKNQR)	49		50		50	48	49	49	48	48	48	50	46	48	48	46	43			43	41

Sina Kordes*, Sergio Romero-Romero*, Leonie Lutz, Birte Höcker
A newly introduced salt bridge cluster improves structural and biophysical properties of de novo TIM
barrels
Protein Science, 2021, 1-15
*equal contribution

Received: 14 October 2021 | Revised: 30 November 2021 | Accepted: 3 December 2021

DOI: 10.1002/pro.4249

FULL-LENGTH PAPER



A newly introduced salt bridge cluster improves structural and biophysical properties of *de novo* TIM barrels

Sina Kordes  | Sergio Romero-Romero  | Leonie Lutz  | Birte Höcker 

Department of Biochemistry, University of Bayreuth, Bayreuth, Germany

CorrespondenceBirte Höcker, Department of Biochemistry, University of Bayreuth, Bayreuth 95447, Germany.
Email: birte.hoecker@uni-bayreuth.de**Funding information**

Alexander von Humboldt-Stiftung; H2020 European Research Council, Grant/Award Number: 647548

Abstract

Protein stability can be fine-tuned by modifying different structural features such as hydrogen-bond networks, salt bridges, hydrophobic cores, or disulfide bridges. Among these, stabilization by salt bridges is a major challenge in protein design and engineering since their stabilizing effects show a high dependence on the structural environment in the protein, and therefore are difficult to predict and model. In this work, we explore the effects on structure and stability of an introduced salt bridge cluster in the context of three different *de novo* TIM barrels. The salt bridge variants exhibit similar thermostability in comparison with their parental designs but important differences in the conformational stability at 25°C can be observed such as a highly stabilizing effect for two of the proteins but a destabilizing effect to the third. Analysis of the formed geometries of the salt bridge cluster in the crystal structures show either highly ordered salt bridge clusters or only single salt bridges. Rosetta modeling of the salt bridge clusters results in a good prediction of the tendency on stability changes but not the geometries observed in the three-dimensional structures. The results show that despite the similarities in protein fold, the salt bridge clusters differently influence the structural and stability properties of the *de novo* TIM barrel variants depending on the structural background where they are introduced.

KEYWORDS

(β/α)₈ barrel, *de novo* protein design, DeNovoTIMs, protein folding, protein stability, salt bridge cluster, TIM barrel

Abbreviations: CD, circular dichroism; DSC, differential scanning calorimetry; $D_{[1/2]}$, midpoint urea unfolding concentration; IF, intrinsic fluorescence; MALS, multi angle light scattering; REU, Rosetta energy unit; SEC, size exclusion chromatography; T_m , midpoint of thermal unfolding; ΔC_p , change in heat capacity; $\Delta G_{25^\circ C}$, change in Gibbs free energy at 25°C; ΔH , change in enthalpy.

Sina Kordes and Sergio Romero-Romero authors contributed equally to the work.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Protein Science* published by Wiley Periodicals LLC on behalf of The Protein Society.

1 | INTRODUCTION

Protein stability is a fundamental biological attribute that modulates the delicate balance among protein evolvability, expression, solubility, structure, and function.^{1–3} It results from the accumulated balance of forces and interactions between protein and solvent that determines whether the folded conformation is stable over other nonfunctional

competing states. The central role of proteins in the chemistry of life, as well as their increasing application in basic and applied research, makes an understanding of protein stability highly relevant.

The information obtained about the forces that fine-tune protein stability come from numerous studies on natural proteins and have led to the possibility to design proteins from scratch. Those computationally designed proteins differ significantly in sequence and structure from naturally occurring proteins, providing new information to gain a deeper understanding of the relationship between sequence, structure, and stability.⁴

Several strategies to increase protein stability have been explored such as rearrangement of hydrogen-bond networks, introduction of salt bridges, improving of hydrophobic cores, or incorporation of covalent bonds.^{5–9} Among them, the prediction and engineering of salt bridges is challenging, due to their high dependence on the structural environment of the protein and the requirement of accurate geometries.^{1,10} Consequently, it is very difficult to estimate the energetic contribution to stability caused by a salt bridge due to a delicate balance of destabilizing desolvation energy and stabilizing interactions.¹¹

A salt bridge can be defined as an ion-pair interaction between two residues of opposite charge with a distance below 4 Å that combines two noncovalent interactions, hydrogen bonding and ionic bonding.¹² This type of interaction plays an important role in defining protein structure, function and stability^{11,13–16} and has been a valuable strategy in protein engineering to stabilize different proteins and calculate their energetic contributions.^{17–21} In addition, the interaction of one basic residue with multiple acidic residues form clustered or networked salt bridges, which are of special interest due to their complexity and important contribution to protein stability.^{10,22,23}

Among all protein architectures, the TIM barrel is one of the most common folds in nature, as one-tenth of the known proteins adopt this topology and it is found in five out of seven enzyme classes.^{24,25} This ubiquitous and versatile topology has been an important model system to study not only the stability, structure, and function relationships but also for *de novo* protein design. Previously, we demonstrated that increasing the hydrophobic clusters of the first *de novo* TIM barrel sTIM11²⁶ resulted in a highly-stable collection of TIM barrels, which we called DeNovoTIMs.²⁷ In the work presented here, we explore the effects on structure and stability when introducing a salt bridge cluster into members of the DeNovoTIM collection.

2 | RESULTS AND DISCUSSION

2.1 | Introducing a salt bridge cluster into different *de novo* TIM barrels

The effects of introducing a salt bridge cluster were inspired by the presence of a similar cluster in the natural HisF TIM barrel, a subunit of the imidazole glycerophosphate synthase (IGPS),^{28,29} and its observed influence in stabilizing this fold.^{30,31} Since the salt bridge cluster was intended to be evaluated without affecting the previous stabilized regions in our DeNovoTIM collection, we focused on the internal core of the barrel to introduce the salt bridge network. Therefore, considering the environmental and geometrical descriptors of the most common salt bridges found in natural proteins,¹⁰ we found that at the internal core of the TIM barrel, and specifically on the bottom part of it, 4 symmetry-related glutamine residues were suitable to introduce the intended salt bridge cluster. The four residues were alternatively mutated to Arg and Glu in the four quarters as indicated in Figure 1 and Table S1.

We explored the effects of the salt bridge cluster in the context of three different *de novo* TIM barrels previously reported as DeNovoTIM collection,²⁷ all designed using a computational fixed-backbone and modular approach to improve the hydrophobic packing: *sTIM11noCys*, the cysteine-free variant of sTIM11 (the first validated *de novo* TIM barrel) without any extra stabilizing mutations as present in DeNovoTIMs; *DeNovoTIM6*, with stabilizing mutations in the bottom region of the peripheral core located between the outer face of the β -strands and the internal face of the α -helices; and *DeNovoTIM13*, with stabilizing mutations in the bottom and top regions of the peripheral core (Table S1). The TIM-barrel architecture among the three proteins is conserved with an RMSD <1.5 Å. The main differences are related to the size and packing of the hydrophobic clusters due to the introduced mutations.²⁷ Salt bridge variants derived from parental proteins (named as the original design plus the suffix-SB) were biophysically and structurally characterized as follows.

2.2 | Salt bridge cluster variants are soluble monomeric and well-folded TIM barrels

All salt bridge variants were expressed and purified to homogeneity in high amounts with similar yields to the parental proteins. Just as the parental proteins DeNovoTIM6-SB and DeNovoTIM13-SB show about 15% dimer in the preparative size exclusion chromatography (SEC). Multi angle light scattering (MALS) measurements

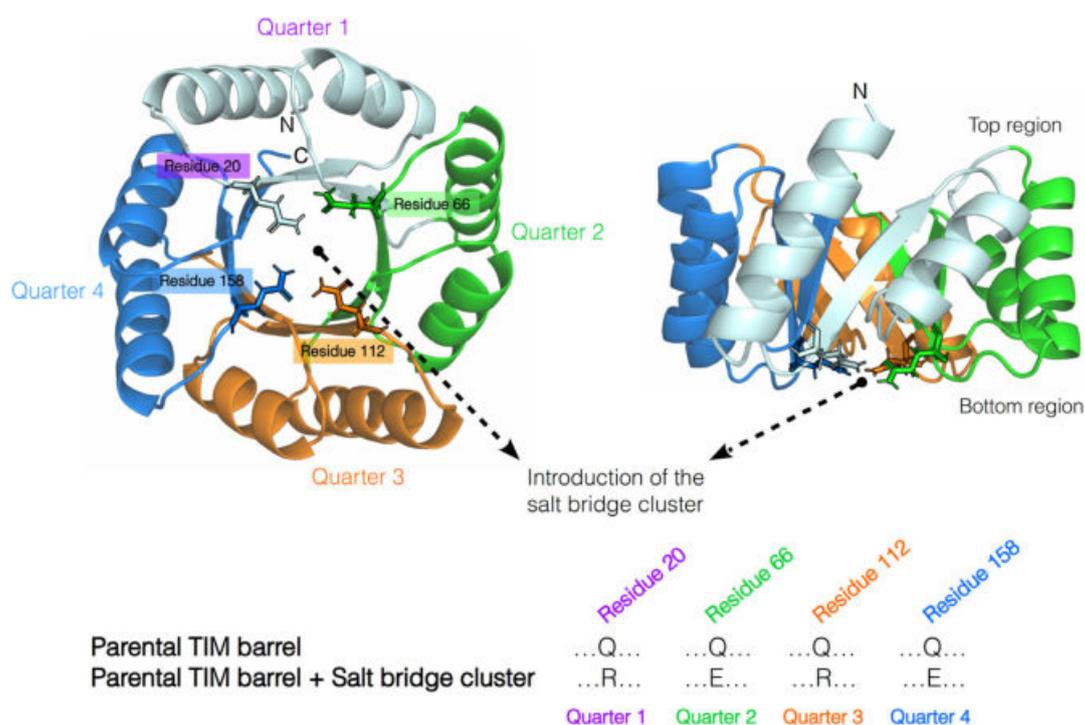


FIGURE 1 Strategy to introduce a salt bridge cluster in *de novo* TIM barrels. Crystal structure of sTIM11noCys is shown with residues 20, 66, 112, and 158 highlighted as sticks, which were used for the introduction of the salt bridge cluster. It was added in the *de novo* TIM barrels replacing Q20 and Q112, belonging to the first and third quarters, by arginine residues, and Q66 and Q158, from the second and fourth quarters, by glutamic acid residues

of the monomer peak for both proteins at different concentrations revealed no concentration dependent dimerization in the range up to 5 mg ml^{-1} . Additionally, the dimer was analyzed and appeared to be stable. All further experiments were done with the monomeric fraction. The molecular weight of all three proteins was determined using MALS and verified as monomers (Figure S1 and Table S2). In contrast to DeNovoTIM13, which showed a tendency for aggregation after purification, DeNovoTIM13-SB did not aggregate in the observed time frame, probably due to the thermal-unfolding reversibility which is not present in DeNovoTIM13.

Circular dichroism (CD) analysis indicated well folded proteins with a mixed α/β secondary structure composition similar to the parental proteins (Figure S2, Table S2). For sTIM11noCys-SB a higher signal at about 222 nm is observed hinting at an increase in the α -helical fraction. For DeNovoTIM13-SB additionally the peak at 208 nm is more pronounced showing gain of overall secondary structure. In contrast, DeNovoTIM6-SB shows an overall decrease of the signal and a more pronounced signal at 208 nm. Deconvolution of the far-UV CD spectra displays slight differences in the secondary structure contents (Table S2), which is confirmed by the three-dimensional structure analysis as discussed below. For

sTIM11noCys-SB and DeNovoTIM13-SB a decrease of random coil connected with an increase of secondary structure content is calculated. In contrast, the deconvolution of DeNovoTIM6-SB CD spectrum indicates a similar content of random coil but with differences in the α -helix and β -sheet composition (Table S2). These data confirm well folded proteins on the basis of their spectroscopy attributes without large structural changes upon the introduction of the salt bridge cluster. To follow up on this, their folding stability behavior was studied by thermal and chemical unfolding experiments.

2.3 | Thermostability is maintained in the salt bridge variants

Thermal stability was initially analyzed by CD (Figure S3). For sTIM11noCys-SB, a melting temperature (T_m) of about 64°C was determined. DeNovoTIM6-SB and DeNovoTIM13-SB both do not completely unfold in the accessible temperature range up to 95°C . Therefore, further analysis was performed by differential scanning calorimetry (DSC). For all three proteins no change of T_m was observed in comparison to the parental proteins (Table 1, Figure 2a), but with some changes in the

TABLE 1 Thermodynamic parameters of salt bridge cluster variants in comparison with the parental proteins

<i>de novo</i> TIM barrel	Thermal unfolding (by DSC)					Chemical unfolding (by CD and IF)				
	T_m (°C)	ΔH (kcal mol ⁻¹)	$\Delta H_{85^\circ\text{C}}$ (kcal mol ⁻¹)	ΔC_p (kcal mol ⁻¹ K ⁻¹)	Calorimetric criterion $\Delta H_{\text{VH}}/\Delta H$	Global thermodynamic stability (kcal K mol ⁻¹)	$\Delta G_{25^\circ\text{C}}$ (kcal mol ⁻¹)	m (kcal mol ⁻¹ M ⁻¹)	$D_{[1/2]}$ (M)	
sTIM11noCys ^a	65.6 ± 0.1	82 ± 1	128 ± 2	2.36 ± 0.08	0.99 ± 0.03	176	3.2 ± 0.2	2.03 ± 0.1	1.9	
sTIM11noCys-SB	67.4 ± 0.4	88 ± 2	123 ± 3	2.35 ± 0.09	1.00 ± 0.01	216	4.8 ± 0.1	2.66 ± 0.01	1.8	
DeNovoTIM6 ^a	92.3 ± 0.1	125 ± 2	108 ± 1	2.38 ± 0.06	1.03 ± 0.02	542	7.9 ± 0.2	1.51 ± 0.08	5.6	
DeNovoTIM6-SB	91.7 ± 0.1	128 ± 1	107 ± 1	2.09 ± 0.03	1.00 ± 0.03	684	9.8 ± 0.3	1.76 ± 0.06	5.6	
DeNovoTIM13 ^a	92.8 ± 0.4	47 ± 5	n.d., E_{act} : 120 ± 3			n.d.	9.5 ± 0.2	1.54 ± 0.02	6.6	
DeNovoTIM13-SB	92.7 ± 0.3	100 ± 2	84 ± 1	1.48 ± 0.06	0.99 ± 0.02	596	8.2 ± 0.3	1.18 ± 0.03	6.9	

Note: n.d., not determined due to irreversibility in the thermal unfolding. Instead, activation energy (E_{act}) was calculated from an irreversible two-state mechanism.

^aParameters reported in Reference 27.

enthalpy ΔH , mainly for DeNovoTIM13-SB. Both, sTIM11noCys-SB and DeNovoTIM6-SB, show thermal unfolding reversibility (Figure S4a, S4c) and were fitted to a reversible two-state model (Figure S4b, S4d).

In contrast, DeNovoTIM13-SB as well as DeNovoTIM13 show thermal-unfolding irreversibility. Moreover, the behavior of DeNovoTIM13-SB is remarkably different to that of DeNovoTIM13: the T_m in the salt-bridge variant is only slightly dependent on the scan rate (Figure S5a) and the area recovered in the DSC reversibility test increased (Figure S4e), changing from 14% for DeNovoTIM13 to 72% for DeNovoTIM13-SB. The lack of significant scan rate effects was demonstrated using a wide scan rate range from 0.5 to 3.0°C min⁻¹ as previously suggested,³² confirming that the degree of recovery observed in DeNovoTIM13-SB does not cause distortions in the baselines or thermal transition, which validate an equilibrium thermodynamics analysis (Figure S4f).

This combination of irreversibility and lack of scan rate dependence has been rarely reported. Typically, calorimetric irreversibility is caused by protein aggregation, but the lack of scan rate effect can be interpreted by assuming that the processes causing irreversibility only take place at very high temperatures where the protein is already completely unfolded.³³ In addition to protein aggregation, there exists the possibility that swapped oligomers are formed in the unfolding state of DeNovoTIM13-SB, which would be thermodynamically more similar to unfolded monomers rather than aggregates, therefore allowing a proper fitting to a reversible model. However, due to the high T_m of this protein, a more in-depth analysis of the unfolded state with techniques such as CD and fluorescence spectroscopy, or SEC-MALS is not possible.

Although the rule of thumb for thermal-unfolding reversibility considers a recovered area higher than 85% for a reversible process, the thermodynamic behavior of DeNovoTIM13-SB, that is, calorimetric irreversibility and no scan rate effects, allows to fit the endotherms to a reversible two-state model as has been reported.³³ In fact, when comparing the fitting for both the irreversible and reversible two-state models (Figure S5b,c), the reversible model fits and explains the experimental data much better than the irreversible one. Also, the calorimetric criterion ($\Delta H_{\text{VH}}/\Delta H$) is very close to 1, which is in agreement with a two-state mechanism (Table 1). All these results confirm the suitability of this model to calculate the thermodynamic parameters for DeNovoTIM13-SB.

Thermodynamic parameters determined for the salt bridge variants indicate no major changes when compared to the parental proteins (Table 1). The main difference is observed in the heat capacity change (ΔC_p), which reshapes the stability curve without modifying the

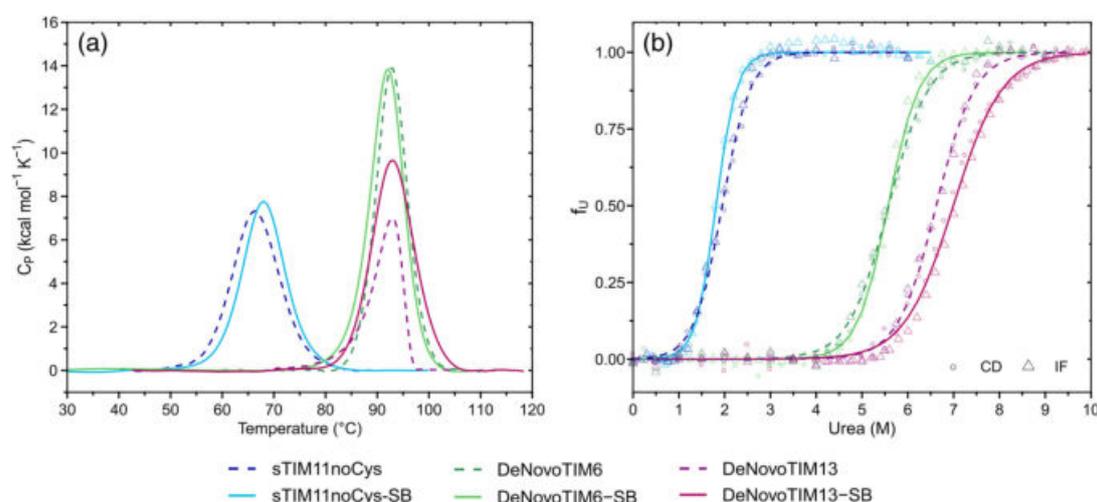


FIGURE 2 Folding stability of the salt bridge variants. (a) Thermal unfolding experiments followed by DSC. Endotherms were collected at $1.5^{\circ}\text{C min}^{-1}$ and protein concentration of 1.0 mg ml^{-1} . Dotted lines show the parental proteins and continuous lines the salt bridge cluster variants. (b) Chemical unfolding with urea at 25°C , circles representing CD data and triangles fluorescence data. Dotted and continuous lines represent the fitting of the data to a reversible two-state model for the parental and salt bridge variants, respectively. Data from sTIM11noCys, DeNovoTIM6, and DeNovoTIM13 are reported in Reference 27. All experiments were collected in 10 mM sodium phosphate pH 8

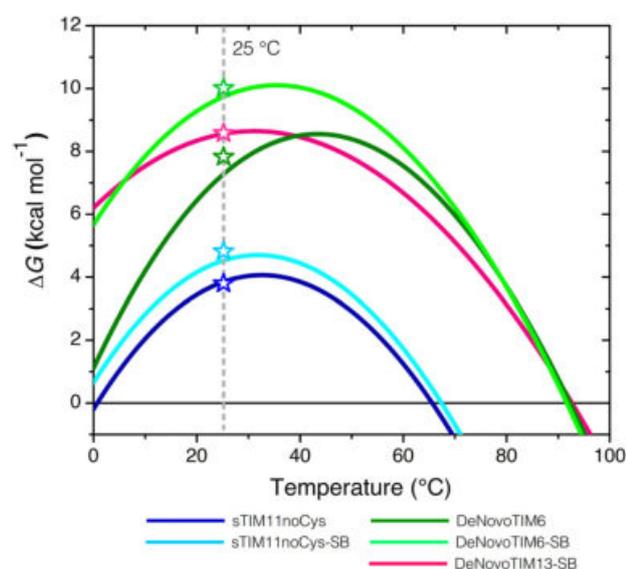


FIGURE 3 Stability curves of the salt bridge variants. Curves were constructed using the parameters from DSC experiments and the Gibbs–Helmholtz equation. Open symbols indicate the ΔG value at 25°C determined by chemical unfolding. Data from sTIM11noCys and DeNovoTIM6 are reported in Reference 27

T_m (Figure 3) but increases the conformational stability at 25°C as discussed in the next section. Interestingly, no major changes in thermodynamic stability are observed regarding the T_m . This stands in contrast to the assumption that salt bridges increase mostly the thermal stability of proteins as observed in thermophilic proteins.^{34–37}

Also, the web server for protein stabilization called Protein Repair One Stop Shop (PROSS) creates thermally stabilized protein by introducing salt bridges,⁶ though with the objective to influence the solvent exposure and to strengthen Coulomb interactions in the low-dielectric protein core.¹ The mutations that introduce salt bridges may have different effects on folding energies depending on the temperature;^{1,38} this can only be answered if conformational stability would be analyzed at different temperatures. Here, upon introduction of the salt bridge cluster most of the differences are observed on the conformational stability at 25°C .

2.4 | Salt bridge cluster variants have a higher conformational stability at 25°C

Changes in the conformational stability at 25°C ($\Delta G_{25^{\circ}\text{C}}$) were studied by chemical unfolding with urea followed by CD and IF. All three proteins showed reversible and cooperative transitions. These fitted well to a two-state model ($N \rightleftharpoons U$) (Figure 2b) with coincident $\Delta G_{25^{\circ}\text{C}}$ values to those calculated from thermal unfolding experiments (Figure 3). Comparison of the salt bridge variants with the parental proteins exhibited different trends: In sTIM11noCys-SB, the salt bridge cluster stabilized the protein by an increase of 1.6 kcal mol^{-1} in $\Delta G_{25^{\circ}\text{C}}$, where the midpoint urea unfolding concentration ($D_{[1/2]}$) stayed unchanged but the m value increased by $0.63\text{ kcal mol}^{-1}\text{ M}^{-1}$ (Table 1), indicating an improved

protein packing. Interestingly, the salt bridge cluster induces in sTIM11noCys-SB a similar stability as observed for sTIM11, namely $4.8 \text{ kcal mol}^{-1}$.²⁷ sTIM11 contains two additional cysteines that are close in proximity but do not appear to form a disulfide bond. However, sTIM11 has a lower m value but an increased $D_{[1/2]}$ compared to sTIM11noCys and sTIM11noCys-SB. For DeNovoTIM6-SB a similar trend is observed but with an even larger stabilizing effect of $1.9 \text{ kcal mol}^{-1}$ resulting in a $\Delta G_{25^\circ\text{C}}$ of $9.8 \text{ kcal mol}^{-1}$ due to an increase of m value by $0.25 \text{ kcal mol}^{-1} \text{ M}^{-1}$. On the other hand, in DeNovoTIM13-SB the salt bridge cluster seems to destabilize the protein slightly by $-1.3 \text{ kcal mol}^{-1}$ resulting in a reduced $\Delta G_{25^\circ\text{C}}$ of $8.2 \text{ kcal mol}^{-1}$. Also, the m value is decreased by $-0.36 \text{ kcal mol}^{-1} \text{ M}^{-1}$ with a slight increase of $D_{[1/2]}$ (Table 1).

This shows that despite a similar context of the basic protein topology, the stability contribution of the salt bridge residues is different. In sTIM11noCys-SB and DeNovoTIM6-SB the salt bridge cluster has a clear stabilizing effect. On the contrary, a destabilizing effect is observed in DeNovoTIM13-SB, although both effects are modulated by changes in the m value. To analyze the thermodynamic contribution of the salt bridge network in these *de novo* TIM barrels in detail, other approaches as computing the electrostatic energies by *in silico* mutation to their hydrophobic isosteres^{39–41} or estimating the

stability changes using a double-mutant cycles^{19,41–44} could be used. However, due to the cluster nature of the salt bridges introduced here, the complexity of the analysis would complicate determining the contribution of each residue. Nonetheless, the stability changes in the salt bridge variants were correlated with the structural rearrangements that took place in the barrel when the salt bridge cluster was introduced.

2.5 | The mutations improve crystallization properties of the *de novo* TIM barrels

In order to analyze the structural effects of the salt bridge mutations on the *de novo* TIM barrels, the three-dimensional structures were solved by protein crystallography (Figures 4, S6, and Table S3). For all three parental proteins, the three-dimensional structures were solved previously.²⁷ For sTIM11noCys and DeNovoTIM13 crystallization was straightforward and structures were solved at high resolution. In contrast, crystallization and structure determination of DeNovoTIM6 was challenging. Despite a large number of crystals in screening and trials of several post crystallization treatments, only a low-resolution structure could be solved. Based on this experience, we wanted to analyze and compare the

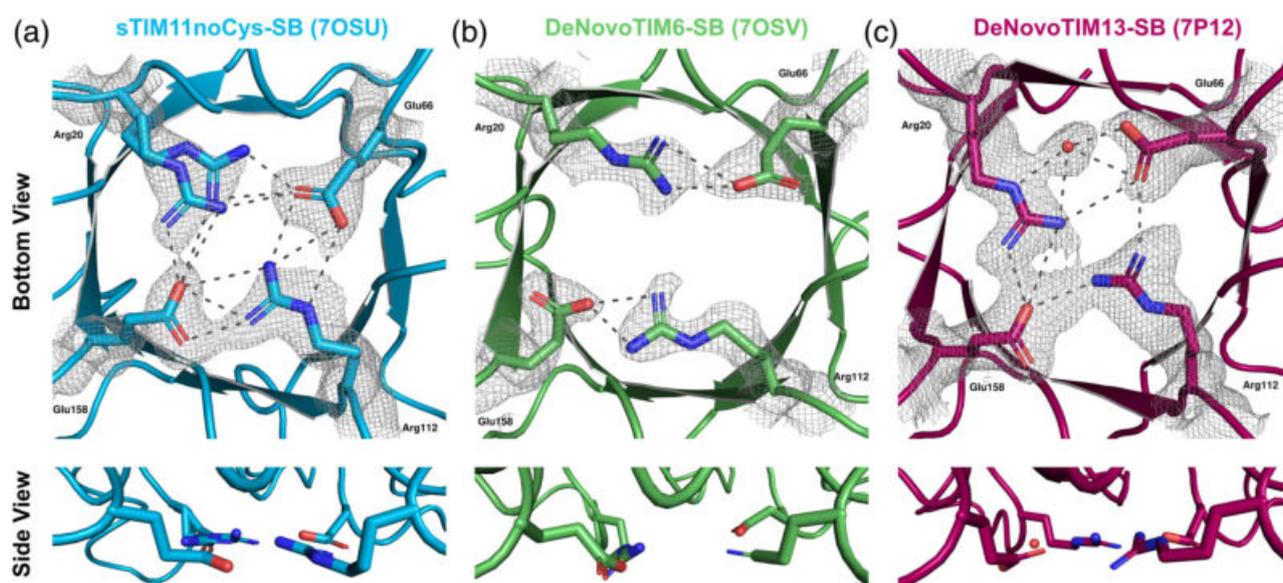


FIGURE 4 Structural conformations of the salt bridge interactions in the *de novo* TIM barrels. (a) sTIM11noCys-SB (crystal form 1, PDB ID: 7OSU). (b) DeNovoTIM6-SB (crystal form 1, PDB ID: 7OSV). (c) DeNovoTIM13-SB (PDB ID: 7P12). In all panels, upper figures indicate the view from the bottom of the barrel with the salt bridge residues highlighted in sticks. $2\text{Fo}-\text{Fc}$ electron density maps contoured at 1σ are shown as a gray mesh for all the residues/water involved in the salt bridge cluster. Lower figures show the side view of the salt bridge interactions to analyze their planarity. Dotted lines indicate the salt bridge interactions between the mutated residues whose measures are reported in Table 2

crystallization properties of the salt bridge variants with those of the parental barrels.

Crystallization of all three salt bridge variants started with screening of several conditions. sTIM11noCys-SB crystallized in the same space group as sTIM11noCys but the resolution was considerably improved from 1.88 to 1.37 and 1.51 Å (resolution for the two different crystal forms observed at different pH conditions, respectively).

Similar to its parental protein, DeNovoTIM6-SB yielded a high number of crystals from screening with a hit rate of about 15% and crystallization conditions favoring ammonium sulfate as precipitant. Nevertheless, the crystals showed improved diffraction quality as is reflected in a decreased mosaicity, improved diffraction patterns and particularly in an improved resolution from 2.9 Å for DeNovoTIM6 to 1.66 and 2.22 Å (crystal form 1 and 2, respectively, crystallized in different conditions and space groups).

Also DeNovoTIM13-SB crystallized in many conditions during screening. The best diffracting crystal was found in a condition similar to the one of DeNovoTIM13 but had a different space group. We observed an interesting anomaly resulting from the high symmetry of the *de novo* TIM barrels: initial data processing assumed the space group I4, but in the associated unit cell dimensions only a quarter of the TIM barrel could be fitted (Figure S7a). Molecular replacement using only a quarter of DeNovoTIM13 yielded good scores, and symmetry operations resulted in a nicely reconstructed TIM barrel using the corresponding symmetry operations with the center of the protein located in the vertex of the unit cell (Figure S7b). Nevertheless, for structure determination, the dataset had to be processed using space group P1, which produced a unit cell with a volume large enough for a complete TIM barrel in the asymmetric unit. To achieve higher completeness and low radiation damage in the P1 space group, datasets from two crystals from the same condition were merged. Despite this interesting observation, it did not affect the diffraction quality as the structures of DeNovoTIM13 and DeNovoTIM13-SB have similar resolutions of 1.64 and 1.69 Å, respectively.

Since DeNovoTIM6-SB and DeNovoTIM13-SB show a small dimer population during protein purification and SEC-MALS measurements, a possible dimer generation in the crystal structures was analyzed using symmetry operators and noncrystallographic symmetry. However, in none of the cases the formation of the oligomer could be recapitulated that would hint at the structural properties of DeNovoTIM6-SB and DeNovoTIM13-SB dimers observed in solution. For DeNovoTIM6-SB, even though crystallization was set up only with the monomeric protein fraction, one of the crystal structures (PDB ID: 7OT8) shows two molecules in the asymmetric unit;

nevertheless, no tight interface interactions are observed in this crystal form.

These observations of improved crystallization can be correlated with the conformational stability of the proteins: sTIM11noCys-SB and DeNovoTIM6-SB both have an increased stability that in both cases is mediated by the *m* value, which is connected to the protein packing. These effects on the crystallization behavior and the data quality additionally raise interest in the exact geometry and how it might influence the protein fold. Therefore, the solved structures were analyzed in detail regarding the geometry of the cluster residues, and additional changes of the TIM barrel fold were investigated.

2.6 | The salt bridge residues arrange with different geometries on the *de novo* TIM barrels

Several previous studies have pointed out that the contribution of a salt bridge to stability is connected with its geometry, that is, distances and angles.^{10,42,45,46} A comprehensive overview of possible geometric orientations of salt bridges and a statistical analysis of their frequency is given by Donald et al.¹⁰ Analysis of the formed geometries of these residues in the solved crystal structures showed a varying behavior depending on the structural context of the *de novo* TIM barrel where the residues were introduced (Table 2).

The solved crystal structures of the salt bridge variants did not show deviations in their TIM-barrel topology. Interestingly, in all three proteins different salt bridge geometries are observed. In sTIM11noCys-SB, a highly ordered salt bridge cluster is formed quite similar to the intended geometry (Figure 4a). Arg20 shows two alternative conformations, providing two different salt bridge sets. One conformation (Arg20-A) forms two salt bridge interactions with Glu66 and one with Glu158. The alternative conformation Arg20-B similarly forms two salt bridges with Glu158 and one with Glu66. Arg112 forms a highly coordinated side-on salt bridge with Glu66 and additionally interacts with Glu158 in an end-on configuration, in both cases via three interactions (Table 2). These two configurations are expected to be lowest-energy states based on quantum mechanics calculations.¹⁰ A single energetically favorable salt bridge must have a good balance between the unfavorable entropic cost and favorable coulombic interactions. In the case of a salt bridge cluster formation as observed in sTIM11noCys-SB, the entropic cost should be lower than for a single one, as one side chain is already reduced in its degrees of freedom.⁴²

No	Residue 1	Residue 2	Distance (Å)	Angle (°)
sTIM11noCys-SB (7OSU)				
1	Arg20A-Nη1	Glu66-Oε1	3.3	164.0
2	Arg20A-Nη2	Glu66-Oε1	2.7	
3	Arg20A-Nη1	Glu158-Oε2	3.8	108.0
4	Arg20B-Nη2	Glu158-Oε2	3.1	101.4
5	Arg20B-Nη2	Glu158-Oε2	3.6	158.8
6	Arg20B-Nη1	Glu158-Oε2	2.6	
7	Arg112-Nη2	Glu66-Oε1	3.2	95.9
8	Arg112-Nη2	Glu66-Oε2	3.4	65.8
9	Arg112-Nε	Glu66-Oε2	3.6	
10	Arg112-Nη1	Glu158-Oε1	2.7	137.8
11	Arg112-Nη1	Glu158-Oε2	3.2	170.0
12	Arg112-Nη2	Glu158-Oε2	3.7	
DeNovoTIM6-SB (7OT8)				
1	Arg20-Nη1	Glu66-Oε2	3.5	163.2
2	Arg20-Nη2	Glu66-Oε2	2.9	
3	Arg112-Nη1	Glu158-Oε2	3.0	158.2
4	Arg112-Nη2	Glu158-Oε2	2.8	
DeNovoTIM13-SB (7P12)				
1	Arg20-Nε	H2O29	2.7	52.3
2	Arg20-Nη2	H2O29	3.1	
3	Arg20-Nη2	Glu66-Oε1	3.7	93.2
4	Arg20-Nη1	Glu158-Oε2	2.7	157.3
5	Arg20-Nη2	Glu158-Oε2	3.7	
6	H2O29	Glu66-Oε2	2.6	107.3
7	H2O29	Glu66-Oε1	3.1	81.1
8	Arg112-Nη1	Glu66-Oε1	3.1	134.6
9	Arg112-Nη2	Glu158-Oε2	3.0	110.0

TABLE 2 Salt bridge geometries of the *de novo* TIM barrels analyzed in this work

The structural analysis of the DeNovoTIM6-SB salt bridge residues revealed a different configuration. In contrast to sTIM11noCys-SB, no cluster is formed between the four residues and only two independent salt bridge pairs are observed between Arg20-Glu66 and Arg112-Glu158. In both cases, two interactions between arginine and glutamate in a monodentate backside configuration are made (Figure 4b). This reduced number of interactions in comparison to sTIM11noCys-SB does not correlate directly with the changes in stability for these two proteins, where for sTIM11noCys-SB the stability is increased by 1.6 kcal mol⁻¹ compared to sTIM11noCys, and DeNovoTIM6-SB increased by 1.9 kcal mol⁻¹ compared to DeNovoTIM6-SB (Table 1).

For DeNovoTIM13-SB, again a different configuration of the introduced network is observed in this case involving a water molecule (Figure 4c). A bidentate water-

mediated salt bridge is formed between Arg20 and Glu66 additionally to a monodentate direct salt bridge. In addition, Arg20 interacts with Glu158, which also interacts via a single salt bridge with Arg112. Closing the network another single salt bridge is formed between Glu66 and Arg112 (Table 2). The reduction on the conformational stability of DeNovoTIM13-SB compared to DeNovoTIM13 would indicate possible negative influences of the salt bridge cluster on the topology. Therefore, the highly coordinated salt bridge network observed in the crystal structure indicates that other structural rearrangements take place in different regions of the barrel. Considering the approach followed to design DeNovoTIM13 that aimed on the improvement of hydrophobic clusters,²⁷ changes in these clusters were analyzed in DeNovoTIM13-SB as a possible cause for the reduction in stability. In fact, we observed that DeNovoTIM13-SB

exhibits a reduction of 409 Å² in the total area of the hydrophobic clusters in comparison to DeNovoTIM13 (5,972 vs. 6,381 Å², respectively). Since epistatic effects play an important role in the stabilization of DeNovoTIMs, a likely reason why DeNovoTIM13-SB reduces in stability is that the introduction of the salt bridge cluster promotes rearrangements that modify the hydrophobic clusters.

Another suggested factor for the stability of a salt bridge is its planarity, meaning a higher planarity implies a better coordination of the charged residues resulting in a higher contribution to the stability.^{10,47} Analysis of the three different salt bridge clusters shows highest planarity for sTIM11noCys-SB (Figure 4a bottom), followed by DeNovoTIM13-SB with a slight distortion of Arg112 (Figure 4C bottom) and lowest planarity in DeNovoTIM6-SB, which is most likely due to the absence of a well-formed cluster geometry (Figure 4b bottom). In addition to the absence of a full cluster in DeNovoTIM6-SB and the therefore low overall planarity between the four residues, also the interacting residue pairs have a low planarity indicating a poor coordination of the separated salt bridges.

Possible influence of the crystal packing on the salt bridge geometries were analyzed by comparing the Matthews coefficient (V_m) and solvent content of the different variants (Table S3). Crystals of sTIM11noCys-SB and DeNovoTIM6-SB have a similar V_m of about 1.9 leading to a solvent content of 36% and 37%, respectively. DeNovoTIM13-SB on the other hand has a higher V_m of 2.25 with a solvent content of 45%. We deduce that the better coordinated geometry of sTIM11noCys is not induced by a tighter crystal packing, as DeNovoTIM6-SB has a similar solvent content.

Finally, since the design and prediction of salt bridge networks and their corresponding changes in stability are open challenges, we tested if the determined changes in stability observed in the salt bridge variants could be predicted by Rosetta scoring.

2.7 | Rosetta recapitulates the changes in stability but not the salt bridge geometries

Prediction and modeling of salt bridges is challenging due to the importance of a well-formed geometry of the involved residues. As we did not perform any preceding modeling of the introduced salt bridge cluster but based it on similar clusters observed in the natural TIM barrel HisF,^{28,29} we were interested if the introduced electrostatic interactions could be accurately modeled and scored according to our experimental data. Therefore, the

mutated residues were introduced into the parental protein structures using Rosetta Remodel followed by relaxation of the models.⁴⁸ Analysis of the created models shows that a complete network is rarely built and also does not score best. Nevertheless, most of the decoys show partially formed networks with one arginine frequently pointing out. Obviously, generating the geometry observed in NovoTIM13-SB involving a water molecule is not possible with this design approach, as no water molecules are included. Still, a more optimized approach might have led to a better performance.

To analyze the influence of a well-placed salt bridge network on the Rosetta score, the crystal structures of the salt bridge variants as well as the parental proteins were scored using the most recent and default scoring function *ref2015* in Rosetta.⁴⁹ First, all structures were idealized using *rosetta.relax* with constrained backbone and sidechain geometries to the starting structure.⁵⁰ Due to varying residue numbers of the structures, the Rosetta scores were normalized to the total number of residues. Comparison of the scores of each parental protein with its corresponding salt bridge variant showed a decrease of the total score for sTIM11noCys-SB and DeNovoTIM6-SB by -1.24 and -1.37 Rosetta energy units (REU), respectively. However, DeNovoTIM6 scores really low with a total of -0.12 REU per residue due to the low resolution of the structure and the associated poor model quality including missing side chains. In contrast, DeNovoTIM13-SB shows an increase of the total score by 0.23 REU compared to DeNovoTIM13 (Table S4).

Comparing the experimentally determined stability for these proteins with the corresponding scores shows that Rosetta is predicting a similar trend for all proteins regarding their stability. For sTIM11noCys and DeNovoTIM6 the score indicates an increase in stability with the addition of the salt bridge cluster, which was verified in the unfolding experiments. Also, the scores of DeNovoTIM13 and DeNovoTIM13-SB are in agreement with the experiments, showing that in this context the salt bridge cluster does not have a positive effect on the stability.

Collectively, using a minimalistic design approach with Rosetta it was not possible to generate models similar to the final solved structures, which might be improved with the introduction of specific constraints and scores. Nevertheless, scoring of the native salt bridge clusters reveals clearly that the most recent scoring function perceives the cluster and considers it positively.

3 | CONCLUSIONS

The engineering, design and prediction of the stabilizing effect of salt bridges in proteins is a challenging task due

to the high interdependency of various factors. Despite the necessity for an optimal geometry, a stabilizing effect is only achieved with the compensation of the desolvation penalty by the bridging energy and other interactions caused by conformational changes.

Here we studied the effects on structure and stability by the introduction of a salt bridge cluster into three different *de novo* TIM barrels. In contrast to findings of previous studies, which correlated the increased thermostability of many proteins with an increased number of salt bridges, our analysis showed no influence on the T_m values for all three proteins. In contrast, analysis of the conformational stability at 25°C revealed different stabilizing effects: in sTIM11noCys-SB and DeNovoTIM6-SB, a clear stabilization by 1.6 and 1.9 kcal mol⁻¹ was observed, respectively. In contrast, DeNovoTIM13-SB is destabilized through the introduced mutations by -1.3 kcal mol⁻¹. Nevertheless, also in DeNovoTIM13-SB the salt bridge cluster has a positive effect in the reduction of aggregation-propensity most likely by the change from an irreversible thermal unfolding process to a reversible one. Our results highlight the complexity of salt bridges in proteins: despite the high identity in sequence and structure of all three proteins, the similar salt bridge clusters have clearly different stabilizing effects.

In addition, we observed improvements on the crystallization properties of the *de novo* TIM barrels in comparison with the parental proteins. The structural analysis revealed highly diverse geometries for all three proteins, ranging from the absence of a cluster geometry and the formation of single salt bridges, via a water mediated cluster arrangement to a highly coordinated cluster network. Interestingly, the network geometry does not correlate with the corresponding stability. For instance, the crystal structure of DeNovoTIM6-SB revealed only the presence of two single salt bridges but the highest stabilizing effect. Due to these diverse influences of the salt bridge cluster on highly similar *de novo* TIM barrels, the influence of salt bridges could be studied in more detail to partition the stabilizing and destabilizing components under the same topology.

Most *de novo* protein design approaches lack intensive design of salt bridges and especially clusters, despite their proven importance for stability and function. Our analysis of the salt bridge TIM barrel variants indicate that Rosetta is able to predict influences on the stability with the right tendency. Nevertheless, the modeling of an accurate cluster geometry is still challenging in a fully automated approach. The engineering and design of salt bridge clusters in different natural and *de novo* proteins

would benefit from an improved understanding of salt bridges.

4 | MATERIALS AND METHODS

4.1 | Biochemicals

All reagents were analytical grade from Sigma-Aldrich or Carl Roth, except when indicated. All solutions were prepared with double-distilled water.

4.2 | Cloning, overexpression, and protein purification

All genes were synthesized and cloned into pET21b(+) vector by BioCat. *Escherichia coli*, BL21(DE3) (Novagen) were transformed with plasmids and used to inoculate LB precultures supplemented with ampicillin (100 µg ml⁻¹) which were grown at 37°C and 180 rpm overnight. Overexpression was performed in 1 L Terrific Broth (TB) cultures inoculated on OD₆₀₀ 0.08 and then grown at 37°C. At an OD₆₀₀ of 0.8–1 overexpression was induced with 1 mM isopropyl-D-1-thiogalactopyranoside and growth performed at 30°C for 4.5 hr. Afterward, cells were harvested by centrifugation (Beckmann Avanti JLA-8.1000, 15 min, 5,000g, 4°C) and pellets resuspended in 5 ml per gram pellet with buffer A: 35 mM sodium phosphate, 150 mM NaCl, 35 mM Imidazole, pH 8 (supplemented with 100 µl protease inhibitor [Mix-HP, Serva] per 10 ml lysate). Cells were lysed by sonication (Branson Ultrasonics) (output 4, duty cycle 40%, 2 times 2 min) and then centrifuged (Beckmann Avanti JA-25.50, 1 hr, 18,000 rpm, 4°C). The lysate was filtered with a 0.22 µm filter (Merck Millipore) and loaded onto a HisTrap HP column (5 ml, Cytiva) equilibrated with buffer A and coupled to an Äkta system (GE Healthcare Life Sciences). Unbound proteins were washed out with 20 column volumes (CVs) of buffer A. Elution of bound protein was performed with a linear gradient over 20 CV from 35 to 300 mM Imidazole using buffer B (35 mM sodium phosphate, 150 mM NaCl, 500 mM Imidazole, pH 8) followed by a step to 500 mM Imidazole for 5 CV. The peak fractions were pooled, concentrated and loaded onto a HiLoad 16/600 Superdex 500 preparative grade column (GE Healthcare Life Sciences) connected to an Äkta System. Elution was performed with 1 CV buffer C (35 mM sodium phosphate, 150 mM NaCl, pH 8) and the monomeric peak fractions were pooled and stored at room temperature or 4°C for use in subsequent

experiments. For some subsequent experiments the proteins were dialyzed into buffer D (10 mM sodium phosphate, pH 8).

4.3 | Analytical size exclusion chromatography-multi angle light scattering (SEC-MALS)

SEC-MALS measurements were performed using a Superdex 75 Increase 10/300 GL column (GE Healthcare Life Sciences) connected to an Äkta Pure System, and coupled to a miniDAWN multi-angle light scattering detector and an Optilab refractometer (WyattTechnology). All experiments were performed in buffer C with 0.02% sodium azide at room temperature and a flow rate of 0.8 ml min⁻¹, using a protein concentration of 1 and 5 mg ml⁻¹. Data collection and analysis were performed with the ASTRA 7.3.2 software (Wyatt Technology). To check for reproducibility during the SEC-MALS runs, BSA standard sample at 2 mg ml⁻¹ was measured at the beginning and end of each measurement day, obtaining identical results.

4.4 | Far-UV CD

CD spectra were collected in buffer D with a Jasco J-710 using a Peltier device to control the temperature (PTC-348 WI). Far-UV CD spectra were measured with a protein concentration of 0.2 mg ml⁻¹ in the wavelength range 195–260 nm at 25°C with a 1 nm bandwidth in a 2 mm cuvette. Spectra of thermally unfolded states were collected at 95°C. Data were normalized by subtraction of buffer spectra and then converted to mean residue ellipticity using: $[\theta_{\text{MRE}}] = (M \cdot \theta) / (10 \cdot d \cdot c)$ and $M = (MW/n - 1)$, where M is the mean residue weight, MW is the molecular weight in Da, n is the number of residues in the protein, θ is the collected ellipticity in mdeg, d is the path length in mm, and c is the protein concentration in mg ml⁻¹. Far-UV spectra were deconvoluted with CDNN.⁵¹

4.5 | Intrinsic fluorescence

Intrinsic fluorescence (IF) spectra were collected in buffer D with a protein concentration of 0.2 mg ml⁻¹ using a Jasco FP-6500 spectrofluorometer and a Peltier device to control the temperature (Julabo MB). Fluorescence was excited at a wavelength of 295 nm and emission was measured in the wavelength range 310–450 nm with a bandwidth of 1 nm. Spectra of unfolded protein

were measured at a Urea concentration capable of unfolding the protein. The spectral center of mass was calculated using: $\text{SCM} = \sum \lambda^2 I_\lambda$.

4.6 | Thermal unfolding followed by CD

Thermal unfolding was followed by CD at a protein concentration of 0.2 mg ml⁻¹ in buffer D in a 2 mm cuvette. The unfolding was followed in the temperature range 20–95°C at 222 nm with a heating rate of 1.5°C min⁻¹. Spectra were normalized to the fraction of unfolded molecules (f_u) by:

$$f_u = \frac{y_{\text{obs}} - (y_N + m_N T)}{(y_u + m_u T) - (y_N + m_N T)} \quad (1)$$

with y_{obs} the observed CD signal at a given temperature, and $(y_N + m_N T)$ and $(y_u + m_u T)$ the linear fitting equations of the native and unfolded regions, respectively.

4.7 | Thermal unfolding followed by DSC

Temperature-induced unfolding experiments by DSC were collected in a VP-Capillary DSC (Malvern Panalytical). Samples were assayed at 1.5°C min⁻¹ and protein concentration of 1, 2, 3, and 4 mg ml⁻¹ in buffer D, after exhaustive dialysis and buffer degassing. In all cases, proper equilibration was performed by running at least two buffer–buffer scans before sample–buffer experiments. The last buffer–buffer scan was subtracted from each protein–buffer scan to perform all thermodynamic analysis. Reversibility was determined by collecting a second endotherm after the first one was collected. For DeNovoTIM13-SB, endotherms were also collected at 1 mg mL⁻¹ and varying scan rate from 1 to 3°C min⁻¹. DSC scans were fitted to a two-state reversible model (Equation 2):

$$C_P(T) = B_0 + B_1 T + f(T) \Delta C_P + \frac{\Delta H(T)}{RT_m^2} \left[\frac{1 - f(T)}{1 - n + \frac{n}{f(T)}} \right] \quad (2)$$

where B_0 and B_1 are pre- and post-transition constants, n is the number of subunits in the native protein sample (monomer for all the proteins in this work) and $f(T)$ is the protein fraction in the folded monomeric state, yielding the parameters ΔH , ΔC_P , and T_m . To test the accuracy of the fitting, the DeNovoTIM3-SB endotherm at 1 mg ml⁻¹ was also fitted to an irreversible two-state model as indicated in Reference 27. Origin v.7.0 (OriginLab Corporation) with MicroCal software was used for data analysis.

Stability curves were constructed using DSC parameters and the Gibbs–Helmholtz equation:⁵²

$$\Delta G(T) = \Delta H \left(1 - \frac{T}{T_m} \right) - \Delta C_p \left(T_m - T + T \ln \left(\frac{T}{T_m} \right) \right) \quad (3)$$

4.8 | Chemical-induced unfolding followed by CD and IF

For chemical-induced unfolding experiments protein concentration was 0.2 mg ml⁻¹ in buffer D. Initially, the equilibrium time for chemical unfolding was determined by incubation of samples at different urea concentrations (0–9 M). CD and IF spectra were recorded at different incubation times and 2 days are sufficient for all analyzed proteins to reach equilibrium. Chemical unfolding experiments were carried out by incubation of samples with increasing urea concentration for 2 days at 25°C. For all urea concentrations the CD signal at 222 nm was measured for 2 min and IF spectra were recorded as aforementioned at 25°C. IF data were processed considering the intensity ratio at the wavelength of the maximum of the unfolded spectrum ($I_{\lambda u}$) and at the wavelength of the maximum of the native spectrum ($I_{\lambda n}$) at every urea concentration ($r_{\lambda} = \frac{I_{\lambda u}}{I_{\lambda n}}$). IF and CD data at every urea concentration were normalized to the fraction of unfolded protein using Equation (4), where y_{obs} is the experimentally observed CD signal or the calculated ratio of IF data at a given concentration, and ($y_N + m_N[\text{urea}]$) and ($y_U + m_U[\text{urea}]$) are the linear fitting equations of the native and unfolded regions, respectively.

$$f_U = \frac{y_{\text{obs}} - (y_N + m_N[\text{urea}])}{(y_U + m_U[\text{urea}]) - (y_N + m_N[\text{urea}])} \quad (4)$$

Determination of the unfolding free energy $\Delta G^{\text{H}_2\text{O}}$ was performed by fitting of the data to a two-state model ($\text{N} \rightleftharpoons \text{D}$) using the Santoro and Bolen equation (Equation 5)⁵³:

$$f_U = \frac{(y_N + m_N[\text{urea}]) + (y_U + m_U[\text{urea}]) * \exp\left(\frac{-\Delta G^{\text{H}_2\text{O}} - m[\text{urea}]}{RT}\right)}{1 + \exp\left(\frac{-\Delta G^{\text{H}_2\text{O}} - m[\text{urea}]}{RT}\right)} \quad (5)$$

where m is $\Delta G/[\text{urea}]$, a parameter related with the dependence of free energy on denaturant concentration and commonly associated with unfolding cooperativity, proportional to the surface area of protein exposed to solvent upon unfolding;^{54,55} T is the temperature of the

experiment (298.15 K), and R the universal gas constant (0.001987 kcal mol⁻¹ K⁻¹). In addition, the denaturant concentration at the midpoint of the unfolding curve, $D_{[1/2]}$ reported in Table 1, is equivalent to $D_{[1/2]} = \Delta G^{\text{H}_2\text{O}}/m$.⁵⁵ Data analysis and fitting was conducted with R⁵⁶ and graphs were created with package ggplot2.⁵⁷

4.9 | Crystallization and structure determination

For crystallization all proteins were in buffer C. Initial screening was performed with the sitting-drop vapor diffusion method using JCSG Core I-IV, Classics I-II, PEGs I-II (Qiagen) in 96-well Intelli plates (Art Robbins Instruments) using a nano dispensing crystallization robot Phoenix (Art Robbins Instruments). Crystallization drops with a volume of 0.8 μl were prepared with different ratios of mother liquid and protein (1:1, 1:2, 2:1). Screening plates were stored at 20°C in the hotel-based Rock Image RI 182 (Formulatrix). Crystallization hits were optimized using sitting and hanging drop vapor diffusion in MRC Maxi 48-well plates and VDXm 24-well plates, respectively, with a crystallization drop size of 2 μl . Initial screening was performed with 10 mg ml⁻¹ (sTIM11noCys-SB), 5.9, 8.6, 9, and 12 mg ml⁻¹ (DeNovoTIM6-SB) and 8.25 mg ml⁻¹ (DeNovoTIM13-SB).

In the following conditions good diffracting crystals were found: sTIM11noCys-SB (crystal form 1): 50% PEG 200, 0.2 M NaCl, 0.1 M phosphate citrate, pH 5, drop ratio 2:1 (protein: mother liquid); sTIM11noCys-SB (crystal form 2): 34% PEG 200, 0.1 M NaCl, 0.1 M Tris pH 7.78, 0.2 M lithium sulfate, drop ratio 1:1 (protein: mother liquid); both conditions with a protein concentration of 10 mg ml⁻¹. DeNovoTIM6-SB (crystal form 1): 0.2 M ammonium sulfate, 0.2 M sodium acetate, pH 4.6, 28% PEG 4000 with a drop ratio of 1:1; DeNovoTIM6-SB (crystal form 2): 0.2 M ammonium sulfate, 0.2 M sodium acetate, pH 4.3, 31% PEG 4000 with a drop ratio of 1:1; both conditions with a protein concentration of 8.6 mg ml⁻¹. DeNovoTIM13-SB: 0.17 M sodium acetate trihydrate, 0.085 M Tris, pH 8.9, 23% PEG 4000, 15% glycerol with a protein concentration of 8.4 mg ml⁻¹ and a drop ratio of 1:1.

For DeNovoTIM6-SB diffraction data were collected at 100 K at the Swiss Light Source at the Paul Scherrer Institute in Villigen (Switzerland) (PXI beamline) using a wavelength of 1.00 and a EIGER 16 M X Detector (Dectris).⁵⁸ For sTIM11noCys-SB and DeNovoTIM13-SB diffraction data were collected at 100 K at the Berlin Electron Storage Ring Society for Synchrotron Radiation beamline 14.1 and 14.2 (BESSY 14.1 and 14.2) operated

by Helmholtz-Zentrum Berlin using a wavelength of 0.9184 Å and a PILATUS3 S 6 M or PILATUS3S 2 M detector, respectively.⁵⁹

The datasets were processed with the X-ray detector software (XDS) using XDSAPP v3.0^{60,61} or command line. For DeNovoTIM13-SB two datasets of two different crystals from the exact same condition were merged using XSCALE to achieve a higher completeness in space group P1. Molecular replacement was performed with PHASER in the PHE-NIX software suite v.1.19.2⁶² using sTIM11noCys (PDB ID: 6YQY) as a starting model for sTIM11noCys-SB and DeNovoTIM13 (PDB ID: 6YQX) for DeNovoTIM6-SB and DeNovoTIM13-SB. Structure refinement was performed with phenix.refine⁶³ and iterative manual model improvement by rebuilding in COOT v.0.9.⁶⁴ Coordinates and structure factors were deposited in the PDB database <https://www.rcsb.org/>⁶⁵ with the accession codes: 7OSU (sTIM11noCys-SB, crystal form 1), 7OT7 (sTIM11noCys-SB, crystal form 2), 7OSV (DeNovoTIM6-SB, crystal form 1), 7OT8 (DeNovoTIM6-SB, crystal form 2), and 7P12 (DeNovoTIM13-SB). Secondary structure composition of crystal structures was calculated using the STRIDE Web Interface,⁶⁶ (<http://webclu.bio.wzw.tum.de/cgi-bin/stride/stridecgi.py>). The figures were created using PyMOL Molecular Graphics System v.4.6.0 (Schrodinger, LLC).

4.10 | Geometric analysis of the salt bridge cluster

In detail analysis of the salt bridge cluster geometries was performed in PyMol (Schrodinger, LLC). Distances reported were calculated by measuring the distance between the corresponding heavy atoms. Measurements of the angles for a certain salt bridge were performed as previously suggested¹⁰ measuring the angle between $\angle(\text{Arg-N}^{\epsilon}, \text{Arg-C}^{\zeta}, \text{Glu-O}^{\epsilon})$. In case a water molecule was involved in the salt bridge, either $\angle(\text{Arg-N}^{\epsilon}, \text{Arg-C}^{\zeta}, \text{H}_2\text{O})$ or $\angle(\text{Glu-C}^{\delta}, \text{Glu-O}^{\epsilon}, \text{H}_2\text{O})$ were determined depending on the involved residue type.

4.11 | Rosetta calculations

Crystal structures of the parental proteins (sTIM11noCys—6YQY, DeNovoTIM6—6YQX, DeNovoTIM13—6Z21) as well as of the salt bridge cluster variants (sTIM11noCys-SB—7OSU, DeNovoTIM6-SB—7OSV, DeNovoTIM13-SB—7P12) were scored with Rosetta using the *ref2015* scoring function.⁴⁹ The PDB structures were initially cleaned with the *clean_pdb.py* script from Rosetta tools followed by a relax with constraining the structure, backbone and side chains, to input coordinates.⁵⁰

Rosetta models were created by mutation of the salt bridge cluster residues using Remodel.⁴⁸ As a starting model, for sTIM11noCys-SB and DeNovoTIM13-SB the parental protein structures were used, whereas for NovoTIM6 the previously created Rosetta model was used.²⁷ Subsequently, the models and crystal structures of the parental as well as of the salt bridge cluster variants were relaxed via an iterative approach until no further decrease of the score was observed.

ACKNOWLEDGMENTS

We acknowledge financial support and allocation of beamtime by HZB and thank the beamline staff at BESSY and SLS for assistance. We thank Sabrina Wischt for her competent technical support and appreciate the feedback and suggestions from Michael Weyand and Sooruban Shanmugaratnam on crystallization and structure determination. This work was supported by scholarships from Foundations Alexander von Humboldt and Bayer Science & Education (Humboldt-Bayer Research Fellowship for Postdoctoral Researchers to Sergio Romero-Romero). Birte Höcker gratefully acknowledges financial support by the European Research Council (ERC Consolidator Grant 647548 “Protein Lego”). Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors declare no potential conflict of interest.

AUTHOR CONTRIBUTIONS

Sina Kordes: Conceptualization (equal); data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); software (equal); supervision (equal); validation (equal); visualization (equal); writing – original draft (equal); writing – review and editing (equal). **Sergio Romero-Romero:** Conceptualization (equal); data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); supervision (equal); validation (equal); visualization (equal); writing – original draft (equal); writing – review and editing (equal). **Leonie Lutz:** Formal analysis (equal); investigation (equal); methodology (equal); writing – review and editing (equal). **Birte Höcker:** Conceptualization (equal); funding acquisition (equal); project administration (equal); resources (equal); supervision (equal); writing – original draft (equal); writing – review and editing (equal).

DATA AVAILABILITY STATEMENT

All data to support the conclusions of this manuscript are included in the main text and supporting information. Coordinates and structure factors have been deposited to the Protein Data Bank (PDB) with accession codes: 7OSU (sTIM11noCys-SB, crystal form 1), 7OT7 (sTIM11noCys-

SB, crystal form 2), 7OSV (DeNovoTIM6-SB, crystal form 1), 7OT8 (DeNovoTIM6-SB, crystal form 2), 7P12 (DeNovoTIM13-SB).

ORCID

Sina Kordes  <https://orcid.org/0000-0001-7489-3380>

Sergio Romero-Romero  <https://orcid.org/0000-0003-2144-7912>

Leonie Lutz  <https://orcid.org/0000-0003-1078-645X>

Birte Höcker  <https://orcid.org/0000-0002-8250-9462>

REFERENCES

1. Goldenzweig A, Fleishman SJ. Principles of protein stability and their application in computational design. *Annu Rev Biochem.* 2018;87:105–129.
2. Nisthal A, Wang CY, Ary ML, Mayo SL. Protein stability engineering insights revealed by domain-wide comprehensive mutagenesis. *Proc Natl Acad Sci U S A.* 2019;116:16367–16377.
3. Bloom JD, Labthavikul ST, Otey CR, Arnold FH. Protein stability promotes evolvability. *Proc Natl Acad Sci U S A.* 2006;103:5869–5874.
4. Huang PS, Boyken SE, Baker D. The coming of age of de novo protein design. *Nature.* 2016;537:320–327.
5. Islam MM, Kobayashi K, Kidokoro SI, Kuroda Y. Hydrophobic surface residues can stabilize a protein through improved water–protein interactions. *FEBS J.* 2019;286:4122–4134.
6. Goldenzweig A, Goldsmith M, Hill SE, et al. Automated structure- and sequence-based design of proteins for high bacterial expression and stability. *Mol Cell.* 2016;63:337–346.
7. Kwok SC, Hodges RS. Clustering of large hydrophobes in the hydrophobic core of two-stranded α -helical coiled-coils controls protein folding and stability. *J Biol Chem.* 2003;278:35248–35254.
8. Malakauskas SM, Mayo SL. Design, structure and stability of a hyperthermophilic protein variant. *Nat Struct Biol.* 1998;5:470–475.
9. Fersht AR, Serrano L. Principles of protein stability derived from protein engineering experiments. *Curr Opin Struct Biol.* 1993;3:75–83.
10. Donald JE, Kulp DW, DeGrado WF. Salt bridges: Geometrically specific, designable interactions. *Proteins.* 2011;79:898–915.
11. Kumar S, Nussinov R. Salt bridge stability in monomeric proteins. *J Mol Biol.* 1999;293:1241–1255.
12. Barlow DJ, Thornton JM. Ion-pairs in proteins. *J Mol Biol.* 1983;168:867–885.
13. Bosshard HR, Marti DN, Jelesarov I. Protein stabilization by salt bridges: Concepts, experimental approaches and clarification of some misunderstandings. *J Mol Recog.* 2004;17:1–16.
14. Kursula I, Partanen S, Lambeir AM, Wierenga RK. The importance of the conserved Arg191-Asp227 salt bridge of triosephosphate isomerase for folding, stability, and catalysis. *FEBS Lett.* 2002;518:39–42.
15. Lounnas V, Wade RC. Exceptionally stable salt bridges in cytochrome P450cam have functional roles. *Biochemistry.* 1997;36:5402–5417.
16. Hendsch ZS, Tidor B. Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci.* 1994;3:211–226.
17. Dong F, Vijayakumar M, Zhou HX. Comparison of calculation and experiment implicates significant electrostatic contributions to the binding stability of barnase and barstar. *Biophys J.* 2003;85:49–60.
18. Strop P, Mayo SL. Contribution of surface salt bridges to protein stability. *Biochemistry.* 2000;39:1251–1255.
19. Marqusee S, Sauer RT. Contributions of a hydrogen bond/salt bridge network to the stability of secondary and tertiary structure in lambda repressor. *Protein Sci.* 1994;3:2217–2225.
20. Sun DP, Sauer U, Nicholson H, Matthews BW. Contributions of engineered surface salt bridges to the stability of T4 lysozyme determined by directed mutagenesis. *Biochemistry.* 1991;30:7142–7153.
21. Fersht AR. Conformational equilibria in α - and δ -chymotrypsin. The energetics and importance of the salt bridge. *J Mol Biol.* 1972;64:497–509.
22. Basu S, Mukharjee D. Salt-bridge networks within globular and disordered proteins: Characterizing trends for designable interactions. *J Mol Model.* 2017;23:206.
23. Missimer JH, Steinmetz MO, Baron R, et al. Configurational entropy elucidates the role of salt-bridge networks in protein thermostability. *Protein Sci.* 2007;16:1349–1359.
24. Romero-Romero S, Kordes S, Michel F, Höcker B. Evolution, folding, and design of TIM barrels and related proteins. *Curr Opin Struct Biol.* 2021;68:94–104.
25. Sterner R, Höcker B. Catalytic versatility, stability, and evolution of the (beta α)₈-barrel enzyme fold. *Chem Rev.* 2005;105:4038–4055.
26. Huang PS, Feldmeier K, Parmeggiani F, Velasco D, Höcker B, Baker D. De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nat Chem Biol.* 2016;12:29–34.
27. Romero-Romero S, Costas M, Silva Manzano DA, et al. The stability landscape of de novo TIM barrels explored by a modular design approach. *J Mol Biol.* 2021;433:167153.
28. Lang D, Thoma R, Henn-Sax M, Sterner R, Wilmanns M. Structural evidence for evolution of the beta/alpha barrel scaffold by gene duplication and fusion. *Science.* 2000;289:1546–1550.
29. Chaudhuri BN, Lange SC, Myers RS, Chittur SV, Davisson VJ, Smith JL. Crystal structure of imidazole glycerol phosphate synthase: A tunnel through a (beta/alpha)₈ barrel joins two active sites. *Structure.* 2001;9:987–997.
30. Seitz T, Bocola M, Claren J, Sterner R. Stabilisation of a (beta α)₈-barrel protein designed from identical half barrels. *J Mol Biol.* 2007;372:114–129.
31. Höcker B, Claren J, Sterner R. Mimicking enzyme evolution by generating new (beta α)₈-barrels from (beta α)₄-half-barrels. *Proc Natl Acad Sci U S A.* 2004;101:16448–16453.
32. Thóroflsson M, Ibarra-Molero B, Fojan P, Petersen SB, Sanchez-Ruiz JM, Martínez A. L-phenylalanine binding and domain organization in human phenylalanine hydroxylase: A differential scanning calorimetry study. *Biochemistry.* 2002;41:7573–7585.
33. Ibarra-Molero B, Naganathan AN, Sanchez-Ruiz JM, Muñoz V. Modern analysis of protein folding by differential scanning calorimetry. *Methods Enzymol.* 2016;567:281–318.
34. Hennig M, Darimont B, Sterner R, Kirschner K, Jansonius JN. 2.0 a structure of indole-3-glycerol phosphate synthase from

- the hyperthermophile *Sulfolobus solfataricus*: Possible determinants of protein stability. *Structure*. 1995;3:1295–1306.
35. Yip KS, Stillman TJ, Britton KL, et al. The structure of *Pyrococcus furiosus* glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. *Structure*. 1995;3:1147–1158.
 36. Rüegg C, Ammer D, Lerch K. Comparison of amino acid sequence and thermostability of tyrosinase from three wild type strains of *Neurospora crassa*. *J Biol Chem*. 1982;257:6420–6426.
 37. Perutz MF. Electrostatic effects in proteins. *Science*. 1978;201:1187–1191.
 38. Elcock AH. The stability of salt bridges at high temperatures: Implications for hyperthermophilic proteins. *J Mol Biol*. 1998;284:489–502.
 39. Kumar S, Ma B, Tsai CJ, Nussinov R. Electrostatic strengths of salt bridges in thermophilic and mesophilic glutamate dehydrogenase monomers. *Proteins*. 2000;38:368–383.
 40. Xiao L, Honig B. Electrostatic contributions to the stability of hyperthermophilic proteins. *J Mol Biol*. 1999;289:1435–1444.
 41. Serrano L, Horovitz A, Avron B, Bycroft M, Fersht AR. Estimating the contribution of engineered surface electrostatic interactions to protein stability by using double-mutant cycles. *Biochemistry*. 1990;29:9343–9352.
 42. Gvritshvili AG, Gribenko AV, Makhatazde GI. Cooperativity of complex salt bridges. *Protein Sci*. 2008;17:1285–1290.
 43. Sindelar CV, Hendsch ZS, Tidor B. Effects of salt bridges on protein structure and design. *Protein Sci*. 1998;7:1898–1914.
 44. Horovitz A, Serrano L, Avron B, Bycroft M, Fersht AR. Strength and co-operativity of contributions of surface salt bridges to protein stability. *J Mol Biol*. 1990;216:1031–1044.
 45. Sarakatsannis JN, Duan Y. Statistical characterization of salt bridges in proteins. *Proteins*. 2005;60:732–739.
 46. Musafia B, Buchner V, Arad D. Complex salt bridges in proteins: Statistical analysis of structure and function. *J Mol Biol*. 1995;254:761–770.
 47. Singh J, Thornton JM, Snarey M, Campbell SF. The geometries of interacting arginine-carboxyls in proteins. *FEBS Lett*. 1987;224:161–171.
 48. Huang PS, Ban YE, Richter F, et al. RosettaRemodel: A generalized framework for flexible backbone protein design. *PLoS One*. 2011;6:e24109.
 49. Alford RF, Leaver-Fay A, Jeliakov JR, et al. The Rosetta all-atom energy function for macromolecular modeling and design. *J Chem Theory Comput*. 2017;13:3031–3048.
 50. Conway P, Tyka MD, DiMaio F, Komerding DE, Baker D. Relaxation of backbone bond geometry improves protein energy landscape modeling. *Protein Sci*. 2014;23:47–55.
 51. Böhm G, Muhr R, Jaenicke R. Quantitative analysis of protein far UV circular dichroism spectra by neural networks. *Protein Eng*. 1992;5:191–195.
 52. Kumar S, Tsai CJ, Nussinov R. Maximal stabilities of reversible two-state proteins. *Biochemistry*. 2002;41:5359–5374.
 53. Santoro MM, Bolen DW. Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl α -chymotrypsin using different denaturants. *Biochemistry*. 1988;27:8063–8068.
 54. Myers JK, Pace CN, Scholtz JM. Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Sci*. 1995;4:2138–2148.
 55. Pace CN. Measuring and increasing protein stability. *Trends Biotechnol*. 1990;8:93–98.
 56. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2020.
 57. Wickham H. *Ggplot2: Elegant graphics for data analysis*. New York, NY: Springer-Verlag, 2016.
 58. Mueller M, Wang M, Schulze-Briese C. Optimal fine ϕ -slicing for single-photon-counting pixel detectors. *Acta Cryst D*. 2012;68:42–56.
 59. Helmholtz-Zentrum Berlin für Materialien und Energie. The MX Beamlines BL14.1-3 at BESSY II. *J Large-scale Res Facil* 2016;2:A47.
 60. Kabsch W. XDS. *XDS Acta Cryst D*. 2010;66:125–132.
 61. Sparta KM, Krug M, Heinemann U, Mueller U, Weiss MS. XDSAPP2.0. *J Appl Cryst*. 2016;49:1085–1092.
 62. Liebschner D, Afonine PV, Baker ML, et al. Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Cryst D*. 2019;75:861–877.
 63. Adams PD, Afonine PV, Bunkóczi G, et al. PHENIX: A comprehensive python-based system for macromolecular structure solution. *Acta Cryst D*. 2010;66:213–221.
 64. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Cryst D*. 2010;66:486–501.
 65. Berman HM, Battistuz T, Bhat TN, et al. The Protein Data Bank. *Acta Cryst D*. 2002;58:899–907.
 66. Heinig M, Frishman D. STRIDE: A web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res*. 2004;32:W500–W502.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Kordes S, Romero-Romero S, Lutz L, Höcker B. A newly introduced salt bridge cluster improves structural and biophysical properties of *de novo* TIM barrels. *Protein Science*. 2021;1–15. <https://doi.org/10.1002/pro.4249>

Supplementary material for:

A newly introduced salt bridge cluster improves structural and biophysical properties of *de novo* TIM barrels

Sina Kordes,^{1†} Sergio Romero-Romero,^{1†} Leonie Lutz,¹ Birte Höcker^{1*}

¹ Department of Biochemistry, University of Bayreuth, 95447 Bayreuth, Germany.

† These authors contributed equally to the work.

* Corresponding author: Birte Höcker. Department of Biochemistry, University of Bayreuth, 95447 Bayreuth, Germany, e-mail address: birte.hoecker@uni-bayreuth.de

This file includes:

- Supplementary Tables 1-4.
- Supplementary Figures 1-7.

Supplementary Tables

Supplementary Table 1. Amino acid sequences of parental DeNovoTIMs and salt bridge cluster variants.

<i>de novo</i> TIM barrel	TIM barrel properties	Sequence
sTIM11noCys	Construct derived from the first validated <i>de novo</i> TIM barrel sTIM11 with the two cysteines reverted to the residues in the original four-fold design (C8Q and C181V). This TIM barrel does not contain any extra stabilizing mutation as present on DeNovoTIMs collection.	MDKDEAWKQVEQLRREGATQIAYRSDDWRDLKEAWKKGADILIVDAT DKDEAWKQVEQLRREGATQIAYRSDDWRDLKEAWKKGADILIVDAT DKDEAWKQVEQLRREGATQIAYRSDDWRDLKEAWKKGADILIVDAT DKDEAWKQVEQLRREGATQIAYRSDDWRDLKEAWKKGADILIVDATGLEHHHHHH
sTIM11noCys-SB	Parental protein sTIM11noCys + mutations to introduce a salt bridge cluster.	MDKDEAWKQVEQLRREGATRIAYRSDDWRDLKEAWKKGADILIVDAT DKDEAWKQVEQLRREGATEIAYRSDDWRDLKEAWKKGADILIVDAT DKDEAWKQVEQLRREGATRIAYRSDDWRDLKEAWKKGADILIVDAT DKDEAWKQVEQLRREGATEIAYRSDDWRDLKEAWKKGADILIVDATGLEHHHHHH
DeNovoTIM6	Construct derived from sTIM11noCys with stabilizing mutations in the bottom peripheral core formed by the outer face of the strands and the internal face of the helices. Stabilizing mutations in each quarter: Q11I, E15L, T18K, K31Q, V34L.	MDKDEAWKQVEILRRLGAKQIAYRSDDWRDLQEALKGGDILIVDAT DKDEAWKQVEILRRLGAKQIAYRSDDWRDLQEALKGGDILIVDAT DKDEAWKQVEILRRLGAKQIAYRSDDWRDLQEALKGGDILIVDAT DKDEAWKQVEILRRLGAKQIAYRSDDWRDLQEALKGGDILIVDATGLEHHHHHH
DeNovoTIM6-SB	Parental protein DeNovoTIM6 + mutations to introduce a salt bridge cluster.	MDKDEAWKQVEILRRLGAKRIAYRSDDWRDLQEALKGGADILIVDAT DKDEAWKQVEILRRLGAKIEAYRSDDWRDLQEALKGGADILIVDAT DKDEAWKQVEILRRLGAKRIAYRSDDWRDLQEALKGGADILIVDAT DKDEAWKQVEILRRLGAKIEAYRSDDWRDLQEALKGGADILIVDATGLEHHHHHH
DeNovoTIM13	Construct derived from DeNovoTIM6 with stabilizing mutations in the top and bottom peripheral core formed by the outer face of the strands and the internal face of the helices. Stabilizing mutations in each half (included those already present on DeNovoTIM6): K2V, A5M, W6L, Q11I, E15L, T18K, Y22V, D29I, K31Q, V34L.	MDVDEMLKQVEILRRLGAKQIAYRSDDWRILQEALKGGDILIVDAT DVDEMLKQVEILRRLGAKQIAYRSDDWRILQEALKGGDILIVDAT DVDEMLKQVEILRRLGAKQIAYRSDDWRILQEALKGGDILIVDAT DVDEMLKQVEILRRLGAKQIAYRSDDWRILQEALKGGDILIVDATGLEHHHHHH
DeNovoTIM13-SB	Parental protein DeNovoTIM13 + mutations to introduce a salt bridge cluster.	MDVDEMLKQVEILRRLGAKRIAYRSDDWRILQEALKGGDILIVDAT DVDEMLKQVEILRRLGAKIEAYRSDDWRILQEALKGGDILIVDAT DVDEMLKQVEILRRLGAKRIAYRSDDWRILQEALKGGDILIVDAT DVDEMLKQVEILRRLGAKIEAYRSDDWRILQEALKGGDILIVDATGLEHHHHHH

Supplementary Table 2. Biochemical and biophysical properties of salt bridge cluster variants in comparison with the parental proteins.

<i>de novo</i> TIM barrel	Expression/Solubility properties		Spectroscopic properties						Hydrodynamic properties		
			Circular dichroism Predicted secondary structure content			Intrinsic Fluorescence					
	Soluble overexpression	Aggregation-prone	α -helix	β -strand	Random coil	Native λ_{\max} (nm)	Unfolded λ_{\max} (nm)	$\Delta\lambda_{\max N-U}$ (nm)	Theoretical MW (kDa)	Experimental MW (kDa)	Exp./Theor. Ratio
sTIM11noCys*	+++	+	47.6 (49.4)	19.2 (21.9)	33.2 (28.7)	338	349	11	22.9	20.7 ± 0.1	0.9
sTIM11noCys-SB	+++	+	48.3 (48.6)	19.0 (21.3)	32.7 (30.1)	345	353	8	22.9	21.2 ± 1.2	0.9
DeNovoTIM6*	+++	+	46.5 (47.0)	18.8 (20.8)	34.8 (32.1)	336	350	14	22.6	22.0 ± 0.4	1.0
DeNovoTIM6-SB	+++	+	44.2 (45.7)	21.0 (23.1)	34.9 (31.2)	343	351	8	22.6	21.0 ± 1.1	0.9
DeNovoTIM13*	++	+++	48.3 (53.0)	18.1 (21.0)	33.5 (26.0)	334	351	17	22.1	21.6 ± 0.7	1.0
DeNovoTIM13-SB	+++	+	50.2 (52.2)	16.9 (20.7)	32.8 (27.2)	341	352	11	22.1	19.1 ± 1.8	0.9

* Parameters reported in ref. 27.

Numbers in brackets in the CD data are the calculated values from the three-dimensional structure.

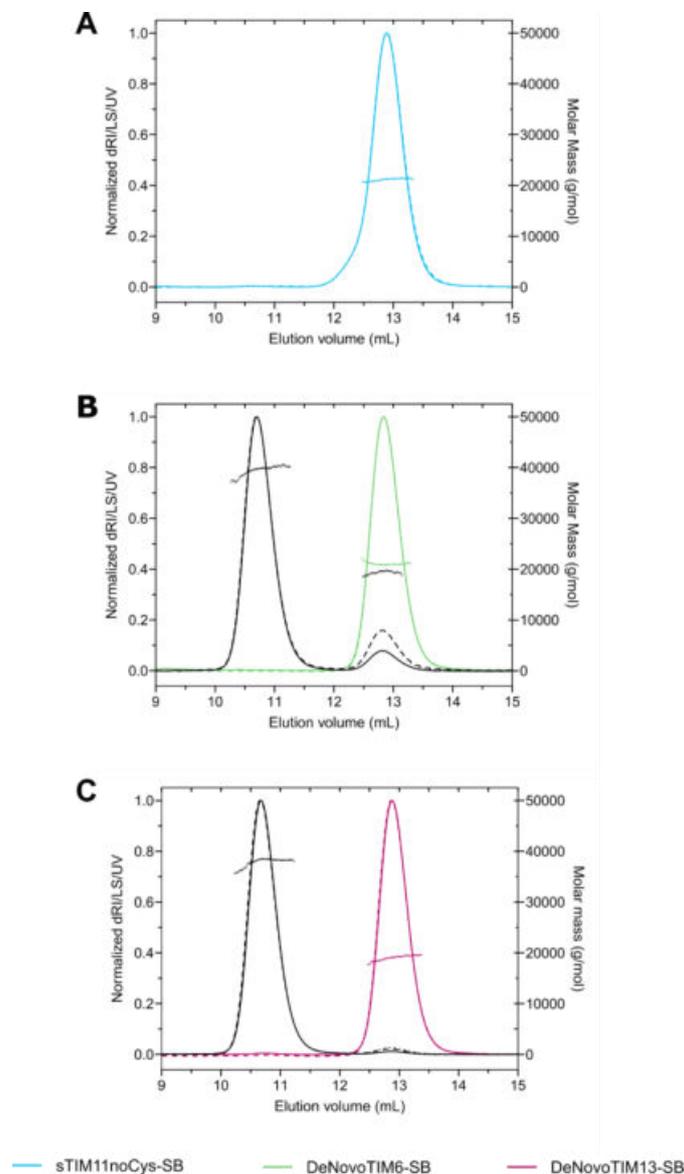
Supplementary Table 3. Crystallographic data collection and refinement statistics for the salt bridge cluster variants. Statistics for the highest resolution shell are shown in parentheses.

Protein	STIM11noCys-SB (crystal form 1)	STIM11noCys-SB (crystal form 2)	DeNovoTIM6-SB (crystal form 1)	DeNovoTIM6-SB (crystal form 2)	DeNovoTIM13-SB
PDB ID	7OSU	7OT7	7OSV	7OT8	7P12
Wavelength	0.9184	0.9184	1.0000	1.0000	0.9184
Resolution range	47.3 – 1.37 (1.42 – 1.37)	47.33 – 1.513 (1.57 – 1.51)	32.81 – 1.66 (1.72 – 1.66)	42.56 – 2.223 (2.31 – 2.22)	32.65 – 1.69 (1.75 – 1.69)
Space group	P 41 21 2	P 41 21 2	P 1 21 1	P 2 21 21	P 1
Unit cell [a, b, c (Å) / α , β , γ (°)]	50.72 50.72 131.09 / 90 90 90	50.70 50.70 132.10 / 90 90 90	29.37 65.63 44.61 / 90 96 90	37.86 74.90 155.15 / 90 90 90	39.40 39.47 39.40 / 113 102 113
Total reflections	518976 (50251)	384388 (30364)	136326 (13716)	290597 (27974)	136123 (13697)
Unique reflections	36942 (3614)	27801 (2677)	19751 (1947)	22258 (2081)	20418 (1981)
Multiplicity	14.1 (13.9)	13.8 (11.3)	6.9 (7.0)	13.1 (13.4)	6.8 (6.9)
Completeness (%)	99.92 (99.92)	99.87 (99.04)	98.89 (97.45)	98.38 (94.97)	99.67 (99.85)
Mean I/ σ (I)	13.30 (0.94)	18.90 (1.01)	7.77 (0.96)	10.90 (0.95)	9.66 (1.13)
Wilson B-factor	23.47	28.66	27.41	47.03	38.3
R-merge	0.0870 (2.04)	0.0596 (1.25)	0.1385 (1.77)	0.1409 (1.18)	0.1995 (2.80)
R-meas	0.0904 (2.11)	0.0619 (1.35)	0.1501 (1.99)	0.1471 (1.26)	0.2165 (3.02)
R-pim	0.024 (0.558)	0.017 (0.683)	0.057 (1.114)	0.041 (0.606)	0.083 (1.125)
CC1/2	0.995 (0.660)	0.999 (0.425)	0.997 (0.339)	0.998 (0.464)	0.987 (0.596)
CC*	0.999 (0.684)	1 (0.772)	0.999 (0.712)	1 (0.796)	0.997 (0.864)
Matthews coefficient V_m (Å ³ Da ⁻¹)	1.93	1.94	1.96	2.52	2.25
Solvent content (%)	36.4	36.8	37.3	51.2	45.4
Protein molecules per asymmetric unit	1	1	1	2	1
Reflections used in refinement	36920 (3613)	27787 (2676)	19731 (1946)	22197 (2060)	20140 (1980)
Reflections used for R-free	1847 (181)	1390 (134)	986 (97)	1066 (93)	1008 (99)
R-work	0.208 (0.359)	0.217 (0.351)	0.210 (0.392)	0.271 (0.368)	0.266 (0.414)
R-free	0.247 (0.350)	0.247 (0.362)	0.233 (0.394)	0.323 (0.420)	0.315 (0.486)
CC(work)	0.962 (0.784)	0.956 (0.642)	0.959 (0.658)	0.923 (0.526)	0.946 (0.713)
CC(free)	0.938 (0.636)	0.951 (0.694)	0.921 (0.636)	0.916 (0.486)	0.857 (0.469)
Number of non-hydrogen atoms	1598	1513	1495	2601	1279
macromolecules	1484	1430	1359	2542	1252
ligands	14	0	36	10	6
solvent	100	83	116	49	21
Protein residues	183	183	173	346	184
RMS(bonds)	0.005	0.010	0.006	0.002	0.004
RMS(angles)	0.73	1.00	0.60	0.40	0.69
Ramachandran favored (%)	98.9	95.5	100.0	96.7	98.4
Ramachandran allowed (%)	1.1	4.5	0.0	3.3	1.6
Ramachandran outliers (%)	0.0	0.0	0.0	0.0	0
Rotamer outliers (%)	0.8	0.8	0.8	0.0	0
Clashscore	3.5	8.1	4.8	2.5	5.9
Average B-factor	36.6	40.8	33.7	56.4	59.1
macromolecules	35.7	40.4	32.5	56.1	58.8
ligands	71.5		67.7	86.1	95.6
solvent	45.9	47.9	41.1	62.3	64.8
Number of TLS groups	5	1	1	1	1

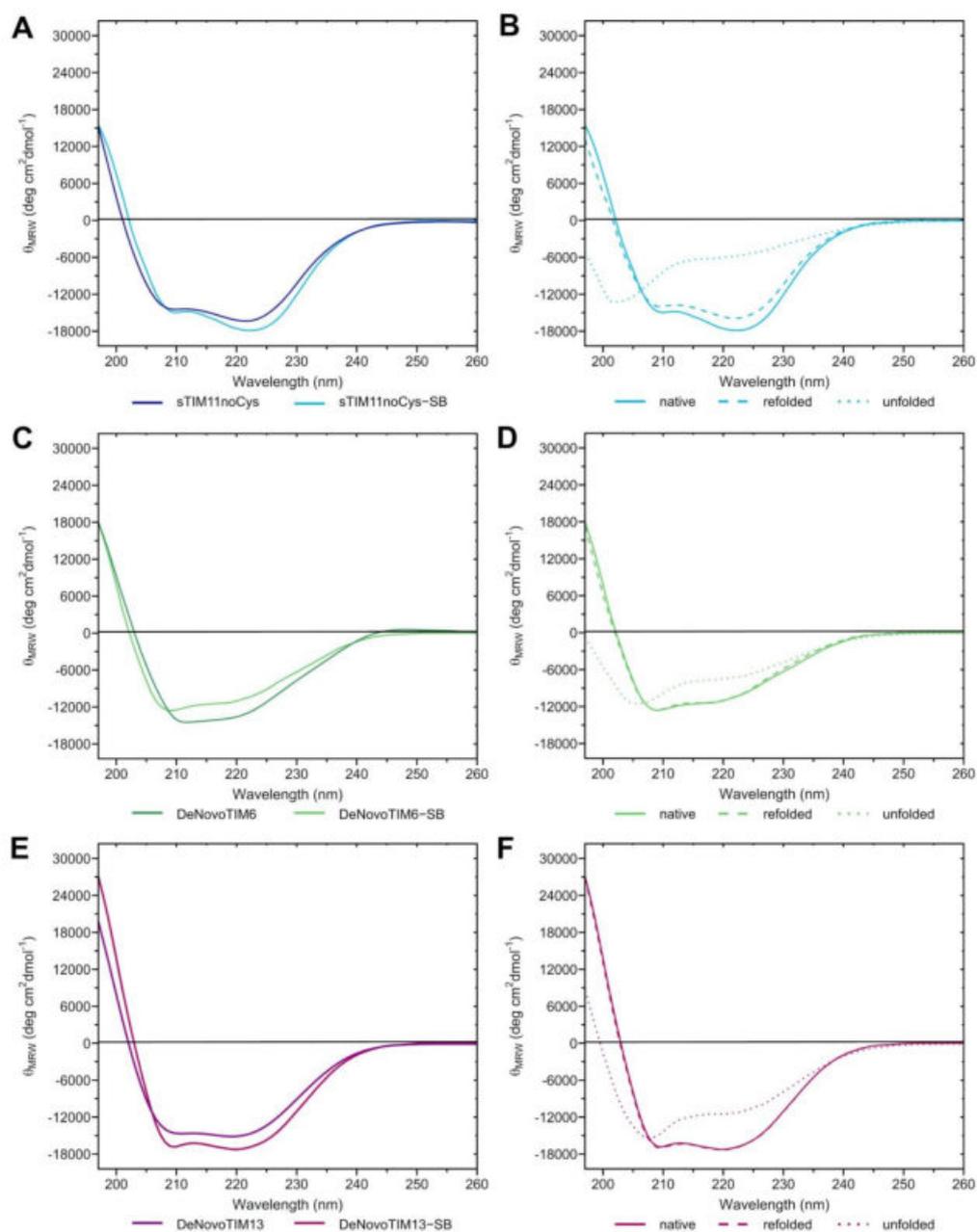
Supplementary Table 4. Rosetta scores for the salt bridge cluster variants in comparison with the parental proteins.

<i>de novo</i> TIM barrel	total score (REU)	normalized score (REU per residue)
sTIM11noCys	-231.32	-1.30
sTIM11noCys-SB	-463.32	-2.53
DeNovoTIM6	-20.52	-0.12
DeNovoTIM6-SB	-258.14	-1.49
DeNovoTIM13	-446.71	-2.47
DeNovoTIM13-SB	-409.61	-2.24

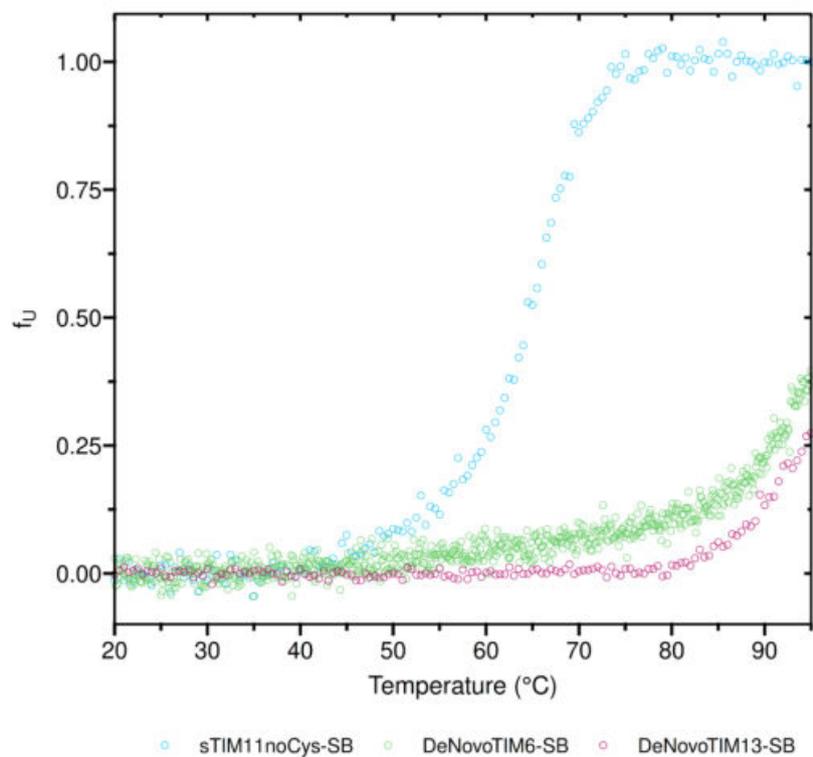
Supplementary Figures



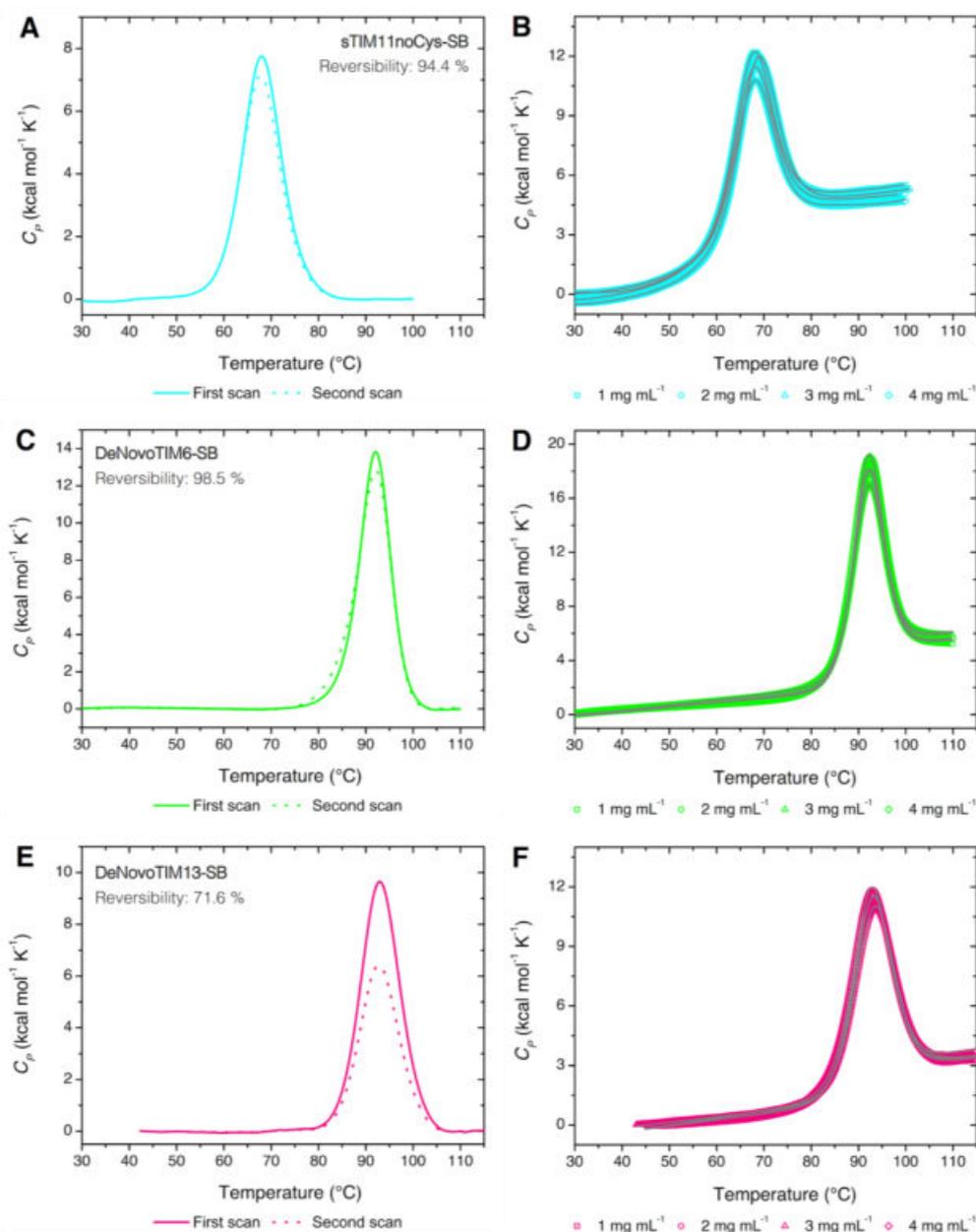
Supplementary Figure 1. SEC-MALS of the salt bridge cluster variants. **A)** Monomeric peak for sTIM1noCys-SB, **B)** Dimer (black lines) and monomer peak (green lines) for DeNovoTIM6-SB, **C)** Dimer (black lines) and monomer peak (pink lines) for DeNovoTIM13-SB. All plots show the elution volume versus either the normalized data for UV absorbance at 280 nm (UV, solid line), the differential refractive index (dRI, dashed line) and the multi-angle light scattering (LS, dotted line) (left axis), or the calculated molar mass (right axis, dot symbols). Values derived from the experiments are reported in the Supplementary Table 2. All experiments were performed in 35 mM sodium phosphate, 150 mM NaCl, 0.02% sodium azide, pH 8.



Supplementary Figure 2. Far-UV CD spectra of the salt bridge variants. The comparison between the parental proteins and the salt bridge variants in their native form are shown in panels **A)** (sTIM11noCys/sTIM11noCys-SB), **C)** (DeNovoTIM6/DeNovoTIM6-SB), and **E)** (DeNovoTIM13/DeNovoTIM13-SB). The experimental data for the native protein (solid lines), the unfolded protein at 95 °C (dotted lines), and the refolded protein after the T-melt (dashed lines) for the salt bridge variants are shown in panels **B)** (sTIM11noCys-SB), **D)** (DeNovoTIM6-SB), and **F)** (DeNovoTIM13-SB). All experiments were performed in 10 mM sodium phosphate pH 8.



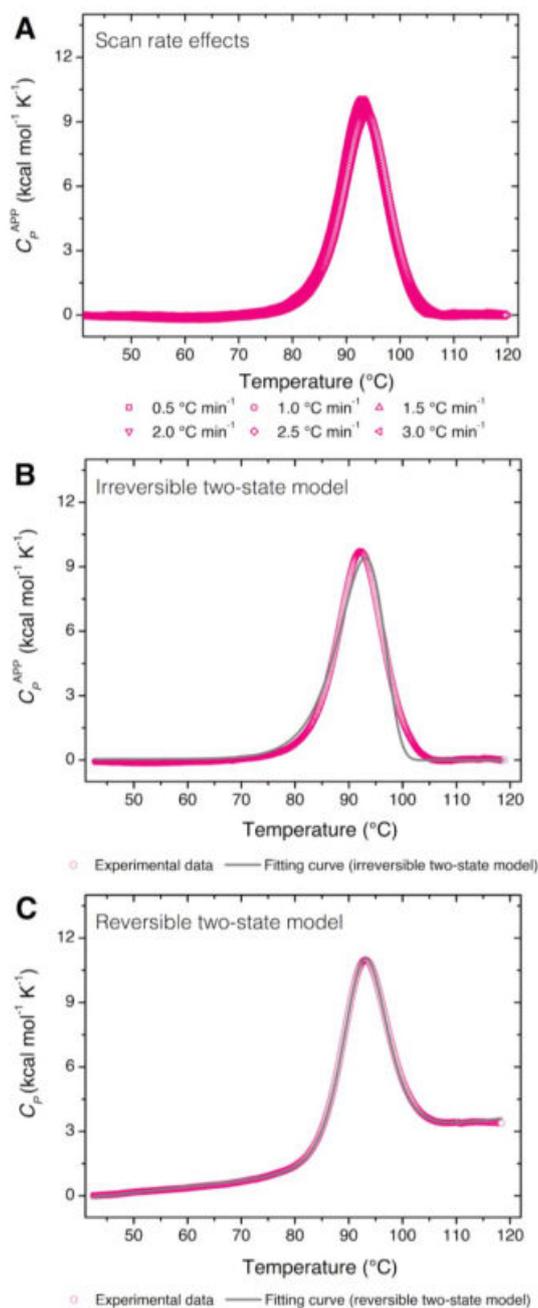
Supplementary Figure 3. Far-UV CD T-melt of the salt bridge variants. All experiments were collected following the CD signal at 222 nm with a protein concentration of 0.2 mg mL^{-1} , a scan rate of $1.5 \text{ }^\circ\text{C min}^{-1}$, and 10 mM sodium phosphate pH 8.



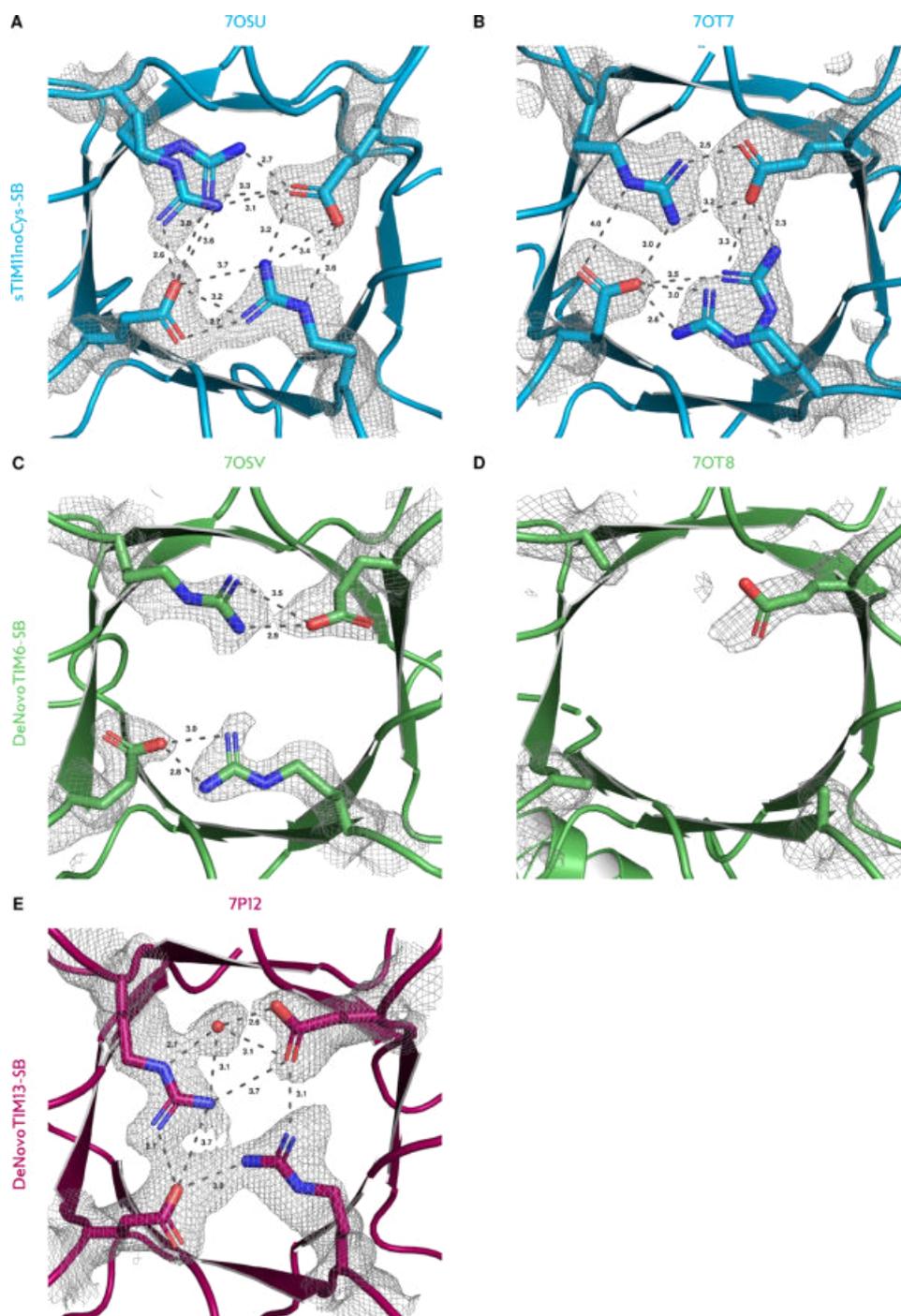
Supplementary Figure 4. Thermal unfolding of the salt bridge variants followed by DSC.

Thermal unfolding reversibility was determined by the recovered area percentage by comparing the first and second scan (continuous and dotted lines, respectively in panels A, C, E).

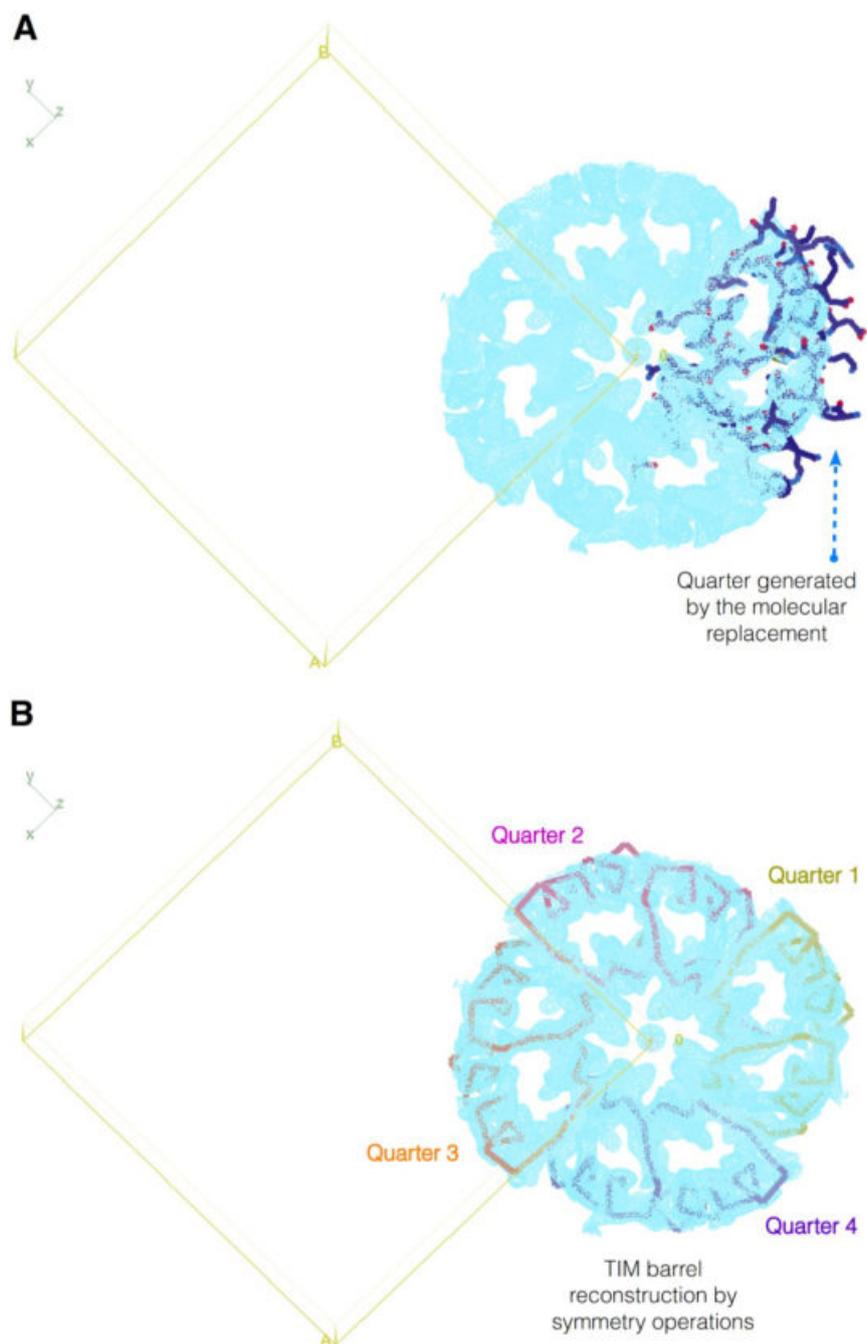
Thermodynamic parameters were calculated fitting the endotherms collected at different protein concentrations (open symbols) to a reversible two-state model (continuous lines in panels B, D, F). In panels A, C, and E, protein concentration was 1.0 mg mL⁻¹. All scans were collected in 10 mM sodium phosphate pH 8.



Supplementary Figure 5. Thermal unfolding of DeNovoTIM13-SB followed by DSC. A) Scan rate effects on the melting temperature. Endotherms were collected at scan rates from 0.5 to 3.0 °C min⁻¹. **B)** Fitting example of an endotherm collected at 1.5 °C min⁻¹ (symbols) fitted to the irreversible two-state model (continuous line). **C)** Fitting example of an endotherm collected at 1.5 °C min⁻¹ (symbols) fitted to the reversible two-state model (continuous line). In all experiments protein concentration was 1 mg mL⁻¹ in 10 mM sodium phosphate pH 8.



Supplementary Figure 6. Comparison of the salt bridge interactions for all crystallography datasets. Top panels show the comparison between the sTIM1InoCys-SB datasets for the crystal form 1 (A) and crystal form 2 (B). Middle panels indicate the comparison between the DeNovoTIM6-SB datasets for the crystal form 1 (C) and crystal form 2 (D). Bottom panel shows the salt bridge cluster for the DeNovoTIM13-SB dataset (E).



Supplementary Figure 7. Molecular Replacement of DeNovoTIM13-SB structure in space group I4 (#79). **A)** TIM-barrel quarter generated by the molecular replacement processing of the data in this space group. **B)** TIM barrel reconstructed by using the corresponding symmetry operations. The four quarters are shown in different colors to highlight the proper formation of the barrel. For both panels, the unit cell is indicated in yellow.

Paper III

Sina Kordes, Merle Flecks, Leonie Lutz, Birte Höcker
Extension of a *de novo* TIM barrel with rationally designed antiparallel coiled coils
Manuscript

Extension of a *de novo* TIM barrel with rationally designed antiparallel coiled coils

Sina Kordes^{1,2}, Merle Flecks¹, Leonie Lutz¹, and Birte Höcker¹

¹Department of Biochemistry, University of Bayreuth, 95447 Bayreuth, Germany

²Current address: Proteros biostructures GmbH, 82515 Martinsried, Germany

Corresponding author:

Birte Hoecker¹

Email address: birte.hoecker@uni-bayreuth.de

ABSTRACT

The *de novo* design promises the production of tailor-made proteins. Nevertheless, so far the majority of designed proteins have idealized scaffolds without larger cavities which are necessary for incorporation of binding sites and enzymatic activities. One interesting target for enzyme design is the TIM-barrel fold, due to its ubiquity in nature and capability to host versatile functions. With the successful *de novo* design of a four-fold symmetric TIM barrel, sTIM11, an idealized, minimalistic scaffold was created. In this work, we attempted to extend sTIM11 by incorporating into its $\beta\alpha$ -loops *de novo* antiparallel coiled coils, which could provide large cavities to host diverse functionalities. A modular design approach using Rosetta was applied to determine suitable, antiparallel coiled-coil sequences, introduce them into the TIM barrel and perform optimization to frame single coiled-coil TIM barrels. Further diversification was performed by exploiting the symmetry of the scaffold to integrate two coiled coils into the TIM barrel. Biochemical characterization demonstrates the formation of additional α -helical secondary structure elements and conformational stability indicate the formation of additional stabilizing interactions. Therefore, the design of a TIM barrel with increased surface areas and large cavities by introduction of stable coiled coils is substantiated by experimental data. The successful designs can be further adapted to incorporate ligand binding and active sites for the ultimate functionalization of the *de novo* TIM barrels.

1 Introduction

The art of *de novo* protein design aims to expand the protein universe by creating new molecules with predefined properties. Recent progress in the design of specific topologies from scratch, including all- α [1, 2], all- β [3, 4] and $\alpha\beta$ proteins [5] has been made by recapitulating natural folds or creating complete new topologies [6, 7, 8]. One particular protein fold has challenged the field for many years: the $(\beta\alpha)_8$ - or TIM-barrel fold. Due to its ubiquity among enzymes it is the most common protein topology in nature and catalyzes a multitude of reactions [9]. Evaluation of natural enzymes revealed a proportion of about 10 % of the $(\beta\alpha)_8$ -topology. Therefore, this fold is able to catalyze five of the seven Enzyme Commission (EC) reaction classes [9]. A TIM barrel is organized in eight $\beta\alpha$ -subunits which form a central eight stranded, parallel β -barrel surrounded by eight α -helices [10, 11, 12]. A key feature of this fold is the spatial separation of stability and catalytic function. Regions at the N-terminal end of the β -barrel associated with stability are named the "stability face" [13, 14]. On the opposite site, at the "top" of the barrel, the "catalytic face" is located [15]. Supported by the extended $\beta\alpha$ -loops substrate binding is typically focused in a cavity formed at the

central surface of the β -barrel where catalytic residues are situated [16].

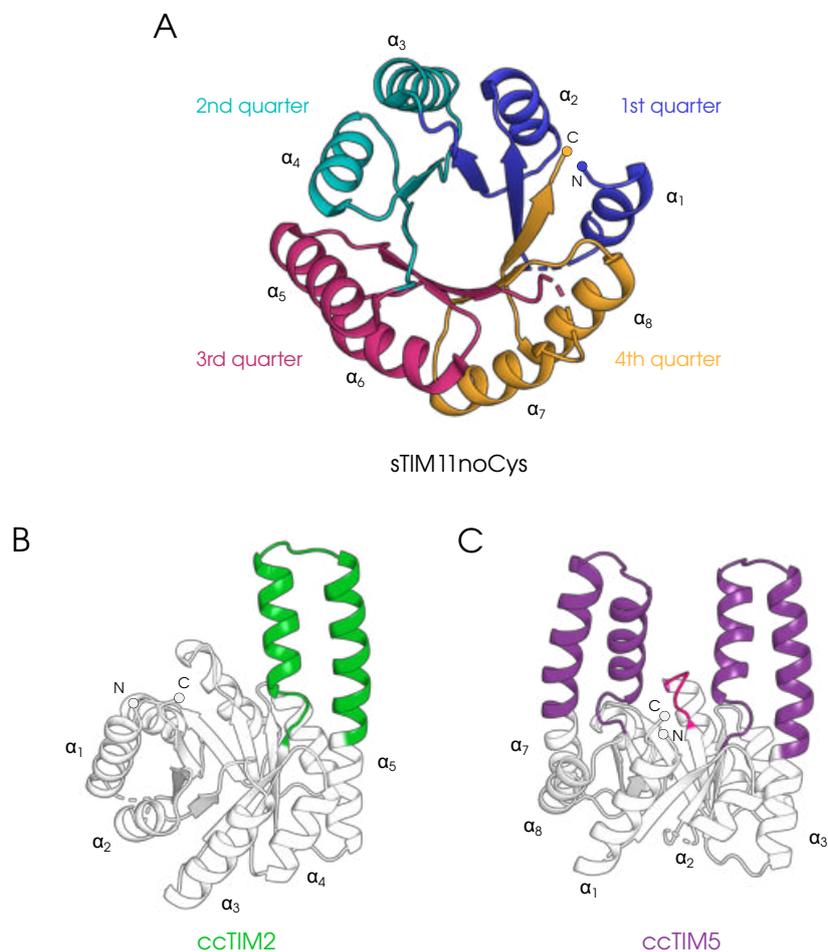


Figure 1. The *de novo* designed TIM barrels. **A** Crystal structure of the cysteine free idealized *de novo* TIM barrel is shown in cartoon representation (PDB ID: 6ZQZ [17]). The four symmetric quarters are shown in different colors and the α -helices are denoted according to their order in sequence. **B** The Rosetta model of ccTIM2 is shown in cartoon representation. The introduced coiled coil is highlighted in green. **C** The Rosetta model of ccTIM5 is shown in cartoon representation with introduced coiled coils highlighted in purple. In pink the β α -loop is highlighted which contains three additional residues in the 'GLE' variants. **B** and **C**. Numbering of α -helices according to sTIM11noCys notation.

Due to its unique structural organization and success in nature, the TIM-barrel fold is a highly interesting scaffold for enzyme design. For more than two centuries researchers endeavoured to identify the main structural determinants of this fold to create an idealized TIM barrel from scratch [18], until Huang et al. [5] succeeded by creating a four-fold symmetric TIM barrel using Rosetta. Their design, called sTIM11, was verified by a X-ray structure and demonstrated the formation of an idealized TIM barrel with minimal loops (Figure 1A). With the creation of this idealized

TIM barrel new opportunities for *de novo* enzyme design opened up. Nevertheless, comparison of sTIM11 with natural TIM barrels highlights, besides its exceptionally well-formed barrel structure, also improvement possibilities. In contrast to natural TIM barrels, sTIM11 comprises no cavities, pockets or extended loops on the "catalytic face" suitable for ligand binding or protein interactions.

Thus, for the creation of a functionalized *de novo* TIM barrel extended surfaces or hydrophobic pockets are required. Different strategies could be applied to achieve this, ranging from basic elongation of unstructured loops, over the introduction of small structural elements, up to the insertion of complete domains. In a recent approach Wiese et al. [19] moved the diversification of sTIM11 forward by the incorporation of a small α -helix into one of the $\beta\alpha$ loops. Even though their design did not match with the solved X-ray structure, due to a distortion induced by the extension, it highlights the rigidity of the *de novo* TIM barrel and its capability for further diversification. In another work, a previously designed ferredoxin fold was fused to a further stabilized TIM barrel variant to create a cavity between both domains which was functionalized by introducing a metal-binding site [20]. These two approaches highlight the ability of sTIM11 to serve as a base structure for further diversification by adding extensions in the $\beta\alpha$ -loops.

In this work we aimed to extend the family of *de novo* TIM barrels by adding larger secondary structure elements to the presumable catalytic face. We decided to use coiled coils for this purpose, due to their well understood design principles, their wide variety of possible applications and imagined the formation of increased hydrophobic cavities. *De novo* antiparallel coiled coils were incorporated into the $\beta\alpha$ -loops creating different ccTIMs. Taking advantage of the four-fold symmetry of sTIM11 a second coiled coil could be inserted analogously, creating a TIM barrel with enlarged surface area and possible hydrophobic pockets. This emphasizes also a possible modular extension strategy for the diversification of *de novo* TIM barrels. Experimental characterization of the ccTIM designs featured an increase of α -helical content in CD spectra and stability analysis indicate stable protein with increased number of stabilizing interactions highlighting the formation of additional structural elements. Even though no three-dimensional structure could be determined, these data specify the formation of additional helical structures and emphasize the successful diversification of *de novo* TIM barrels with coiled coils.

2 Materials and Methods

2.1 Design Protocol

All modelling and design steps were performed with the Rosetta molecular modelling suite using the scoring function ref2015 [21, 22]. Detailed command lines, blueprints and options can be found in the supplementary information. An overview over the design protocol is given in Figure 2.

2.1.1 Antiparallel Coiled Coil - *Ab Initio Relax*

Initially, an antiparallel coiled-coil model was created by structure prediction using Rosetta *Ab Initio Relax* [23, 24]. The sequence from Myszka and Chaiken [25] was adapted to fit our requirements: N- and C-termini were changed from Cys to Val, the sequence was tailored to two heptad repeats for each helix, and the loop length was reduced to three residues (Gly-Gly-Pro). Fragment files for *ab initio* prediction were created using the Robetta server (<http://old.robetta.org/fragmentsubmit.jsp>) [26, 23]. The predicted models were tested for convergence by calculating the RMSD (root mean square deviation) for all decoys against the lowest scoring model. Plotting the RMSD against the Rosetta total score produced a so called "folding funnel". The best scored model was used for subsequent design steps.

2.1.2 Initial Insertion - Rosetta Remodel

Insertion of the best *ab initio* coiled-coil model into sTIM11noCys, the cysteine free variant of sTIM11 (PDB ID: 6YQY [17]), was performed with Rosetta Remodel [27] using the -remodel:domainFusion:insert_segment_from_pdb option. Initially, different $\beta\alpha$ -loops were tested: $\beta_3\alpha_4$, $\beta_4\alpha_5$ and $\beta_5\alpha_6$ (Figure 1 A). In the supplementary only the blueprint for $\beta_4\alpha_5$ -loop is shown, equivalent blueprints were also used for the other two loops (Suppl. Methods). In detail, N- and C-terminally of the coiled-coil three or two flanking residues were flexible during the modelling step, respectively. Following all created models were relaxed using Rosetta relax [28]. Selection criteria were the Rosetta score and geometrical features of the coiled-coil. One model was selected for further adaptation and optimization.

2.1.3 Optimization and Adaptation - Remodel

The antiparallel coiled coil as well as the transition regions to the TIM barrel were further optimized using Rosetta Remodel. Following adjustments were done: (1) sequence redesign of surface residues with restriction to polar amino acids, (2) sequence redesign according to helical wheel (Figure 2), (3) adjustment of transition regions and (4) elongation of coiled-coil loop. Details can be gathered from Blueprint Continuous Helix (Suppl. Methods). Subsequent, all models were relaxed and ranked based on the score and manual investigation. At this point the first two single coiled-coil designs, ccTIM2 and ccTIM3, were selected. Further adaptation was performed for ccTIM1 by changing the $\alpha - \alpha$ -hairpin of the coiled coil. Additionally, helical capping based on the Schellmann motif was introduced for ccTIM4 [29]. In all cases final relaxation was done before selection of the single coiled-coil designs, ccTIM1 to 4, for experimental characterization.

2.2 Biochemical Materials

All reagents were analytical grade from Sigma-Aldrich or Carl Roth unless otherwise stated. All solutions were prepared with double-distilled water. All primers were ordered and synthesized from Eurofins. All enzymes and chemicals for cloning were obtained from New England BioLabs GmbH. Cloned constructs were verified by sequencing at myGATC or Eurofins. All other constructs were synthesized and cloned into pET21b(+) vector by BioCat.

2.3 Cloning Methods

2.3.1 ccTIM1 to 4

Initial constructs with one coiled coil, namely ccTIM1 to 4, were cloned using primers complementary to the insertion points of the TIM barrel with encoding coiled-coil sequence as overhangs. In a first step only one ccTIM primer in combination with forward T7 primer were used to amplify the first half of sTIM11noCys with the first part of the coiled-coil. In the same way, the second ccTIM primer was used in combination with the T7 reverse primer to create the second half of the TIM barrel. In a second step, these amplified sequences were used as megaprimers to amplify the pET21a vector with ccTIM as insert. For further amplification, both according ccTIM primers were used to multiply the pET21a::ccTIM vector. All polymerase chain reactions (PCR) were performed using Q5 Polymerase according to the manufacturer's standard protocol with 1 μ M primer mix. The PCR was performed with an elongation time of 20 sec and 30 cycles. Produced megaprimers were purified with the PCR clean-up kit from Macherey-Nagel. For megaprimer reaction annealing was performed at 72 °C with an increased elongation time of 3.5 min. The empty plasmid was digested with DpnI for 2 h at 37 °C. Amplified segments were ligated at 4 °C over night and following transformed into chemical competent *E. coli* Top10 cells (Novagen). Constructs were verified by sequencing.

2.3.2 ccTIM5-GLE

The ccTIM5-GLE construct was created via a two step approach. Initially, a half TIM-barrel with the coiled coil in the center was created via Gibson Assembly. Therefore, pET21a::ccTIM2 and the empty pET28c vector were used as template. Using the according forward and reverse primers the fragments were amplified with a standard PCR reaction using Q5 polymerase. Gibson assembly was performed with a 1:2 ratio of vector to insert by incubation at 37 °C for 5 h. The assembled product was transformed into *E. coli* Top10 cells and verified by sequencing. Using this verified construct pET28c::ccTIM2-half, a construct with two coiled-coils was produced by duplicating the ccTIM2 half barrel. Therefore, the primers fw_ccTIM5-GLE_Insert and rv_ccTIM5-GLE_Insert were used to amplify the insert and primers fw_ccTIM5-GLE_Vector and rv_ccTIM5-GLE_Vector to amplify the complete pET28c::ccTIM2-half construct. These segments were assembled via the sticky ends of the generated fragments as previously described and following transformed in Top10.

2.4 Overexpression and protein purification

The plasmids were transformed in *E. coli* BL21(DE3) (Novagen). A single clone from transformed cells was used to inoculate LB precultures supplemented with ampicillin (100 µg µL⁻¹) or kanamycin (50 µg µL⁻¹) and were grown at 37 °C and 180 rpm overnight. Terrific Broth (TB) cultures of 1 L were inoculated on an OD₆₀₀ of 0.08 and grown at 37 °C. Overexpression was induced at an OD₆₀₀ of 0.8-1 with 1 mM isopropyl-D-1-thiogalactopyranoside (IPTG). For ccTIM2, 3 and 4 expression was performed for 20 h at 20 °C, for all other constructs at 30 °C for 4.5 h. Cells were harvested by centrifugation (Beckmann Avanti JLA-8.1000, 15 min, 5000 g, 4 °C) and pellets resuspended in 5 mL per gram pellet buffer A: 50 mM sodium phosphate, 150 mM NaCl, 20 mM Imidazole, pH7.5 supplemented with 100 µL protease inhibitor (Mix-HP, Serva) per 10 mL lysate. Cells were lysed by sonication (Branson Ultrasonic) (output 4, duty cycle 40 %, 2 times 2 min) and then centrifuged (Beckmann Avanti JA-25.50, 1 h, 18000 rpm, 4 °C).

All purifications were performed at room temperature. The lysate was filtered with a 0.22 µm filter (Merck Millipore) and loaded on a HisTrap HP column (5 mL, Cytiva) coupled to an Äkta system (GE Healthcare Life Sciences). Removal of unbound protein was done with 20 column volumes (CV) of buffer A. Bound protein was eluted with a linear gradient over 20 CV from 50 to 300 mM Imidazole using buffer B (50 mM sodium phosphate, 150 mM NaCl, 500 mM Imidazole, pH7.5). The peak fractions were pooled, concentrated and loaded onto a HiLoad 16/600 Superdex 75 preparative grade column (GE Healthcare Life Sciences) connected to an Äkta System. Elution was performed with 1 CV buffer C (50 mM sodium phosphate, 150 mM NaCl, pH7.5) and the monomeric peak fractions were pooled and stored at room temperature for use in subsequent experiments. For ccTIM5, ccTIM5-SB, and ccTIM5-GLE-SB buffer A and B contained 300 mM NaCl for increased yields. Purification of ccTIM2, 3 and 4 was performed with buffers at pH8.0. For some subsequent experiments proteins were dialysed into buffer D (50 mM sodium phosphate pH7.5 or pH8.0).

2.5 Analytical Size Exclusion Chromatography/Multi Angle Light Scattering (SEC-MALS)

SEC-MALS measurements were performed using a Superdex 75 Increase 10/300 GL column (GE Healthcare Life Sciences) connected to an Äkta Pure System, coupled to a miniDAWN multi-angle light scattering detector and a Optilab refractometer (Wyatt Technology). All experiments were performed in buffer C with 0.02 % sodium azide at room temperature and a flow rate of 0.8 mL/min using protein concentrations of 1 and 5 mg mL⁻¹. Data collection and analysis was performed with ASTRA 7.3.2 software (Wyatt Technology). BSA standard sample at 2 mg mL⁻¹ was measured at

the beginning and end of each measurement day to check reproducibility.

2.6 Far-UV Circular Dichroism

Circular Dichroism (CD) spectra were collected in buffer C with a Jasco J-710 using a peltier device to control the temperature (PTC/348 WI). Far-UV spectra were recorded in the wavelength range 195-260 nm at 25 °C with a protein concentration of 0.2 mg mL⁻¹ with a 1 nm bandwidth in a 2 mm cuvette. Spectra of thermally unfolded protein were collected at 95 °C and 10 iterations were averaged. Normalization was performed by subtraction of buffer spectra and conversion to mean residue molar ellipticity using:

$$[\theta_{MRE}] = \frac{(M \cdot \theta)}{10 \cdot d \cdot c} \quad (1)$$

with $M = \frac{MW}{n-1}$, where M is the mean residue weight, MW the molecular weight in Da, n the number of residues in the protein, θ the collected ellipticity in mdeg, d the path length in mm and c the protein concentration in mg mL⁻¹. Deconvolution of the far-UV spectra was done with CDNN [30].

2.7 Intrinsic Fluorescence

Intrinsic fluorescence (IF) spectra were collected at 25 °C in buffer C with a protein concentration of 0.2 mg mL⁻¹ using a Jasco FP-6500 spectrofluorometer and a Peltier device to control the temperature (Julabo MB). Excitation of fluorescence was done at a wavelength of 295 nm and emission measured in the wavelength range 310-450 nm with a bandwidth of 1 nm.

2.8 Thermal unfolding followed by Circular Dichroism

Thermal unfolding was followed by CD at a protein concentration of 0.2 mg mL⁻¹ in buffer C in a 2 mm cuvette in the temperature range 20-95 °C at 222 nm with a heating rate of 1.5 °C/min. Spectra were normalized to the fraction of unfolded molecules (f_u) by:

$$f_u = \frac{y_{obs} - (y_N + m_N T)}{(y_U + m_U T) - (y_N + m_N T)} \quad (2)$$

with y_{obs} the observed CD signal at a given temperature, and $(y_N + m_N T)$ and $(y_U + m_U T)$ the linear fitting equation of the native and unfolded regions, respectively.

2.9 Thermal unfolding followed by Differential Scanning Calorimetry

In addition, Differential Scanning Calorimetry (DSC) was used to follow the thermal unfolding of the proteins using a VP-Capillary DSC (Malvern Panalytical). Samples were measured with a scanning speed of 1.5 °C/min and at protein concentrations in the range between 0.5 and 4 mg mL⁻¹ in buffer D after exhaustive dialysis and buffer degassing. Buffer-buffer scans were performed prior to all sample-buffer experiments and the last buffer-buffer scan was subtracted from each protein-buffer scan for thermodynamic analysis. Reversibility was determined by collecting a second endotherm after the first one was collected. The percentage of reversibility was determined by comparison the calorimetric ΔH for first scan and rescan ($\Delta H_{rescan} / \Delta H_{firstscan} \cdot 100$). DSC scans were fitted to a two-state reversible model:

$$C_p(T) = B_0 + B_1T + f(T)\Delta C_p + \frac{\Delta H(T)}{RT_m^2} \left(\frac{1 - f(T)}{1 - n + \frac{n}{f(T)}} \right) \quad (3)$$

where B_0 and B_1 are pre- and post-transition constants, n the number of subunits in the native protein sample (monomer for all ccTIMs) and $f(T)$ the protein fraction in the folded monomeric state, yielding the parameters ΔH , ΔC_p and T_m .

The van't Hoff enthalpy (ΔH_{vH}) was calculated with:

$$\Delta H_{vH} = \frac{4RT_m^2 C_{P,T_m}}{\Delta H} \quad (4)$$

with R the universal gas constant in $\text{cal mol}^{-1} \text{K}^{-1}$, T_m the melting temperature in K , C_{P,T_m} the heat capacity value at T_m and ΔH the total calorimetric enthalpy of the endotherm ([31]). Data analysis was done with Origin v.7.0 (OriginLab Corporation) with MicroCal software.

2.10 Chemical-induced unfolding followed by CD and IF

Chemical-induced unfolding was performed with Urea at a protein concentration of 0.2 mg mL^{-1} in buffer D. First, the equilibrium time was determined by incubation of samples at different urea concentration (0-6 M). CD and IF spectra were recorded at different incubation times and 2 days are sufficient for all analysed proteins to reach equilibrium. Following samples with increasing urea concentration were incubated for 2 days at 25°C . For all samples CD signal at 222 nm was followed for 2 min and IF spectra were recorded as aforementioned at 25°C . IF data were processed considering the intensity ratio at the wavelength of the maximum of the unfolded spectrum (I_{λ^u}) and at the wavelength of the maximum of the native spectrum (I_{λ^n}) at every urea concentration ($r_\lambda = \frac{I_{\lambda^u}}{I_{\lambda^n}}$). IF and CD data at every urea concentration were normalized to the fraction of unfolded protein using equation 5, where y_{obs} is the experimentally observed CD signal or the calculated ratio of IF data at a given concentration, and $(y_f + m_f[\text{urea}])$ and $(y_u + m_u[\text{urea}])$ are the linear fitting equations of the native and unfolded regions, respectively.

$$f_U = \frac{(y_f + m_f[\text{urea}]) - y_{obs}}{(y_f + m_f[\text{urea}]) - (y_u + m_u[\text{urea}])} \quad (5)$$

The unfolding free energy ΔG^{H_2O} was determined by fitting of the data to a two-state model ($\text{N} \rightleftharpoons \text{D}$) using the Santoro Bolen equation [32]:

$$f_U = \frac{(y_F + m_F[\text{urea}]) + (y_U + m_U[\text{urea}]) \cdot \exp\left(-\frac{\Delta G^{H_2O} - m[\text{urea}]}{RT}\right)}{1 + \exp\left(-\frac{\Delta G^{H_2O} - m[\text{urea}]}{RT}\right)} \quad (6)$$

where m is $\Delta G/[\text{urea}]$, T is the temperature of the experiment (298.15 K) and R the universal gas constant ($0.001987 \text{ kcal mol}^{-1} \text{K}^{-1}$). Data analysis and fitting was conducted with R [33, 34] and graphs were created with package ggplot2 [35].

3 Results

Detailed structural analysis of sTIM11 illustrates a flat surface with minimalistic $\beta\alpha$ -loops lacking structural features required for the design of functional sites. Hence, increasing its surface area and creating cavities are the first steps required for the design of a functionalized *de novo* TIM barrel. It was already shown in previous works that the $\beta\alpha$ -loops are suitable for the incorporation of additional structural elements [19, 20]. To create hydrophobic cavities and interaction areas we aimed to introduce larger structural elements. Coiled coils are a particularly suitable supersecondary structure element for this approach due to the well understood design principles and the high versatility of applications [36]. Therefore, we aimed to utilize these as building blocks in the extension of the *de novo* TIM barrel. As sTIM11 holds two cysteine residues which do not form the intended disulfide bond in the solved structure, we used the cysteine free variant, sTIM11noCys, which is reported in Romero-Romero et al. [17].

3.1 Design Strategy

The design of a coiled-coil TIM barrel, termed ccTIM, was accomplished using a multi-step design strategy (Figure 2). The design started with the creation of a backbone for the coiled coil (1), which was subsequently inserted into sTIM11noCys (2), in the following modified and optimized to create a rigid coiled-coil (3) and finally four designs were selected for experimental characterization (4) (Figure 2). Henceforth, a detailed description of these steps is given.

In a first assessment of our design goal, the introduction of coiled coils into the $\beta\alpha$ -loops of sTIM11noCys, we analysed the existing set of *de novo* designed coiled coils: although a large number of parallel arrangements of different lengths and oligomerization states have been designed, a shortage of *de novo* antiparallel coiled coils exists [37, 36]. Despite achievements in the design of antiparallel coiled coils as homodimers or higher oligomers [38, 39, 40, 41], we did not find examples of *de novo* designed two-stranded, intramolecular coiled coils in an antiparallel orientation in literature. As we aimed to introduce a coiled coil of this composition, no suitable model was available. Nevertheless, Myszka and Chaiken [25] proposed a sequence of a two-stranded, intramolecular, antiparallel coiled coil and showed experimental evidence for this arrangement. Using this sequence we performed *ab initio* predictions with Rosetta and successfully obtained a high quality model with the intended topology. With its 56 residues this coiled-coil sequence is relatively large in proportion to sTIM11noCys, therefore the repetitive sequence was reduced from 3.5 to two heptad repeats and the loop was halved to three amino acids (Figure 2-AbInitio Model). The tailored sequence was again applied to *ab initio* predictions and analysis revealed a folding funnel with low scoring models having low RMSDs, substantiating the sequences compatibility to the antiparallel coiled-coil geometry.

The best scoring *ab initio* model was selected and in the next step inserted into sTIM11noCys. We assumed that the $\beta\alpha$ -loops of the second and third quarter are most suitable for insertion of the coiled coil due to their lower B-factors in the crystal structures of sTIM11 and sTIM11noCys [5, 17]. Additionally, the spatial distance to the termini potentially reduces interference with the folding and closing of the TIM barrel. Therefore, we tested the four loops in the second and third quarter of the TIM barrel as possible insertion points for the coiled coils by using Rosetta Remodel. Premised on the overall rosetta score the $\beta_4\alpha_5$ -loop was selected for further design steps, which was also used by Wiese et al. [19]. For the design of the transition regions between TIM barrel and coiled coil a larger set of models was calculated, revealing highly diverse orientations of the coiled coil in respect to the TIM barrel, which might reflect possible flexibility (Figure 2-Insertion).

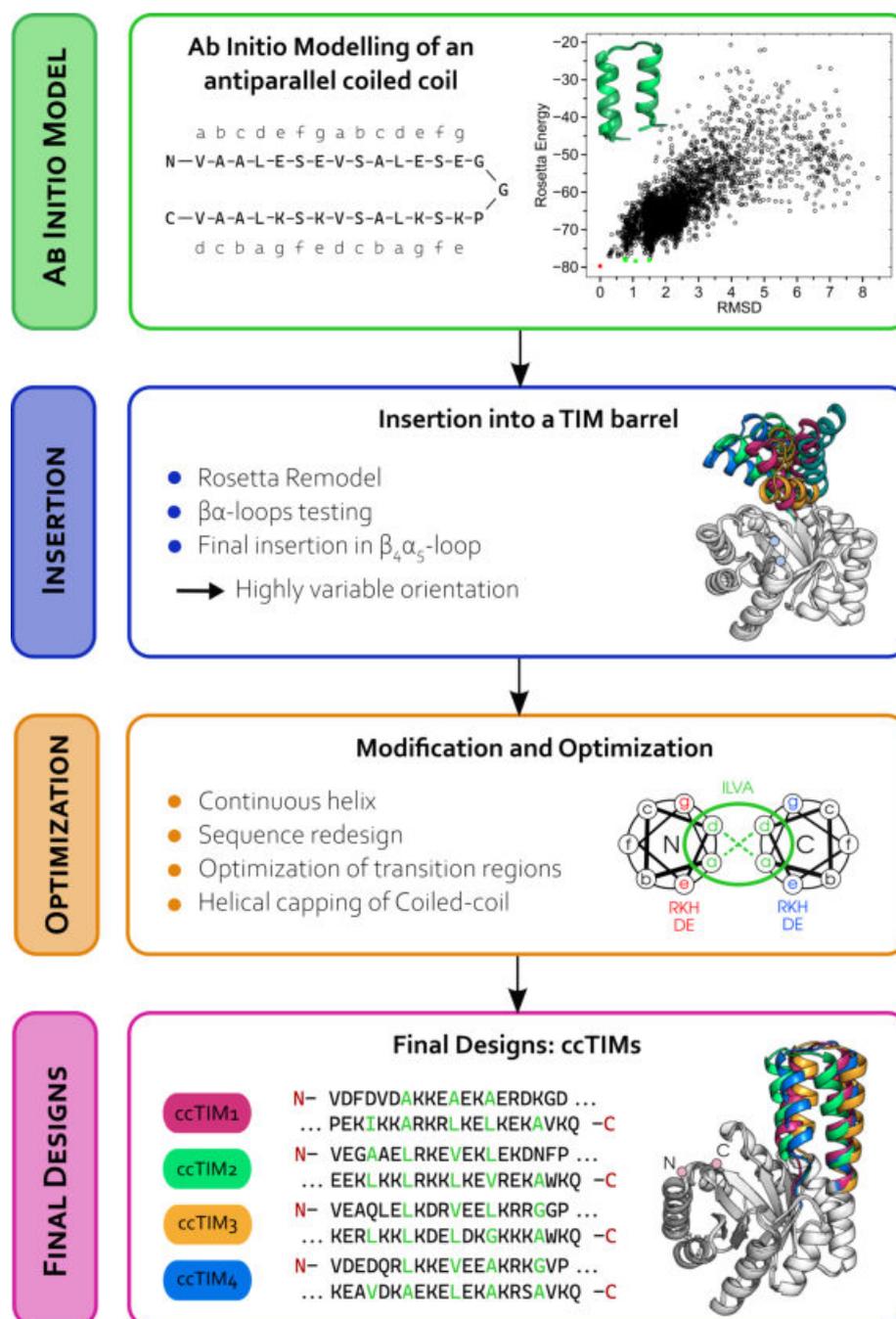


Figure 2. Design Strategy of ccTIMs. **Ab Initio Model.** Customised sequence of Myszka and Chaiken [25] for the antiparallel coiled coil used for the *ab initio* prediction is shown with notations for the heptad repeat positions. On the right folding funnel for models is shown with best scoring model highlighted in red and following three in green. Structure of best scored model is shown in cartoon representation. **Insertion.** In a second step designed coiled coils were inserted into the TIM barrel. Best five models are shown in cartoon representation illustrating high diversity of coiled-coil orientation. **Optimization** The orientation of the coiled coil was particularized by continuous helix and further optimization of the coiled coil sequences with amino acid restrictions based on the shown helical wheel diagram. **Final Designs** Coiled coil sequences as well as Rosetta models of the four selected designs are shown. Directionality is indicated by red letters for termini.

By elongating the second helix of the coiled coil to form a continuous helix with the α -helix of the TIM barrel we wanted to address this and rigidify its orientation. Therefore, the sequence of the transition region was shortened and restricted to an helical secondary structure. Only models with a continuous helix and low score were selected by manual analysis of the geometry for further optimization (Figure 2-Optimization). Based on previously determined design principles, the sequence of the coiled coil was redesigned [37, 42]: The characteristic heptad repeat of coiled coils, typically notated as $a-g$, can be depicted using a helical wheel diagram, which displays the orientation of the residues to each other in the assembly. Based on this representation, the coiled-coil interface is formed by the a and d residues, which were restricted in our sequence design to Ile, Leu, Val or Ala in order to form a stable hydrophobic core. The flanking residues, e and g , were restricted to charged residues in order to shield the hydrophobic interface from the solvent and to enable ionic interactions stabilizing the two helices in an antiparallel orientation.

Furthermore, the transition regions between the TIM barrel and the coiled coil were redesigned using only secondary structure restraints. The designs were filtered by the overall score and the final selection was based on geometrical analysis and two designs were selected: ccTIM2 and ccTIM3. In a final optimization round the $\alpha\alpha$ -hairpin of the coiled coil was addressed, and helical capping using the Schellman motif was installed resulting in the designs ccTIM1 and ccTIM4. These single coiled-coil designs, ccTIM1-4, were characterized as described in 3.2. The experimental data show only for ccTIM2 potential formation of a coiled coil.

A main design feature of sTIM11 was the four-fold symmetry, which we aimed to utilize by introducing multiple coiled coils in the different TIM barrel quarters in order to further increase the available area for functionalization and to create larger cavities. Therefore, we repurposed the coiled-coil sequence of ccTIM2 and grafted it into two opposing $\beta\alpha$ -loops to reduce the likelihood of steric clashes between them. In ccTIM2 the coiled coil was introduced in the $\beta_4\alpha_5$ -loop, however on the opposite side the protein termini are located. Thus for the new variants the coiled coils were introduced into the $\beta_2\alpha_3$ and $\beta_6\alpha_7$ (Figure 1B and C). This led to two constructs with double coiled-coil designs, namely ccTIM5 and ccTIM5-GLE. The later one comprises three additional residues (Gly-Leu-Glu) in the $\beta_4\alpha_5$ -loop which were unintentionally introduced during cloning. Finally, we improved the stability and crystallization behaviour of a set of *de novo* TIM barrels by the installation of a salt bridge cluster at the bottom face of the β -barrel by mutation of four residues [43]. In order to achieve a similar effect on the ccTIMs we introduced identical mutations into ccTIM5 and ccTIM5-GLE generating ccTIM5-SB and ccTIM5-GLE-SB, respectively. Sequences of all ccTIM variants are shown in Figure S1.

3.2 Characterization of single coiled-coil TIM barrels

The single coiled-coil designs ccTIM2, 3 and 4 were expressed soluble in *E. coli* and could be purified with sufficient yields. All three constructs exhibit monomeric, homogeneous peaks in an analytical size exclusion chromatography (SEC) and calculated weights correspond to a monomeric species (Figure S2A). Analysis of secondary structure content by circular dichroism spectroscopy revealed well folded protein only for ccTIM2 and ccTIM3. Both show typical $\beta\alpha$ -protein spectra similar to sTIM11noCys, though only for ccTIM2 an increased α -helical signal indicates the formation of the coiled coil (data for ccTIM2 shown in Figure 3B, comparison between the three proteins is shown in Figure S2B).

3.3 Characterization of ccTIM5 variants

All constructs with two coiled coils could be expressed soluble and purified to homogeneity. In comparison to sTIM11noCys and ccTIM2, reduced yields were observed for ccTIM5, ccTIM5-SB and ccTIM5-GLE-SB. Additionally, ccTIM5 and ccTIM5-SB were highly prone to aggregation after purification and precipitated rapidly. SEC-MALS measurements evince homogeneous, monomeric species with concentration independent behaviour for all constructs (Figure 3A) and calculated masses correlated with theoretical values of monomeric proteins (Table S1).

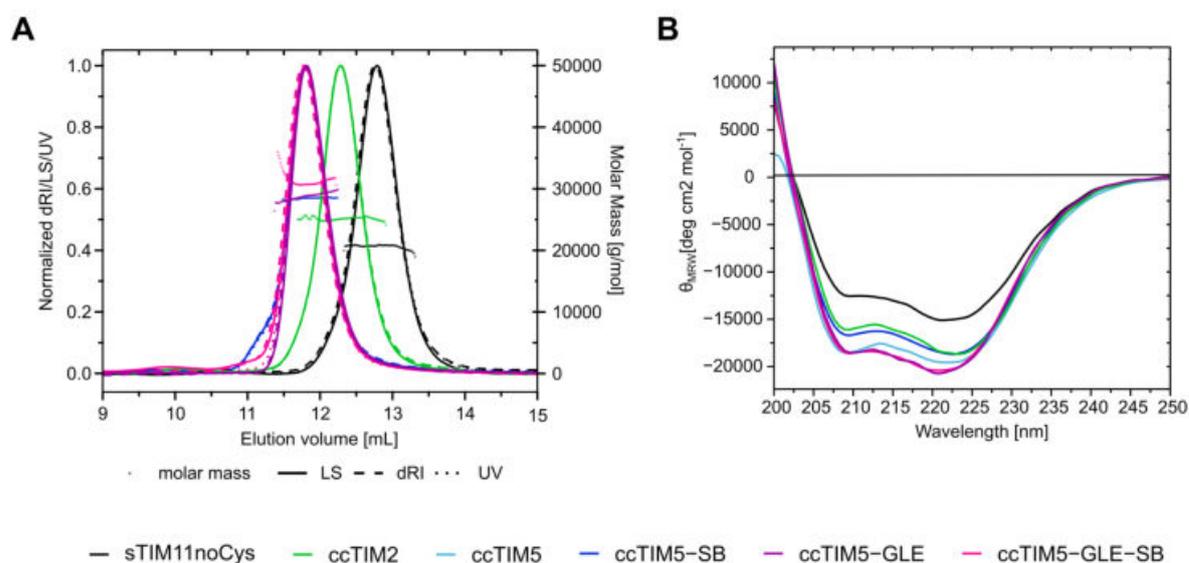


Figure 3. Biophysical analysis of ccTIMs. **A.** SEC-MALS measurements of sTIM11noCys and ccTIMs at 1 mg mL^{-1} . Plot showing the elution volume versus either normalized data for UV absorbance at 280 nm (UV, dotted lines), the differential refractive index (dRI, dashed line), and the multi-angle scattering (LS, solid lines), or the calculated molar mass (right axis, dot symbols). Values derived from this experiments are reported in the Supplementary Table 1. **B.** Comparison of native Far-UV CD spectra of sTIM11noCys and ccTIM variants. All spectra were collected at 0.2 mg mL^{-1} and were normalized to mean residue ellipticity.

CD spectra are compatible with $\alpha\beta$ -proteins with an overall comparable secondary structure composition to sTIM11noCys (Figure 3B). For all variants an increase in α -helical signal is observed, though for ccTIM5-SB it is merely in the magnitude of ccTIM2. In contrast, further gain of α -helical content is exhibited by ccTIM5, ccTIM5-GLE and ccTIM5-GLE-SB with latter two showing highly similar spectra. Additionally, CD spectra were deconvoluted and compared to the secondary structure content of the Rosetta models (Table S1). The β -strand content is similar for all ccTIMs and sTIM11noCys and corresponds to the predicted amounts. The highest α -helical content is predicted for ccTIM5-GLE and ccTIM5-GLE-SB with about 63 %, being even higher than calculated from the Rosetta models. The value for ccTIM5 and ccTIM5-SB is lower, but still indicates a small increase of α -helices compared to ccTIM2 and correlates well with the models. Interestingly, the content of random coils is similar for all ccTIMs and significantly lower than for sTIM11noCys. Therefore, the deconvolution data reinforces the observed increase in α -helical content in the CD spectra,

especially for ccTIM2, ccTIM5-GLE and ccTIM5-GLE-SB. Additionally, it highlights the positive effect of the 'GLE' residues on the formation of additional helical structures.

3.4 Coiled-coil insertions alter the thermostability

A first assessment of the thermal stability of ccTIMs was performed by monitoring the temperature induced unfolding by CD (Figure 4A). All analysed ccTIM variants show reversibility of unfolding as determined by comparison of CD spectra of native, unfolded and refolded protein (Figure S3). Only ccTIM5 and ccTIM5-GLE-SB exhibit reduced CD signal after refolding (Figure S3B and D) indicating reduced reversibility. Comparison with sTIM11noCys ($T_m = 73.6^\circ\text{C}$) revealed only a marginal destabilization of most ccTIM variants (Table 4). Notably, ccTIM5 and ccTIM5-SB exhibit reduced thermal stability by about 11 and 6 $^\circ\text{C}$, respectively. In contrast, for ccTIM2, ccTIM5-GLE and ccTIM5-GLE-SB only minor changes in melting temperature of up to 2.5 $^\circ\text{C}$ were determined. Due to its low stability and aggregation prone behaviour, ccTIM5 was not included in further analysis.

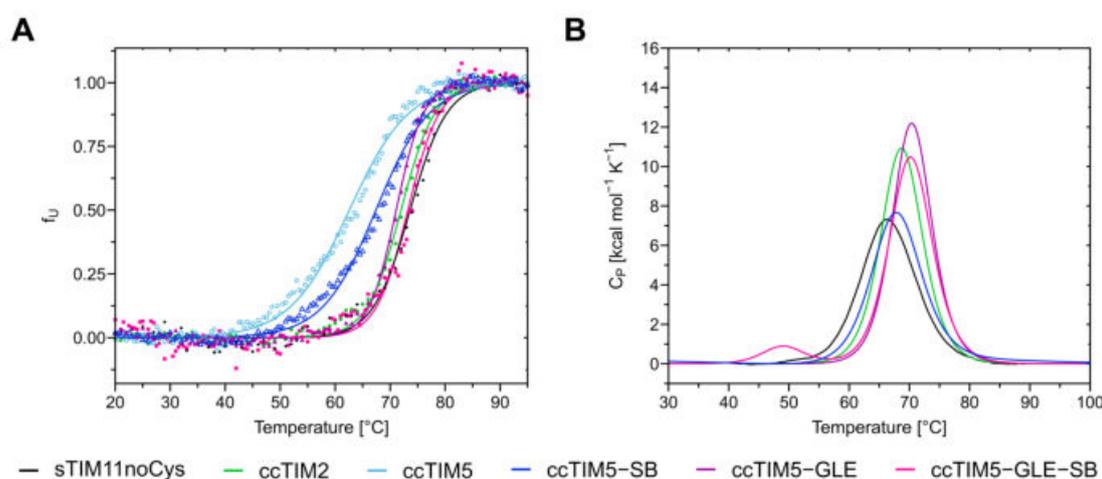


Figure 4. Thermodynamic Stability analysis of ccTIMs. **A** Thermal unfolding followed by CD. **B** Thermal unfolding followed by DSC. Endotherms were collected at 1.5 $^\circ\text{C}/\text{min}$ and a protein concentration of 1.0 mg mL^{-1} . Data from sTIM11noCys was reported in [17].

For a more accurate analysis of the thermal stability differential scanning calorimetry (DSC) was used. Interestingly, a contrary trend is observed in comparison to the CD data. All ccTIM variants exhibit a higher melting temperature in comparison to sTIM11noCys and the ccTIM ranking by DSC melting temperature differs from rank order by CD derived T_m values (Figure 4 and table 1). One possible reason is the different salt concentration used for CD and DSC measurements, but it is intriguing that a higher thermal stability was determined in conditions without salt. With a T_m^{DSC} of 67.30 $^\circ\text{C}$ ccTIM5-SB is the least stable variant followed by ccTIM2 and ccTIM5-GLE. In contrast, ccTIM5-GLE-SB exhibits a three state-unfolding with a first transition at about 50 $^\circ\text{C}$ and complete unfolding at 70.25 $^\circ\text{C}$, which is the highest T_m^{DSC} for all ccTIMs. Reversible thermal unfolding was verified for ccTIM2, ccTIM5-GLE and ccTIM5-SB as the recovered area is above 85 % and data was fitted to a reversible two-state model (Figure S4A, C, E). However, ccTIM5-GLE-SB shows reduced reversibility of unfolding with a recovery rate of 81.2 % and therefore was not fitted to a

Table 1. Stability properties of ccTIMs

de novo TIM barrel	Thermal unfolding						Chemical unfolding (by CD and IF)		
	by CD		by DSC				$\Delta G_{25^\circ\text{C}}$ (kcal mol ⁻¹)	<i>m</i> (kcal mol ⁻¹ M ⁻¹)	D[1/2] (M)
	<i>T_m</i> (°C)	<i>T_m</i> (°C)	ΔH (kcal mol ⁻¹)	ΔCP (kcal mol ⁻¹ K ⁻¹)	Calorimetric criterion $\Delta H_{\text{VH}}/\Delta H$				
sTIM11noCys*	73.6 ± 0.2	65.6 ± 0.1	82 ± 1	2.36 ± 0.08	0.99	3.2 ± 0.2	2.03 ± 0.1	1.9	
ccTIM2	72.0 ± 0.1	67.81 ± 0.03	97.4 ± 0.2	3.76 ± 0.07	1.02	5.3 ± 0.2	2.23 ± 0.1	2.4	
ccTIM5-GLE	71.0 ± 0.1	69.87 ± 0.04	107 ± 0.4	3.81 ± 0.13	0.98	5.3 ± 0.2	2.57 ± 0.1	2.1	
ccTIM5-SB	67.4 ± 0.1	67.30 ± 0.06	86.5 ± 0.3	1.42 ± 0.10	0.94	4.5 ± 0.1	1.89 ± 0.1	2.4	
ccTIM5-GLE-SB	73.2 ± 0.1	70.26 ± 0.002 (49.10 ± 0.02)	99.5 ± 0.05 (8.3 ± 0.05)	n.d. ^b		CD: 4.1 ± 0.2 IF: 6.3 ± 0.2	CD: 1.76 ± 0.1 IF: 2.94 ± 0.1	CD: 2.33 IF: 2.13	

± indicates the standard error from global fitting

^a n.d. - not determined due to inaccuracy of fit and therefore low confidence of ΔC_p

^b n.d. - not determined due to irreversibility in thermal unfolding and a non two-state mechanism

reversible model (Figure S4G). For all four ccTIM variants a concentration independence of thermal unfolding was verified (Figure S4B, D, F, H).

Analysis of the unfolding enthalpy ΔH , which is related to the number of stabilizing interactions in a protein and therefore is proportional to the molecular weight, can give insights in the fold organization and indicates the formation of well-folded proteins. For sTIM11noCys, ccTIM2, ccTIM5-GLE and ccTIM5-GLE-SB ΔH increases with the molecular weight indicating the formation of additional bonds in these ccTIM variants (Table 1). In contrast, ccTIM5-SB exhibits only a minor ΔH increase compared to sTIM11noCys thus, being inconsistent with its higher molecular weight and implying a lower number of stabilizing interactions compared to the other ccTIM variants. This is also in agreement with the overall reduced CD signal (Figure 3 B) for ccTIM5-SB in comparison to the other double coiled-coil designs, indicating an imperfect folding of the coiled-coil insertions.

Data of ccTIM2 and ccTIM5-GLE could be well fitted to a two-state model and further thermodynamic parameters were determined. The heat capacity change ΔC_p of both proteins increases significantly in comparison to sTIM11noCys (Table 1). The heat capacity of a protein is related to its hydration state, with a positive contribution for hydration of non-polar groups and a negative contribution for polar groups. The overall heat capacity change upon unfolding is therefore positive for globular proteins [44]. The increased ΔC_p for ccTIM2 and ccTIM5-GLE indicates the formation of additional structures with hydrophobic cores which are hydrated upon unfolding. This is in agreement with ΔH values and indicates the formation of globular proteins with additional structural regions, as expected for successful designs. Fitting of ccTIM5-SB to a two-state model was insufficient due to poor baselines, therefore the determined parameters are not reliable and further analysis of the global thermodynamic stability was not performed.

The ratio of the van't Hoff enthalpy and the measured enthalpy is determined as the Calorimetric criterion, which can be used to judge reversibility, unfolding process and influences of protein aggregation of DSC measurements. Since sTIM11noCys, ccTIM2, and ccTIM5-GLE exhibit values around 1 a two-state unfolding mechanism was confirmed and protein aggregation can be eliminated. For ccTIM5-SB the calorimetric criterion was also determined, even though the data fits only poorly to a two-state model. Nevertheless, as ΔC_p is highly influenced by baseline fits and consequently influences the calculation of the van't Hoff enthalpy, this has to be considered with caution. The thermodynamic data further substantiates the formation of additional stable structures in ccTIM2 and ccTIM5-GLE. The insertion of the coiled coil reshapes the thermodynamic stability of the TIM barrel with a T_m raise by around 2 °C and an increase of ΔH by 10 to 15 kcal mol⁻¹ per additional coiled coil. Also the additivity of these effects with the insertion of the second coiled-coil assists the assumption that the coiled-coils are well formed in ccTIM2 and ccTIM5-GLE.

3.5 Conformational stability of ccTIMs is increased

Conformational stability of ccTIMs was investigated by chemical unfolding with Urea and followed by CD and IF. All proteins show a reversible, cooperative unfolding with coinciding CD and IF transitions, except for ccTIM5-GLE-SB which exhibits deviating unfolding transitions (Figure 5D and S5). The combined IF and CD data of ccTIM2, ccTIM5-SB and ccTIM5-GLE as well as separated data of ccTIM5-GLE-SB fitted well to a two-state model. For all analysed ccTIM variants the conformational stability was increased in comparison to sTIM11noCys (Table 1). The unfolding of ccTIM2 and ccTIM5-SB show highly similar transitions with identical $D_{[1/2]}$ but slight variances in the cooperativity with a higher value for ccTIM2 and consequently differing $\Delta G_{25\text{ }^\circ\text{C}}$. For ccTIM5-GLE the transition curve is more similar to sTIM11noCys, but with increased values for all parameters. Interestingly, the $\Delta G_{25\text{ }^\circ\text{C}}$ of ccTIM2 and ccTIM5-GLE is identical with

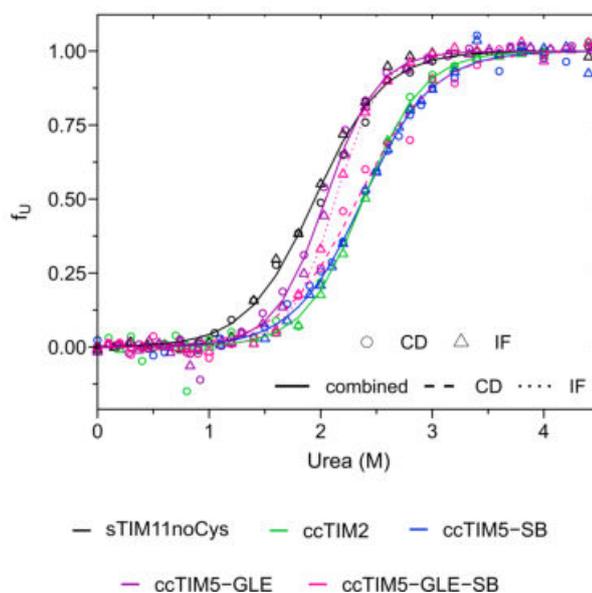


Figure 5. Chemical unfolding of ccTIMs. Chemical unfolding with urea at 25 °C, circles representing CD data and triangles fluorescence data. For ccTIM5-GLE-SB fitting of CD and IF data is shown separately: CD - dotted lines, IF - dashed line. For all other data continuous line shows the combined fit. Data from sTIM11noCys was reported in [17].

5.3 kcal mol⁻¹, but m and $D_{[1/2]}$ vary, showing a lower cooperativity for ccTIM2.

Comparison of ccTIM5-GLE-SB with sTIM11noCys and other ccTIM variants is more difficult. Traditionally, non-coinciding transitions in chemical unfolding are evidence for a non-two state unfolding. Nevertheless, as there is no evidence for an intermediate state in either of the two transitions, we assume a two-state transition which is also assisted by the fact that both techniques monitor different structural changes. CD monitors the decline of secondary structure and IF follows the unfolding of the tertiary structure by monitoring tryptophan fluorescence. As no Trp residues are present in the coiled-coil insertion, only the TIM barrel unfolding is monitored by IF. In contrast, both TIM barrel and the coiled coil insertions contribute to the transition monitored by CD. This is also reinforced by separately comparing the IF and CD transition of sTIM11noCys with the other ccTIMs. The IF transition is more similar to the unfolding of sTIM11noCys and the determined parameters show a further stabilization of the TIM barrel, most likely induced by the incorporation of the salt bridge cluster (Figure S5D). In contrast, in the CD transition also the coiled-coil unfolding can be monitored and comparison shows high accordance of the CD transition of ccTIM5-GLE-SB with ccTIM2 and ccTIM5-SB. Another reason for the non-coinciding transitions could be a not fully reached equilibrium for this protein after a incubation of 2 days. Comparison of both transition curves shows high agreement up to 2 M Urea. At higher denaturant concentrations the TIM barrel, monitored by IF, unfolds faster with higher cooperativity. Contrary, the unfolding of the secondary structure is less cooperative. A more detailed kinetics analysis of the unfolding of ccTIM5-GLE-SB would be required to refine its unfolding pathway.

In summary, the conformational stability of all ccTIM variants is increased in comparison to sTIM11noCys. Especially, ccTIM2 and ccTIM5-GLE show a high stability with a two-state transition and highly similar stability parameters, but also ccTIM5-GLE-SB exhibits promising stability features. This data is also in agreement with previously shown data of the biochemical

characterization and thermal unfolding. Altogether the experimental data of these three designed proteins, ccTIM2, ccTIM5-GLE and ccTIM5-GLE-SB, indicate the formation of additional, stably formed secondary structure elements and promote the assumption of a successful design approach.

4 Discussion

In this work we attempted to further diversify the family of *de novo* TIM barrels by introducing antiparallel coiled coils into the $\beta\alpha$ -loops of sTIM11noCys [5]. This original protein has an idealized topology with minimalistic loops and therefore misses any extended surface areas and cavities for installation of binding and active sites. Due to the well understood design principles and their highly diverse applicability, we chose antiparallel coiled coils as insertion module. By separating the coiled-coil generation, the insertion into the TIM barrel, and the optimization of the final designs the complexity of each step was reduced and adjustments could be easily performed to fit our requirements. Starting with the creation of an antiparallel coiled coil as stable building block we were able to test different loops for optimal insertion points. Additionally, we identified variability of the coiled-coil orientation as a possible concern and aligned our design approach accordingly to ensure a steady transition from the coiled coil into the α -helix of sTIM11noCys. Additional sequence redesign of the coiled coil and the transition regions to the TIM barrel were performed to find low energy sequences and four designs were selected for experimental characterization.

After initial characterization, one coiled coil proved to be most suitable and was therefore utilized for the creation of further extended variants by introduction of two identical coiled coils into opposing $\beta\alpha$ -loops. In the attempt to further stabilize the TIM barrels, the previously reported stabilizing salt bridge was introduced [43]. The final set of ccTIM variants consisted of the best single coiled-coil variant ccTIM2, the double coiled-coil variants ccTIM5 and ccTIM5-GLE, and the variants with additional salt bridge mutations at the N-terminal face of the β -barrel, ccTIM5-SB and ccTIM5-GLE-SB. Biochemical characterization highlighted the leading variants to be ccTIM2, ccTIM5-GLE and ccTIM5-GLE-SB, due to their CD spectra being compatible with the designs, increased thermal stability and a non-aggregation prone behaviour. These three variants exhibited an increase of the α -helical secondary structure content in conformity with the inserted coiled coil. Further on, detailed analysis of thermodynamic and conformational stability of these variants emphasizes the formation of additional, well-folded secondary structure elements. Contrary to frequent observations, that insertions have a destabilizing effect on protein folds ([45, 46], upon insertion of the coiled coils no thermal or conformational destabilization of the TIM barrel was induced. Evaluation of DSC data revealed an increase of ΔH for ccTIM2, ccTIM5-GLE and ccTIM5-GLE-SB which is associated with the higher number of intramolecular interactions. Additionally, ccTIM2 and ccTIM5-GLE could be fitted to a two-state model allowing the determination of ΔC_p , which is increase for both proteins and refers to an increased hydration upon unfolding. Consequently, DSC analysis specifies the stable formation of secondary structure elements without disrupting the TIM barrel scaffold. Further analysis of the conformational stability by chemical unfolding with urea revealed an increased ΔG for ccTIM2 and ccTIM5-GLE. Both proteins show a well defined two-state unfolding process with high cooperativity and increased stability values, which suggests the formation of additional stabilizing interactions with a well formed hydrophobic core.

ccTIM5-GLE-SB is a special case in both, DSC and chemical unfolding, as it exhibits two transitions in both experiments. As CD follows secondary structure and IF tertiary structure changes during chemical unfolding, we speculate that for this protein the coiled coil unfolds independently of the TIM barrel. This is in agreement with DSC analysis which also outline two transitions upon

thermal unfolding. Nevertheless, detailed analysis of this protein would be required to determine its exact unfolding process. Another interesting observation is the importance of the additional 'GLE' residues in the $\beta_4\alpha_5$ -loop for all ccTIM5 variants. Direct comparison of the ccTIM5 variants with their respective 'GLE' variant shows a higher amount of α helical content in CD and a significantly higher T_m for ccTIM5-GLE and ccTIM5-GLE-SB. Additionally, ccTIM5 and ccTIM5-SB were prone to aggregation and only ccTIM5-SB was stable enough to be characterized in detail. However, the thermal and conformational stability parameters of ccTIM-SB were less promising and did not indicate formation of the coiled coils. Following, the 'GLE' insertion seems to have a positive effect on the ccTIMs either by stabilizing the TIM barrel or by direct interaction with the coiled coils.

In summary, our approach to create larger cavities in sTIM11 by incorporating antiparallel coiled coils, delivered a set of promising designs. For three proteins, ccTIM2, ccTIM5-GLE and ccTIM5-GLE-SB, the experimental data emphasize the formation of stable coiled coils. Although the designs could not be verified by a 3D structure yet, the biochemical and thermodynamic measurements highlight the formation of stable additional structures with well formed hydrophobic cores, additional intramolecular interactions and a high α -helical content.

In comparison to recent works, which introduced a full domain or a short α -helix into the *de novo* TIM barrel [20, 19], our approach of introducing coiled coils has some advantages. First, the created interface and cavities are larger compared to a small α -helix and can be adjusted easier than complete domains due to a less complex structure of the coiled coils. Additionally, as shown by introducing one and two coiled coils into the TIM barrel the designed proteins can easily be extended further. On the one hand it should be possible to introduce further coiled coils into the remaining symmetry related $\beta\alpha$ -loops which would result in a four-fold coiled-coil arrangement with a large pocket on top of the TIM barrel. On the other hand it might also be interesting to shuffle different approaches by combining the coiled-coils with the designed α -helix. Furthermore, a large set of diverse TIM barrels with varying stabilities was created already in previous works [18, 43] and introducing the designed coiled coil in these proteins might result in proteins with tailored stabilities and cavities suitable for the design of functional *de novo* TIM barrels. On the way to create tailor-made enzymes a main challenge is the determination of stable scaffolds suitable for the incorporation of a specific active site geometry [47, 48]. The *de novo* TIM barrels exhibit excellent features of a design platform and with the diverse set of extensions created in previous and in the current work make a diverse set of topologies available, which can be combined to meet different geometries.

5 Acknowledgments

We thank Sabrina Wischt for her competent technical support and Sergio Romero-Romero and Stefan Klingl for their comments on the manuscript. We kindly thank all members of the Höcker Lab for their constructive suggestions to improve the research.

References

- [1] Lindsey Doyle, Jazmine Hallinan, Jill Bolduc, Fabio Parmeggiani, David Baker, Barry L. Stoddard, and Philip Bradley. Rational design of α -helical tandem repeat proteins with closed architectures. *Nature*, 528(7583):585–588, 12 2015. ISSN 14764687. doi: 10.1038/nature16191.
- [2] Lynne Regan and William F Degrado. Characterization of a Helical Protein Designed from First Principles. *Science*, 241(4868):976–979, 1988.

- [3] Enrique Marcos, Tamuka M. Chidyausiku, Andrew C. McShan, Thomas Evangelidis, Santrupti Nerli, Lauren Carter, Lucas G. Nivón, Audrey Davis, Gustav Oberdorfer, Konstantinos Tripsianes, Nikolaos G. Sgourakis, and David Baker. De novo design of a non-local β -sheet protein with high stability and accuracy. *Nature Structural and Molecular Biology*, 25(11):1028–1034, 11 2018. ISSN 15459985. doi: 10.1038/s41594-018-0141-6.
- [4] Jiayi Dou, Anastassia A. Vorobieva, William Sheffler, Lindsey A. Doyle, Hahnbeom Park, Matthew J. Bick, Binchen Mao, Glenna W. Foight, Min Yen Lee, Lauren A. Gagnon, Lauren Carter, Banumathi Sankaran, Sergey Ovchinnikov, Enrique Marcos, Po Ssu Huang, Joshua C. Vaughan, Barry L. Stoddard, and David Baker. De novo design of a fluorescence-activating β -barrel. *Nature*, 561(7724):485–491, 9 2018. ISSN 14764687. doi: 10.1038/s41586-018-0509-0.
- [5] Po-Ssu Ssu Huang, Kaspar Feldmeier, Fabio Parmeggiani, D. Alejandro Fernandez Velasco Fernandez, Birte Hocker, David Baker, D Alejandro Fernandez Velasco, Birte Höcker, and David Baker. De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nature Chemical Biology*, 12(1):29–34, 2015. ISSN 1552-4450. doi: 10.1038/nchembio.1966.
- [6] Xingjie Pan, Michael C. Thompson, Yang Zhang, Lin Liu, James S. Fraser, Mark J. S. Kelly, and Tanja Kortemme. Expanding the space of protein geometries by computational design of de novo fold families. *Science*, 369(6507):1132–1136, 8 2020. ISSN 0036-8075. doi: 10.1126/science.abc0881.
- [7] Che Yang, Fabian Sesterhenn, Jaume Bonet, Eva A. van Aalen, Leo Scheller, Luciano A. Abriata, Johannes T. Cramer, Xiaolin Wen, Stéphane Rosset, Sandrine Georgeon, Theodore Jardetzky, Thomas Krey, Martin Fussenegger, Maarten Merckx, and Bruno E. Correia. Bottom-up de novo design of functional proteins with complex structural features. *Nature Chemical Biology*, 17(4):492–500, 1 2021. ISSN 15524469. doi: 10.1038/s41589-020-00699-x.
- [8] Brian Kuhlman, Gautam Dantas, Gregory C Ireton, Gabriele Varani, Barry L Stoddard, and David Baker. Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science*, 302(5649):1364–1368, 2003.
- [9] Reinhard Sterner and Birte Höcker. Catalytic versatility, stability, and evolution of the ($\beta\alpha$)₈-barrel enzyme fold. *Chemical Reviews*, 105(11):4038–4055, 2005. ISSN 00092665. doi: 10.1021/cr030191z.
- [10] D. W. Banner, A. C. Bloomer, G. A. Petsko, D. C. Phillips, C. I. Pogson, I. A. Wilson, P. H. Corran, A. J. Furth, J. D. Milman, R. E. Offord, J. D. Priddle, and S. G. Waley. Structure of chicken muscle triose phosphate isomerase determined crystallographically at 2.5Å resolution: Using amino acid sequence data. *Nature*, 255(5510):609–614, 1975. ISSN 00280836. doi: 10.1038/255609a0.
- [11] Dominique Maes, Johan P Zeelen, Narmada Thanki, Nicola Beaucamp, Marco Alvarez, Thi Minh Hoa Dao, Jan Backmann, Joseph A Martial, Lode Wyns, Rainer Jaenicke, and Rik K Wierenga. The crystal structure of triosephosphate isomerase (TIM) from *Thermotoga maritima*: A comparative thermostability structural analysis of ten different TIM structures. *Proteins: Structure, Function and Genetics*, 37(3):441–453, 1999. ISSN 08873585. doi: 10.1002/(SICI)1097-0134(19991115)37:3<441::AID-PROT111>3.0.CO;2-7.
- [12] R. K. Wierenga. The TIM-barrel fold: A versatile framework for efficient enzymes. *FEBS Letters*, 492(3):193–198, 3 2001. ISSN 00145793. doi: 10.1016/S0014-5793(01)02236-0.
- [13] Roman Urfer and Kasper Kirschner. The importance of surface loops for stabilizing an eightfold $\beta\alpha$ barrel protein. *Protein Science*, 1(1):31–45, 1992. ISSN 1469896X. doi: 10.1002/pro.5560010105.

- [14] M. S. Vijayabaskar and Saraswathi Vishveshwara. Insights into the Fold Organization of TIM Barrel from Interaction Energy Based Structure Networks. *PLoS Computational Biology*, 8(5): e1002505, 5 2012. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1002505.
- [15] Nozomi Nagano, Christine A Orenge, and Janet M Thornton. One Fold with Many Functions : The Evolutionary Relationships between TIM Barrel Families Based on their Sequences , Structures and Functions. *Journal of Molecular Biology*, 2836(02):741–765, 2002. doi: 10.1016/S0022-2836(02)00649-6.
- [16] Ralf Thoma, Michael Hennig, Reinhard Sterner, and Kasper Kirschner. Structure and function of mutationally generated monomers of dimeric phosphoribosylanthranilate isomerase from *Thermotoga maritima*. *Structure*, 8(3):265–276, 3 2000. ISSN 09692126. doi: 10.1016/S0969-2126(00)00106-4.
- [17] Sergio Romero-Romero, Miguel Costas, Daniel Adriano Silva Manzano, Sina Kordes, Erendira Rojas-Ortega, Cinthya Tapia, Yasel Guerra, Sooruban Shanmugaratnam, Adela Rodríguez-Romero, David Baker, Birte Höcker, and D. Alejandro Fernández-Velasco. The Stability Landscape of de novo TIM Barrels Explored by a Modular Design Approach. *Journal of Molecular Biology*, 433(18):167153, 9 2021. ISSN 10898638. doi: 10.1016/j.jmb.2021.167153.
- [18] Sergio Romero-Romero, Sina Kordes, Florian Michel, and Birte Höcker. Evolution, folding, and design of TIM barrels and related proteins. *Current Opinion in Structural Biology*, 68: 94–104, 6 2021. ISSN 1879033X. doi: 10.1016/j.sbi.2020.12.007.
- [19] Jonas Gregor Wiese, Sooruban Shanmugaratnam, and Birte Höcker. Extension of a de novo TIM barrel with a rationally designed secondary structure element. *Protein Science*, 30(5): 982–989, 5 2021. ISSN 1469896X. doi: 10.1002/pro.4064.
- [20] Shane J. Caldwell, Ian C. Haydon, Nikolett Piperidou, Po Ssu Huang, Matthew J. Bick, H. Sebastian Sjöström, Donald Hilvert, David Baker, and Cathleen Zeymer. Tight and specific lanthanide binding in a de novo TIM barrel with a large internal cavity designed by symmetric domain fusion. *Proceedings of the National Academy of Sciences of the United States of America*, 117(48):30362–30369, 12 2020. ISSN 10916490. doi: 10.1073/pnas.2008535117.
- [21] Rebecca F. Alford, Andrew Leaver-Fay, Jeliakzko R. Jeliakzov, Matthew J. O’Meara, Frank P. DiMaio, Hahnbeom Park, Maxim V. Shapovalov, P. Douglas Renfrew, Vikram K. Mulligan, Kalli Kappel, Jason W. Labonte, Michael S. Pacella, Richard Bonneau, Philip Bradley, Roland L. Dunbrack, Rhiju Das, David Baker, Brian Kuhlman, Tanja Kortemme, Jeffrey J. Gray, Matthew J. O’Meara, Frank P. DiMaio, Hahnbeom Park, Maxim V. Shapovalov, P. Douglas Renfrew, Vikram K. Mulligan, Kalli Kappel, Jason W. Labonte, Michael S. Pacella, Richard Bonneau, Philip Bradley, Roland L. Dunbrack, Rhiju Das, David Baker, Brian Kuhlman, Tanja Kortemme, and Jeffrey J. Gray. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *Journal of Chemical Theory and Computation*, 13(6):3031–3048, 6 2017. ISSN 15499626. doi: 10.1021/acs.jctc.7b00125.
- [22] Andrew Leaver-Fay, Michael Tyka, Steven M. Lewis, Oliver F. Lange, James Thompson, Ron Jacak, Kristian W. Kaufman, P. Douglas Renfrew, Colin A. Smith, Will Sheffler, Ian W. Davis, Seth Cooper, Adrien Treuille, Daniel J. Mandell, Florian Richter, Yih-En En Andrew Ban, Sarel J. Fleishman, Jacob E. Corn, David E. Kim, Sergey Lyskov, Monica Berrondo, Stuart Mentzer, Zoran Popović, James J. Havranek, John Karanicolas, Rhiju Das, Jens Meiler, Tanja Kortemme, Jeffrey J. Gray, Brian Kuhlman, David Baker, and Philip Bradley. Rosetta3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules. *Methods in Enzymology*, 487(C):545–574, 1 2011. ISSN 0076-6879. doi: 10.1016/B978-0-12-381270-4.00019-6.

- [23] Srivatsan Raman, Robert Vernon, James Thompson, Michael Tyka, Ruslan Sadreyev, Jimin Pei, David Kim, Elizabeth Kellogg, Frank Dimaio, Oliver Lange, Lisa Kinch, Will Sheffler, Bong Hyun Kim, Rhiju Das, Nick V. Grishin, and David Baker. Structure prediction for CASP8 with all-atom refinement using Rosetta. *Proteins: Structure, Function and Bioinformatics*, 77 (SUPPL. 9):89–99, 2009. ISSN 08873585. doi: 10.1002/prot.22540.
- [24] Kim T. Simons, Charles Kooperberg, Enoch Huang, and David Baker. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *Journal of Molecular Biology*, 268(1):209–225, 4 1997. ISSN 00222836. doi: 10.1006/jmbi.1997.0959.
- [25] David G. Myszka and Irwin M Chaiken. Design and Characterization of an Intramolecular Antiparallel Coiled-Coil Peptide. *Biochemistry*, 33(9):2363–2372, 1994. ISSN 0006-2960. doi: 10.1021/bi00175a003.
- [26] Yifan Song, Frank Dimaio, Ray Yu Rwei Wang, David Kim, Chris Miles, Tj Brunette, James Thompson, and David Baker. High-resolution comparative modeling with RosettaCM. *Structure*, 21(10):1735–1742, 10 2013. ISSN 09692126. doi: 10.1016/j.str.2013.08.005.
- [27] Po Ssu Huang, Yih En Andrew Ban, Florian Richter, Ingemar Andre, Robert Vernon, William R. Schief, and David Baker. RosettaRemodel: A generalized framework for flexible backbone protein design. *PLoS ONE*, 6(8), 2011. ISSN 19326203. doi: 10.1371/journal.pone.0024109.
- [28] Lucas Gregorio Nivón, Rocco Moretti, and David Baker. A Pareto-Optimal Refinement Method for Protein Design Scaffolds. *PLOS ONE*, 8(4):e59004, 4 2013. ISSN 1932-6203. doi: 10.1371/JOURNAL.PONE.0059004.
- [29] Rajeev Aurora and George D Rose. Helix capping. *Protein science : a publication of the Protein Society*, 7(1):21–38, 1998. doi: 10.1002/pro.5560070103.
- [30] Gerald Böhm, Rudolf Muhr, and Rainer Jaenicke. Quantitative analysis of protein far UV circular dichroism spectra by neural networks. *Protein engineering*, 5(3):191–195, 1992. ISSN 0269-2139. doi: 10.1093/PROTEIN/5.3.191.
- [31] Ernesto Freire. Statistical thermodynamic analysis of differential scanning calorimetry data: Structural deconvolution of heat capacity function of proteins. *Methods in Enzymology*, 240 (C):502–530, 1 1994. ISSN 15577988. doi: 10.1016/S0076-6879(94)40062-8.
- [32] Marcelo M. Santoro and D. W. Bolen. Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl .alpha.-chymotrypsin using different denaturants. *Biochemistry*, 27(21):8063–8068, 10 2002. doi: 10.1021/BI00421A014.
- [33] RStudio Team. RStudio: Integrated Development Environment for R, 2021.
- [34] R Core Team. R: A Language and Environment for Statistical Computing, 2021.
- [35] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4.
- [36] Derek N. Woolfson. Coiled-coil design: Updated and upgraded. *Sub-Cellular Biochemistry*, 82:35–61, 1 2017. ISSN 03060225. doi: 10.1007/978-3-319-49674-0_2.
- [37] M. G. Oakley and J. J. Hollenbeck. The design of antiparallel coiled coils. *Current Opinion in Structural Biology*, 11(4):450–457, 2001. ISSN 0959440X. doi: 10.1016/S0959-440X(00)00232-3.
- [38] Daniel G. Gurnon, Jennifer A. Whitaker, and Martha G. Oakley. Design and characterization of a homodimeric antiparallel coiled coil. *Journal of the American Chemical Society*, 125(25): 7518–7519, 2003. ISSN 00027863. doi: 10.1021/ja0357590.
- [39] Kevin Pagel, Karsten Seeger, Bettina Seiwert, Alessandra Villa, Alan E. Mark, Alessandra VillaCurrent address: J. W. Goethe, Alan E. Mark, Stefan Berger, and Beate Kokschi. Ad-

- vanced approaches for the characterization of a de novo designed antiparallel coiled coil peptide. *Organic & Biomolecular Chemistry*, 3(7):1189, 3 2005. ISSN 1477-0520. doi: 10.1039/b418167k.
- [40] Vladimir N. Malashkevich, Chelsea D. Higgins, Steven C. Almo, and Jonathan R. Lai. A switch from parallel to antiparallel strand orientation in a coiled-coil X-ray structure via two core hydrophobic mutations. *Biopolymers*, 104(3):178–185, 5 2015. ISSN 10970282. doi: 10.1002/bip.22631.
- [41] Guto G. Rhys, Christopher W. Wood, Joseph L. Beesley, Nathan R. Zaccai, Antony J. Burton, R. Leo Brady, Andrew R. Thomson, and Derek N. Woolfson. Navigating the Structural Landscape of de Novo α -Helical Bundles. *Journal of the American Chemical Society*, 141(22):8787–8797, 6 2019. ISSN 15205126. doi: 10.1021/jacs.8b13354.
- [42] Erik B Hadley, Oliver D Testa, Derek N Woolfson, and Samuel H Gellman. Preferred side-chain constellations at antiparallel coiled-coil interfaces. *Proceedings of the National Academy of Sciences of the United States of America*, 105(2):530–5, 2008. doi: 10.1073/pnas.0709068105.
- [43] Sina Kordes, Sergio Romero-Romero, Leonie Lutz, and Birte Höcker. A newly introduced salt bridge cluster improves structural and biophysical properties of de novo TIM barrels. *Protein Science*, 12 2021. ISSN 0961-8368. doi: 10.1002/pro.4249.
- [44] Peter L. Privalov and George I. Makhatadze. Contribution of hydration and non-covalent interactions to the heat capacity effect on protein unfolding. *Journal of Molecular Biology*, 224(3):715–723, 4 1992. ISSN 00222836. doi: 10.1016/0022-2836(92)90555-X.
- [45] Anne Juliane Geitner and Franz Xaver Schmid. Combination of the human prolyl isomerase FKBP12 with unrelated chaperone domains leads to chimeric folding enzymes with high activity. *Journal of Molecular Biology*, 420(4-5):335–349, 7 2012. ISSN 10898638. doi: 10.1016/j.jmb.2012.04.018.
- [46] Gabriel Zoldák, Linn Carstensen, Christian Scholz, and Franz X. Schmid. Consequences of Domain Insertion on the Stability and Folding Mechanism of a Protein. *Journal of Molecular Biology*, 386(4):1138–1152, 3 2009. ISSN 00222836. doi: 10.1016/j.jmb.2008.12.052.
- [47] William M. Dawson, Freddie J.O. Martin, Guto G. Rhys, Kathryn L. Shelley, R. Leo Brady, and Derek N. Woolfson. Coiled coils 9-to-5: Rational: de novo design of α -helical barrels with tunable oligomeric states. *Chemical Science*, 12(20):6923–6928, 5 2021. ISSN 20416539. doi: 10.1039/d1sc00460c.
- [48] Katie J. Grayson and J. L. Ross Anderson. Designed for life: Biocompatible de novo designed proteins and components. *Journal of the Royal Society Interface*, 15(145), 2018. ISSN 17425662. doi: 10.1098/rsif.2018.0472.

Supplementary Material

Extension of a *de novo* TIM barrel with rationally designed antiparallel coiled coils

Sina Kordes^{1,2}, Merle Flecks¹, Leonie Lutz¹, Birte Höcker^{1*}

¹ Department of Biochemistry, University of Bayreuth, 95447 Bayreuth, Germany.

² Current address: Proteros biostructures GmbH, 85215 Martinsried, Germany

* Corresponding author: Birte Höcker. Department of Biochemistry, University of Bayreuth, 95447 Bayreuth, Germany, e-mail address: birte.hoecker@uni-bayreuth.de

This file includes:

- Supplementary Methods
- Supplementary Table S1
- Supplementary Figures S1-S4

Supplementary Methods

Specific options for Rosetta modelling is given in the following with exact blueprints in the next section.

For all calculations Rosetta version 2017.34 was used with the scoring function ref2015.

1. Options used for *ab initio* calculations

```
-database Rosetta/main/database
-abinitio:relax
-in:file:fasta coiledcoil.fasta
-in:file:frag3 3fragm
-in:file:frag9 9fragm
-nstruct 5000
```

2. Options used for initial insertion into *sTIM11noCys*

```
-database Rosetta/main/database
-remodel:blueprint blueprint_initial_insertion.txt
-remodel:domainFusion:insert_segment_from_pdb CoiledCoil.pdb
-remodel:num_trajectory 5000
-remodel::quick_and_dirty
-find_neighbors
```

3. Options used for optimization of continuous coiled coil

```
-database Rosetta/main/database
-remodel:blueprint blueprint_ContinuousHelix.txt
-remodel:num_trajectory 5000
-find_neighbors
-no_optH false
-ex1
-ex2
```

4. Insertion of helical capping for *ccTIM1* and *ccTIM4*

```
-database Rosetta/main/database
-remodel:blueprint blueprint_optimization
-remodel:num_trajectory 5000
-find_neighbors
-no_optH false
-ex1
-ex2
```

Blueprints

blueprint_initial_insertion.txt

1 D .	53 Q .	0 x I
2 K .	54 L .	0 x I
3 D .	55 R .	0 x I
4 E .	56 R .	0 x I
5 A .	57 E .	0 x I
6 W .	58 G .	0 x I
7 K .	59 A .	0 x I
8 C .	60 T .	0 x I
9 V .	61 Q .	0 x I
10 E .	62 I .	0 x I
11 Q .	63 A .	0 x I
12 L .	64 Y .	0 x I
13 R .	65 R .	0 x I
14 R .	66 S .	0 x I ALLAA
15 E .	67 D .	0 x I ALLAA
16 G .	68 D .	90 K H PIKAA K
17 T .	69 W .	91 D H PIKAA D
18 Q .	70 R .	92 E .
19 I .	71 D .	93 A .
20 A .	72 L .	94 W .
21 Y .	73 K .	95 K .
22 R .	74 E .	96 Q .
23 S .	75 A .	97 V .
24 D .	76 W .	98 E .
25 D .	77 K .	99 Q .
26 W .	78 K .	100 L .
27 R .	79 G .	101 R .
28 D .	80 A .	102 R .
29 L .	81 D .	103 E .
30 K .	82 I .	104 G .
31 E .	83 L .	105 A .
32 A .	84 I .	106 T .
33 W .	85 V .	107 Q .
34 A .	86 D S PIKAA D	108 I .
35 D .	87 A S PIKAA A	109 A .
36 I .	88 T D ALLAA	110 Y .
37 L .	0 x I ALLAA	111 R .
38 I .	0 x I ALLAA	112 S .
39 V .	0 x I	113 D .
40 D .	0 x I	114 D .
41 A .	0 x I	115 W .
42 T .	0 x I	116 R .
43 D .	0 x I	117 D .
44 K .	0 x I	118 L .
45 D .	0 x I	119 K .
46 E .	0 x I	120 E .
47 A .	0 x I	121 A .
48 W .	0 x I	122 W .
49 K .	0 x I	123 K .
50 Q .	0 x I	124 K .
51 V .	0 x I	125 G .
52 E .	0 x I	126 A .

127 D .	145 Q .	163 D .
128 I .	146 L .	164 L .
129 L .	147 R .	165 K .
130 I .	148 R .	166 E .
131 V .	149 E .	167 A .
132 D .	150 G .	168 W .
133 A .	151 A .	169 K .
134 T .	152 T .	170 K .
135 D .	153 Q .	171 G .
136 K .	154 I .	172 A .
137 D .	155 A .	173 D .
138 E .	156 Y .	174 I .
139 A .	157 R .	175 L .
140 W .	158 S .	176 I .
141 K .	159 D .	177 C .
142 Q .	160 D .	178 D .
143 V .	161 W .	179 A .
144 E .	162 R .	180 T .

Blueprint_ContinuousHelix.txt

1 D .	37 L .	73 K .
2 K .	38 I .	74 E .
3 D .	39 V .	75 A .
4 E .	40 D .	76 W .
5 A .	41 A .	77 K .
6 W .	42 T .	78 K .
7 K .	43 D .	79 G .
8 C .	44 K .	80 A .
9 V .	45 D .	81 D .
10 E .	46 E .	82 I .
11 Q .	47 A .	83 L .
12 L .	48 W .	84 I .
13 R .	49 K .	85 V .
14 R .	50 Q .	86 D E PIKAA D
15 E .	51 V .	87 A E ALLAA
16 G .	52 E .	88 S L ALLAA
17 T .	53 Q .	89 T H ALLAA
18 Q .	54 L .	91 A H POLAR PIKAA KRHDE
19 I .	55 R .	92 L H PIKAA ILVA
20 A .	56 R .	93 E H POLAR PIKAA KRHDE
21 Y .	57 E .	94 S H POLAR PIKAA KRHDE
22 R .	58 G .	95 E H POLAR PIKAA KRHDE
23 S .	59 A .	96 V H PIKAA ILVA
24 D .	60 T .	97 S H POLAR PIKAA KRHDE
25 D .	61 Q .	98 A H POLAR PIKAA KRHDE
26 W .	62 I .	99 L H PIKAA ILVA
27 R .	63 A .	100 E H POLAR PIKAA KRHDE
28 D .	64 Y .	101 S H POLAR PIKAA KRHDE
29 L .	65 R .	102 E D POLAR PIKAA KRHDE
30 K .	66 S .	103 G L PIKAA GFLE
31 E .	67 D .	104 G L PIKAA GDR
32 A .	68 D .	0 x L ALLAA
33 W .	69 W .	105 P L PIKAA P
34 A .	70 R .	106 K H PIKAA DNK
35 D .	71 D .	107 S H POLAR PIKAA KRHDE
36 I .	72 L .	108 K H POLAR PIKAA KRHDE

109 L H PIKAA ILVA	143 D .	177 R .
110 A H POLAR PIKAA KRHDE	144 D .	178 R .
111 S H POLAR PIKAA KRHDE	145 W .	179 E .
112 V H PIKAA ILVA	146 R .	180 G .
113 K H POLAR PIKAA KRHDE	147 D .	181 A .
114 S H POLAR PIKAA KRHDE	148 L .	182 T .
115 K H POLAR PIKAA KRHDE	149 K .	183 Q .
116 L H PIKAA ILVA	150 E .	184 I .
117 A H POLAR PIKAA KRHDE	151 A .	185 A .
118 S H POLAR PIKAA KRHDE	152 W .	186 Y .
119 A H PIKAA ILVA	153 K .	187 R .
120 K H POLAR PIKAA KRHDE	154 K .	188 S .
121 D H POLAR PIKAA KRHDE	155 G .	189 D .
122 E H ALLAA	156 A .	190 D .
123 A H ALLAA	157 D .	191 W .
124 W H	158 I .	192 R .
125 K .	159 L .	193 D .
126 Q .	160 I .	194 L .
127 V .	161 V .	195 K .
128 E .	162 D .	196 E .
129 Q .	163 A .	197 A .
130 L .	164 T .	198 W .
131 R .	165 D .	199 K .
132 R .	166 K .	200 K .
133 E .	167 D .	201 G .
134 G .	168 E .	202 A .
135 A .	169 A .	203 D .
136 T .	170 W .	204 I .
137 Q .	171 K .	205 L .
138 I .	172 Q .	206 I .
139 A .	173 V .	207 C .
140 Y .	174 E .	208 D .
141 R .	175 Q .	209 A .
142 S .	176 L .	210 T .

Blueprint ccTIM1

1 D .	20 A .	39 V .
2 K .	21 Y .	40 D .
3 D .	22 R .	41 A .
4 E .	23 S .	42 T .
5 A .	24 D .	43 D .
6 W .	25 D .	44 K .
7 K .	26 W .	45 D .
8 C .	27 R .	46 E .
9 V .	28 D .	47 A .
10 E .	29 L .	48 W .
11 Q .	30 K .	49 K .
12 L .	31 E .	50 Q .
13 R .	32 A .	51 V .
14 R .	33 W .	52 E .
15 E .	34 A .	53 Q .
16 G .	35 D .	54 L .
17 T .	36 I .	55 R .
18 Q .	37 L .	56 R .
19 I .	38 I .	57 E .

58 G .	110 K .	161 V .
59 A .	111 K .	162 D .
60 T .	112 A .	163 A .
61 Q .	113 R .	164 T .
62 I .	114 K .	165 D .
63 A .	115 R .	166 K .
64 Y .	116 L .	167 D .
65 R .	117 K .	168 E .
66 S .	118 E .	169 A .
67 D .	119 L .	170 W .
68 D .	120 K .	171 K .
69 W .	121 E .	172 Q .
70 R .	122 K .	173 V .
71 D .	123 A .	174 E .
72 L .	124 V .	175 Q .
73 K .	125 K .	176 L .
74 E .	126 Q .	177 R .
75 A .	127 V .	178 R .
76 W .	128 E .	179 E .
77 K .	129 Q .	180 G .
78 K .	130 L .	181 A .
79 G .	131 R .	182 T .
80 A .	132 R .	183 Q .
81 D .	133 E .	184 I .
82 I .	134 G .	185 A .
83 L .	135 A .	186 Y .
84 I .	136 T .	187 R .
85 V .	137 Q .	188 S .
86 D .	138 I .	189 D .
87 F .	139 A .	190 D .
88 D .	140 Y .	191 W .
89 V .	141 R .	192 R .
90 D .	142 S .	193 D .
91 A .	143 D .	194 L .
92 K .	144 D .	195 K .
93 K .	145 W .	196 E .
94 E .	146 R .	197 A .
95 A .	147 D .	198 W .
96 E .	148 L .	199 K .
97 K .	149 K .	200 K .
98 L H PIKAA AL	150 E .	201 G .
99 E H PIKAA EK	151 A .	202 A .
100 K H PIKAA KR	152 W .	203 D .
101 D H PIKAA DKR	153 K .	204 I .
103 R H PIKAA HKL	154 K .	205 L .
104 K L PIKAA G	155 G .	206 I .
105 P L PIKAA D	156 A .	207 C .
106 K H PIKAA PFW	157 D .	208 D .
107 E H PIKAA E	158 I .	209 A .
108 E H PIKAA EK	159 L .	210 T .
109 I H PIKAA LIA	160 I .	

blueprint_ccTIM4

1 D .		113 E .
2 K .	57 E .	114 K .
3 D .	58 G .	115 E .
4 E .	59 A .	116 L .
5 A .	60 T .	117 E .
6 W .	61 Q .	118 K .
7 K .	62 I .	119 A .
8 C .	63 A .	120 K .
9 V .	64 Y .	121 R .
10 E .	65 R .	122 S .
11 Q .	66 S .	123 A .
12 L .	67 D .	124 V .
13 R .	68 D .	125 K .
14 R .	69 W .	126 Q .
15 E .	70 R .	127 V .
16 G .	71 D .	128 E .
17 T .	72 L .	129 Q .
18 Q .	73 K .	130 L .
19 I .	74 E .	131 R .
20 A .	75 A .	132 R .
21 Y .	76 W .	133 E .
22 R .	77 K .	134 G .
23 S .	78 K .	135 A .
24 D .	79 G .	136 T .
25 D .	80 A .	137 Q .
26 W .	81 D .	138 I .
27 R .	82 I .	139 A .
28 D .	83 L .	140 Y .
29 L .	84 I .	141 R .
30 K .	85 V .	142 S .
31 E .	86 D .	143 D .
32 A .	87 E .	144 D .
33 W .	88 D .	145 W .
34 A .	89 Q .	146 R .
35 D .	90 K .	147 D .
36 I .	91 L .	148 L .
37 L .	92 K .	149 K .
38 I .	93 K .	150 E .
39 V .	94 K .	151 A .
40 D .	95 V .	152 W .
41 A .	96 E .	153 K .
42 T .	97 E H PIKAA E	154 K .
43 D .	98 I H PIKAA IAL	155 G .
44 K .	99 K H PIKAA KRHDE	156 A .
45 D .	100 K H PIKAA KRHDE	157 D .
46 E .	101 D H PIKAA HL	158 I .
47 A .	103 R L PIKAA G	159 L .
48 W .	104 A L PIKAA RDN	160 I .
49 K .	105 P H PIKAA P	161 V .
50 Q .	106 K H PIKAA KRHDE	162 D .
51 V .	107 E H	163 A .
52 E .	108 E H PIKAA A	164 T .
53 Q .	109 I H PIKAA ILVA	165 D .
54 L .	110 D H PIKAA D	166 K .
55 R .	111 K .	167 D .
56 R .	112 A .	168 E .

169 A .	183 Q .	197 A .
170 W .	184 I .	198 W .
171 K .	185 A .	199 K .
172 Q .	186 Y .	200 K .
173 V .	187 R .	201 G .
174 E .	188 S .	202 A .
175 Q .	189 D .	203 D .
176 L .	190 D .	204 I .
177 R .	191 W .	205 L .
178 R .	192 R .	206 I .
179 E .	193 D .	207 C .
180 G .	194 L .	208 D .
181 A .	195 K .	209 A .
182 T .	196 E .	210 T .

Blueprint_ccTIM4_FinalAdjustment

1 D .	40 D .	79 G .
2 K .	41 A .	80 A .
3 D .	42 T .	81 D .
4 E .	43 D .	82 I .
5 A .	44 K .	83 L .
6 W .	45 D .	84 I .
7 K .	46 E .	85 V .
8 C .	47 A .	86 D .
9 V .	48 W .	87 E .
10 E .	49 K .	88 D .
11 Q .	50 Q .	89 Q H PIKAA Q
12 L .	51 V .	90 K H PIKAA R
13 R .	52 E .	91 L H PIKAA L
14 R .	53 Q .	92 K H PIKAA K
15 E .	54 L .	93 K H PIKAA K
16 G .	55 R .	94 K H PIKAA E
17 T .	56 R .	95 V H PIKAA V
18 Q .	57 E .	96 E H PIKAA E
19 I .	58 G .	97 E H PIKAA E
20 A .	59 A .	98 A H PIKAA A
21 Y .	60 T .	99 K H PIKAA K
22 R .	61 Q .	100 R H PIKAA R
23 S .	62 I .	101 L H PIKAA HFMEK
24 D .	63 A .	102 G L PIKAA G
25 D .	64 Y .	103 R L PIKAA VI
26 W .	65 R .	104 P L PIKAA P
27 R .	66 S .	105 K H PIKAA K
28 D .	67 D .	106 V H PIKAA ED
29 L .	68 D .	107 A H PIKAA A
30 K .	69 W .	108 V H PIKAA V
31 E .	70 R .	109 D H PIKAA D
32 A .	71 D .	110 K H PIKAA K
33 W .	72 L .	111 A .
34 A .	73 K .	112 E .
35 D .	74 E .	113 K .
36 I .	75 A .	114 E .
37 L .	76 W .	115 L .
38 I .	77 K .	116 E .
39 V .	78 K .	117 K .

118 A .	149 E .	180 A .
119 K .	150 A .	181 T .
120 R .	151 W .	182 Q .
121 S .	152 K .	183 I .
122 A .	153 K .	184 A .
123 V .	154 G .	185 Y .
124 K .	155 A .	186 R .
125 Q .	156 D .	187 S .
126 V .	157 I .	188 D .
127 E .	158 L .	189 D .
128 Q .	159 I .	190 W .
129 L .	160 V .	191 R .
130 R .	161 D .	192 D .
131 R .	162 A .	193 L .
132 E .	163 T .	194 K .
133 G .	164 D .	195 E .
134 A .	165 K .	196 A .
135 T .	166 D .	197 W .
136 Q .	167 E .	198 K .
137 I .	168 A .	199 K .
138 A .	169 W .	200 G .
139 Y .	170 K .	201 A .
140 R .	171 Q .	202 D .
141 S .	172 V .	203 I .
142 D .	173 E .	204 L .
143 D .	174 Q .	205 I .
144 W .	175 L .	206 C .
145 R .	176 R .	207 D .
146 D .	177 R .	208 A .
147 L .	178 E .	209 T .
148 K .	179 G .	

Supplementary Table

Supplementary Table 1: Biochemical and biophysical properties of ccTIM variants in comparison to sTIM11noCys

<i>de novo</i> TIM barrel	Expression/Solubility properties		Hydrodynamic properties			Spectroscopic properties		
						Circular dichroism		
	Soluble overexpression ^a	Aggregation-prone ^b	Theoretical MW (kDa)	Experimental MW (kDa)	Exp./Theor. Ratio	Predicted secondary structure content (%) ^c		
α -helix						β -strand	Random coil	
sTIM11noCys	+++	+	22.88	20.7 ± 0.1	0.9	47.6 (49.4)	19.2 (21.9)	33.2 (28.7)
ccTIM2	+++	+	26.45	25.2 ± 0.1	0.95	54.7 (53.2)	22.3 (25.5)	22 (21.3)
ccTIM5-GLE	+++	+	30.60	29.0 ± 0.2	0.95	63.2 (56.5)	19.1 (21.8)	17.7 (21.8)
ccTIM5-SB	++	++	30.08	28.5 ± 0.6	0.95	56.4 (57.3)	21.6 (25.2)	20.6 (17.5)
ccTIM5-GLE-SB	+	+	30.38	31.2 ± 0.1	1.03	62.6 (57.3)	19.3 (23.8)	17.5 (19.0)
ccTIM5	+	+++	30.02	n.d.	n.d.	59.3	20.5	18.5

^a Symbols represent the yield of pure protein: +++ > 10 mg L⁻¹ culture, ++ 2-10 mg L⁻¹ culture, < 2 mg L⁻¹ culture

^b Symbols represent aggregation behaviour after purification: +++ aggregation visible within hours after purification, ++ aggregation visible within days after purification, + no aggregation observed > 1 month after purification

^c numbers within brackets are the calculated from the three-dimensional structure (sTIM11noCys) or the rosetta models (ccTIMs)

Supplementary Figures

List of supplementary figures

- **Fig S1:** Sequence alignment of ccTIMs and sTIM11noCys
- **Fig S2:** Comparison analytical SEC and CD spectra of ccTIM2, 3 and 4
- **Fig S3:** Far-UV CD spectra of ccTIM variants
- **Fig S4:** Thermal unfolding of ccTIM variants followed by DSC.
- **Fig S5:** Chemical unfolding of ccTIMs followed by CD (normalized data)

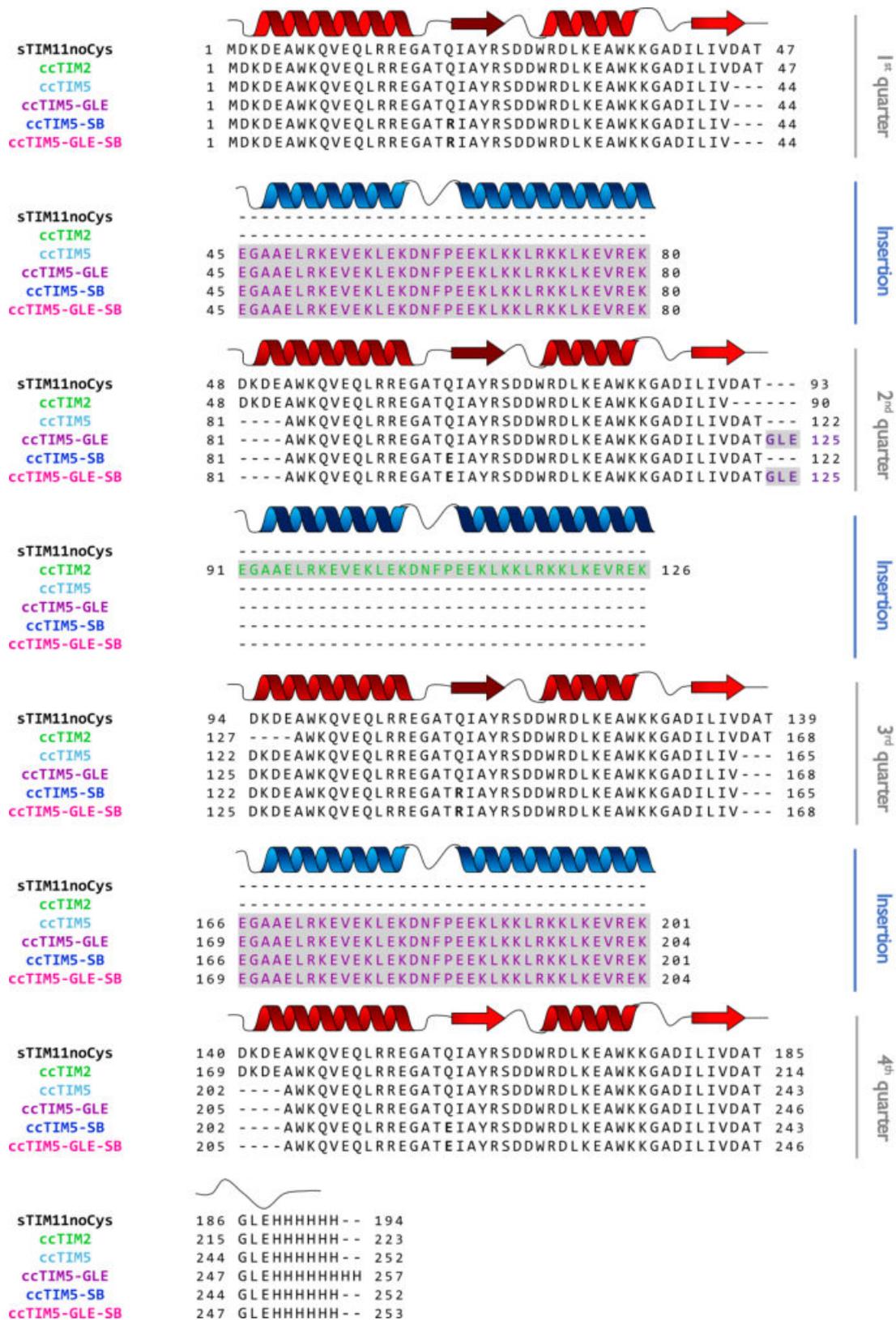


Fig. S1: Sequence alignment of ccTIMs and sTIM11noCys. The according secondary structure of the TIM barrels is shown on top of the sequence, with elements belonging to the TIM-barrel fold in red and coiled coils in blue. Coiled coil sequences are shown with grey background. Notation of the according quarters is shown on the right.

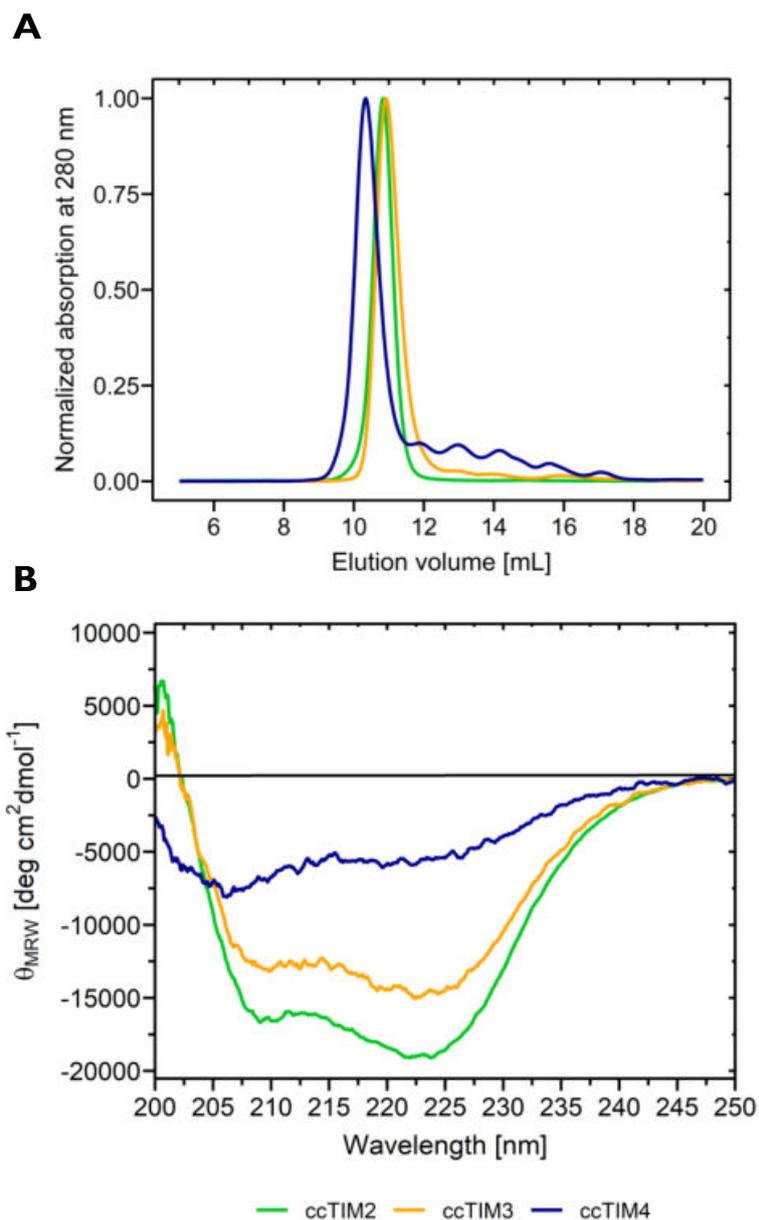


Fig. S2: Analytical SEC and CD spectra of ccTIM2, ccTIM3 and ccTIM4. (A) Comparison of analytical SEC for ccTIM2 (green), ccTIM3 (orange) and ccTIM4 (blue) is shown. Plot showing the elution volume versus the normalized absorption at 280 nm. (B) Comparison of native Far-UV CD spectra of ccTIM2 (green), ccTIM3 (orange) and ccTIM4 (blue). All spectra were collected at 0.2 mg mL⁻¹ and were normalized to mean residue ellipticity.

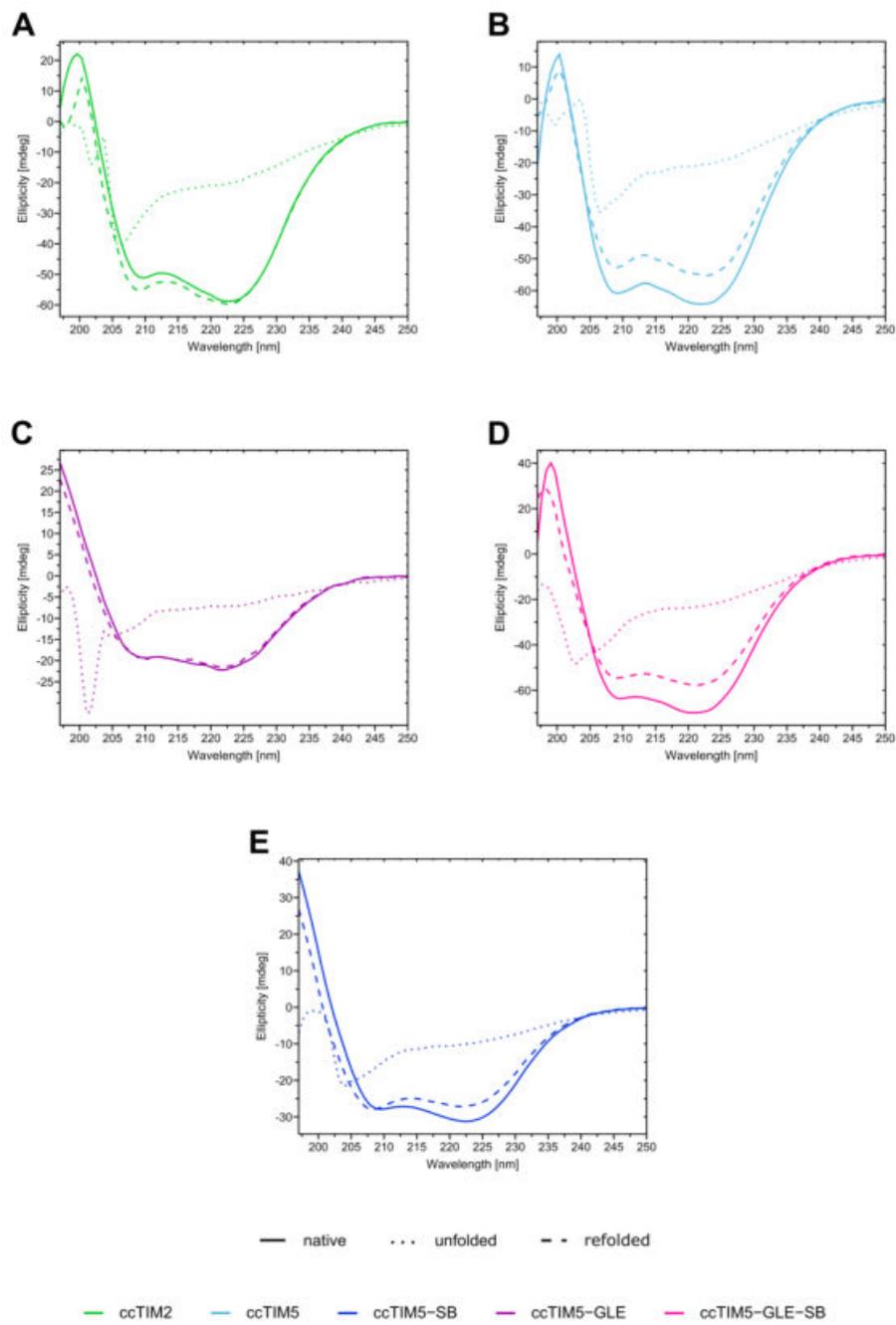


Fig. S2: Far-UV CD spectra of ccTIM variants. Comparison between the native protein (solid line), unfolded protein at 95 °C (dotted line) and refolded protein at 20 °C (dashed line) is shown for (A) ccTIM2, (B) ccTIM5, (C) ccTIM5-GLE, (D) ccTIM5-GLE-SB and (E) ccTIM5-SB.

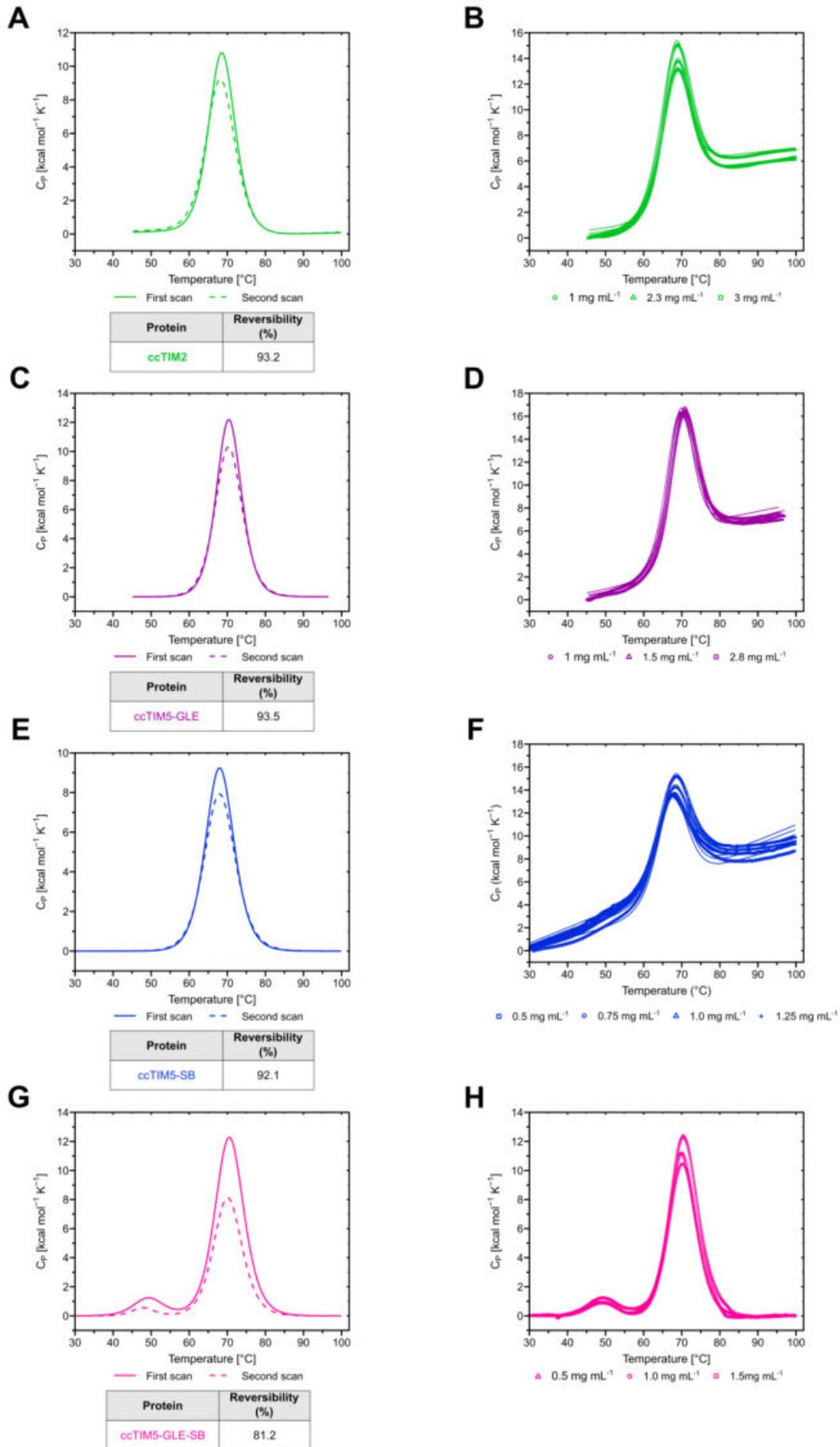


Fig S3: Thermal unfolding of ccTIM variants followed by DSC. (A, C, E, G) Thermal unfolding reversibility was determined by the recovered area percentage by comparing the first (continuous line) and second scan (dashed line). (B, C, F, H) Thermodynamic parameters were calculated fitting the endotherms collected at different protein concentrations to a reversible two-state model (continuous lines in B, C, F). (A,B) ccTIM2; (C,D) ccTIM5-GLE, (E,F) ccTIM5-SB and (G,H) ccTIM5-GLE-SB.

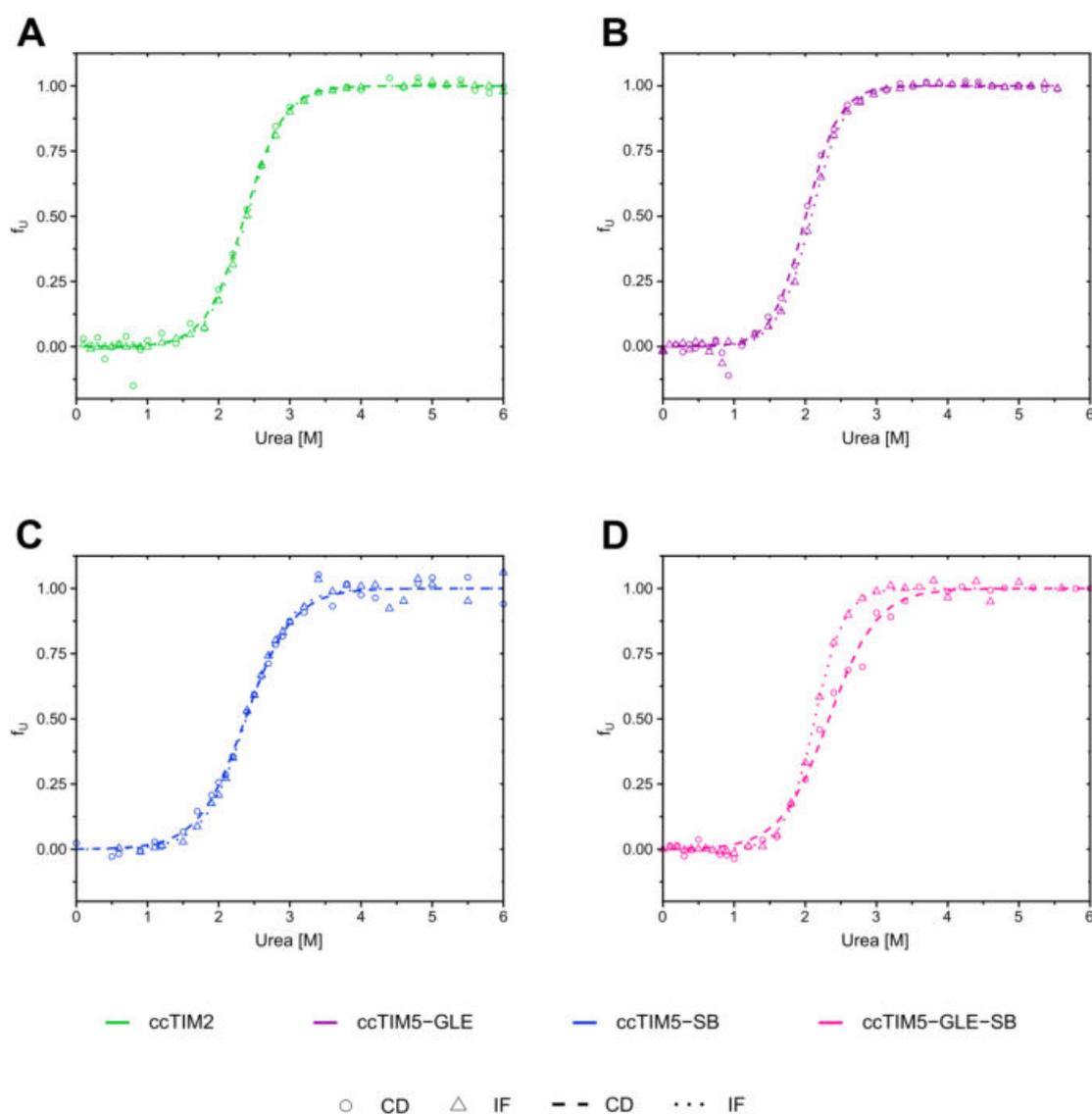


Fig S4: Chemical unfolding of ccTIMs followed by CD (normalized data). Chemical unfolding with urea at 25 °C for ccTIM variants. Circles represent CD data and triangles fluorescence data. Separated fit of CD (dashed lines) and IF (dotted lines) data was performed. High agreement of both fits is shown for ccTIM2 (A), ccTIM5-GLE (B) and ccTIM5-SB (C). For ccTIM5-GLE-SB (D) fits differ significantly.

Paper IV

Sergio Romero-Romero*, **Sina Kordes***, Florian Michel*, Birte Höcker
Evolution, folding, and design of TIM barrels and related proteins
Current Opinion in Structural Biology, 2021, 68, pp. 94-104
*equal contribution



Evolution, folding, and design of TIM barrels and related proteins

Sergio Romero-Romero¹, Sina Kordes¹, Florian Michel¹ and Birte Höcker



Proteins are chief actors in life that perform a myriad of exquisite functions. This diversity has been enabled through the evolution and diversification of protein folds. Analysis of sequences and structures strongly suggest that numerous protein pieces have been reused as building blocks and propagated to many modern folds. This information can be traced to understand how the protein world has diversified. In this review, we discuss the latest advances in the analysis of protein evolutionary units, and we use as a model system one of the most abundant and versatile topologies, the TIM-barrel fold, to highlight the existing common principles that interconnect protein evolution, structure, folding, function, and design.

Address

Department of Biochemistry, University of Bayreuth, 95447 Bayreuth, Germany

Corresponding author: Höcker, Birte (birte.hoecker@uni-bayreuth.de)

¹ Equal contribution.

Current Opinion in Structural Biology 2021, **68**:94–104

This review comes from a themed issue on **Sequences and topology**

Edited by **Nir Ben-Tal** and **Andrei N Lupas**

For a complete overview see the [Issue](#) and the [Editorial](#)

Available online 13th January 2021

<https://doi.org/10.1016/j.sbi.2020.12.007>

0959-440X/© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Introduction

Structural and functional diversity in modern proteins is the result of diversification and optimization processes over the course of evolution. Studying these processes is useful to evaluate how different molecular mechanisms, like duplication and recombination, shape biophysical properties in proteins. Sequence and structural analysis suggest that numerous protein pieces, considered as evolutionary units, have been reused and combined to create higher complexity. In this context, what are the reasons for the recurring success of some of these units? What is their role in protein fold diversification? And how can we use the accumulated information to further our protein design goals?

In this review, we try to unravel these mysteries by integrating different perspectives and approaches (Figure 1). We first discuss the current views of evolutionary units (Section ‘Current views of evolutionary units’). Then, we use the TIM-barrel fold as model system to analyze how our knowledge of the protein-based world is enhanced by the integration of evolutionary analysis (Section ‘Evolutionary events: fragments and natural TIM-barrel proteins’), experimental recreation of evolutionary events (Section ‘Recreating evolutionary events in the lab: chimeragenesis and directed evolution’), folding-function-fitness studies (Section ‘Three *f* determinants in TIM-barrel evolution: folding, function, and fitness’), and protein design approaches (Section ‘Learning from nature towards protein design’). We illustrate how these studies pave the way to a detailed description of existing structure-folding-function-fitness relationships and also boost the design of new proteins with novel molecular properties.

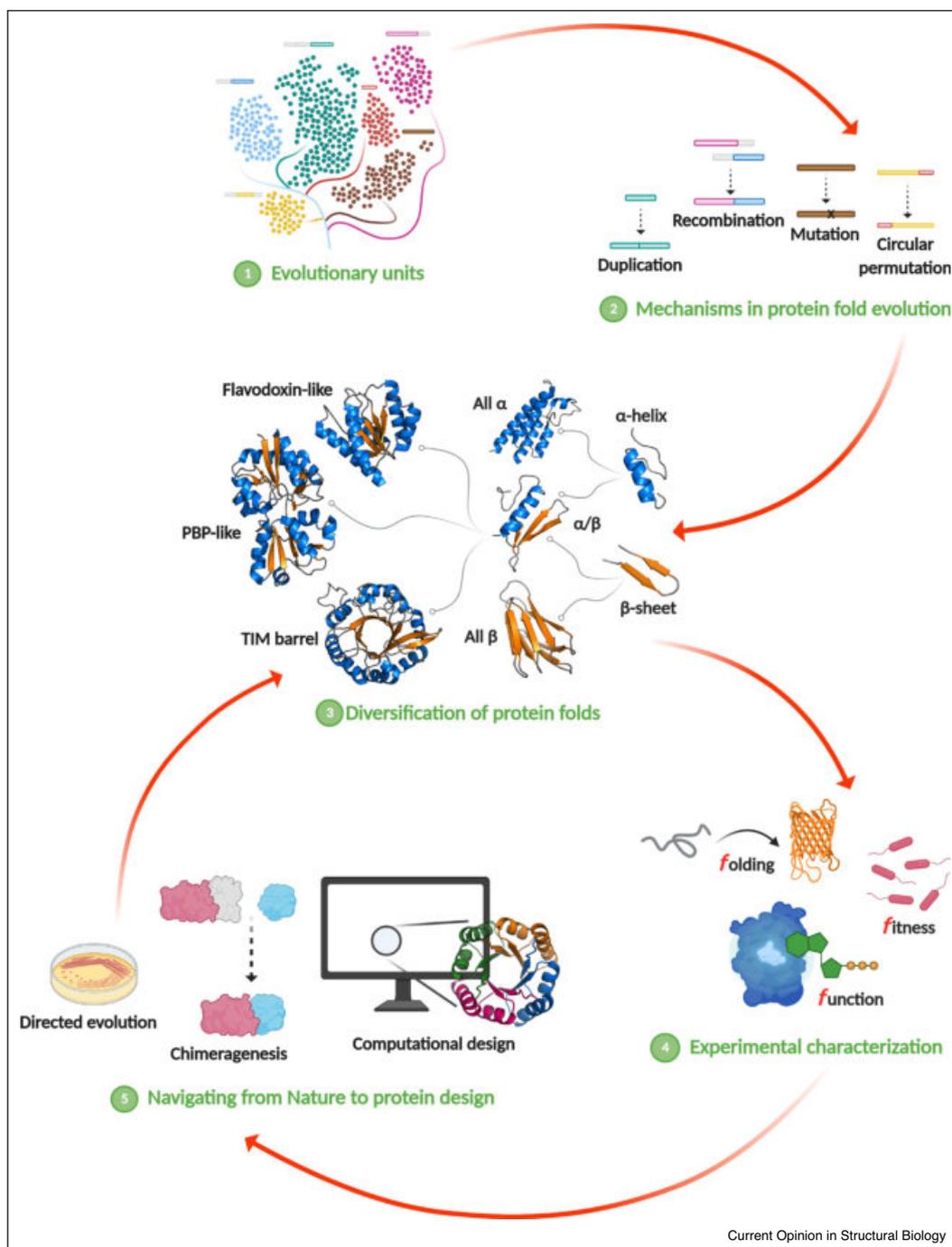
Current views of evolutionary units

Look at any protein and you are bound to find pieces that appear to have been reused either in different proteins or as the modules in a repeat protein. Clearly, reuse of sequences is ubiquitous within the natural fold space as was suggested already early on [1,2]. For protein scientists this beckons the question: how many of these pieces are there and what makes them so successful?

The structural annotation of proteins typically includes consulting at least one of the major databases SCOP, CATH or ECOD [3–5] to append additional information on evolutionary relationships. Molecular evolution studies have shown that different forces and mechanisms such as mutations, duplications, recombinations, deletions, and circular permutations drive the diversification of the protein-based world [6,7]. These mechanisms also hold true for events in the subdomain regime.

In recent years there have been several approaches to define subdomain units as distinguishable building blocks (Figure 2). For example, an evolutionary relationship between the TIM-barrel and flavodoxin-like folds based on a 40-residue fragment was identified by sequence searches [8]. In a large-scale approach, Alva *et al.* identified and defined the reuse of elements within all modern proteins [9]. They generated a vocabulary of 40 subdomain fragments of up to 38 residues, which occur within a

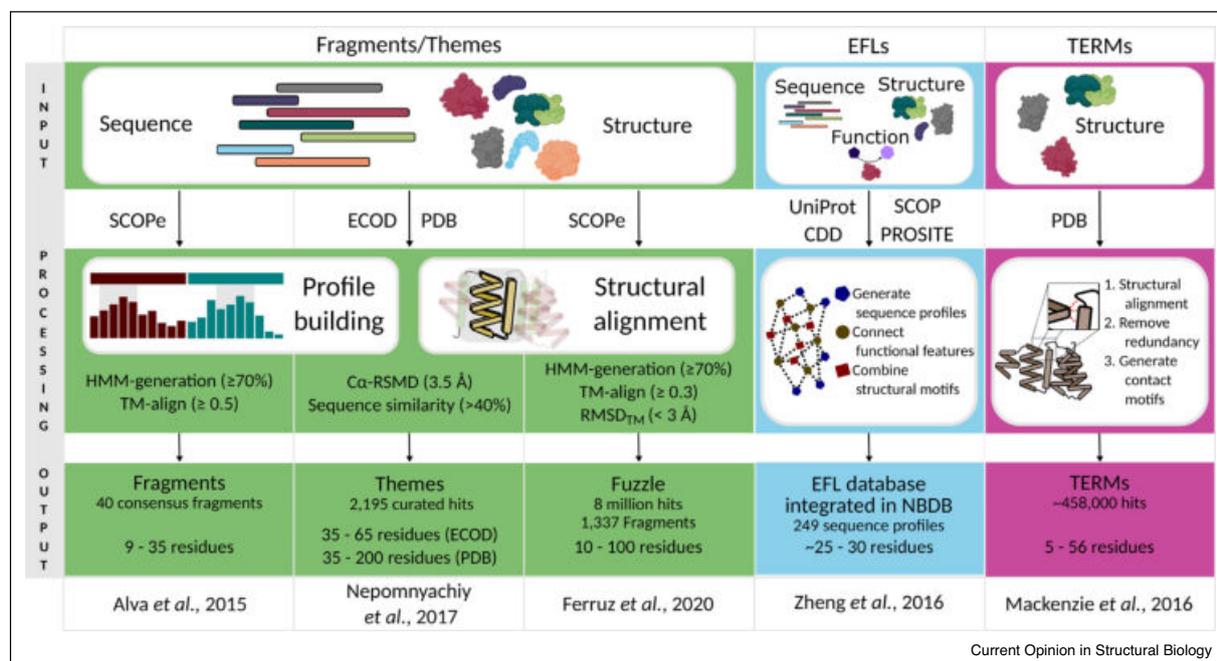
Figure 1



Schematic overview of the relationships between protein fold evolution, experimental characterization, and design approaches discussed in this review. The upper part of the figure shows how evolutionary units are reused through different molecular mechanisms to diversify protein folds. Experimental reconstruction of different evolutionary pathways and the analysis of folding, function, and fitness determinants in evolution increase our knowledge of the protein-based world and allow navigating from Nature to protein design as shown in the bottom part.

96 Sequences and topology

Figure 2



Current subdomain classification approaches. Shown is the generation of available subdomain databases including the different input, data processing, and final output. While Fragment/Themes are continuous sequences and are defined by HMM-profile comparisons and structural alignments, TERMS are non-continuous and focus on contact maps for classification. In contrast, EFLs combine information from structure, sequence and function, but are limited by existing annotation of functional sites.

great number of different folds. Subsequent efforts to expand on these initial fragments led to the description of *themes* – reused fragments of at least 35 residues [10^{**}]. A *theme* is defined whenever a sensitive sequence search using HHsearch suggests remote homology.

Along the same lines, Ferruz *et al.* expanded the fragment universe applying a set of filters to ensure the fragments are related, but not restricting their length [11^{**}]. This generated a dataset of over eight million hits, which are summarized in the *Fuzzle* database (<https://fuzzle.uni-bayreuth.de>). When visualizing the dataset in a network representation a major component is observed that includes many hits between folds thought to be ancestral reinforcing earlier observations on different datasets [12,13]. This might hint not only to a common evolutionary history, but also to the existence of a favorable set of rules for protein folding, function, and fitness.

Another description by Berezovsky defines *elementary functional loops* (EFLs) [14]. These EFLs describe stretches of proteins with a specific sequence profile thought to be defined by the polymer nature of the polypeptide as reviewed recently [15]. Combining this with information

on the conservation of structure and function provides indications, which elements might have proven successful in a primordial peptide-stage of evolution. This concept has been employed for example in the *nucleotide binding database* (NBDB), which contains EFLs involved in binding nucleotide-containing ligands [16]. Phosphate binding signatures obtained by this database were applied in the design of a P-loop protein testing the role of polymer physics in the emergence of basic units of proteins [17].

A fourth view that does not necessarily focus on the evolutionary aspect but rather on protein fold space are the *tertiary structural motifs* (TERMs) [18]. TERMS are 5–56 residue-long, discontinuous structural entities that are generated solely by comparing their environment. While TERMS focus primarily on conserved structural environments, a comparison of motifs generated by simulated evolution on TERMS and those of their natural counterparts showed that TERMS were able to accurately describe nature-like sequence variation.

These examples of either using structural information alone or sensitive in-depth sequence analysis or a combination thereof clearly hint to one thing: there is a subset of

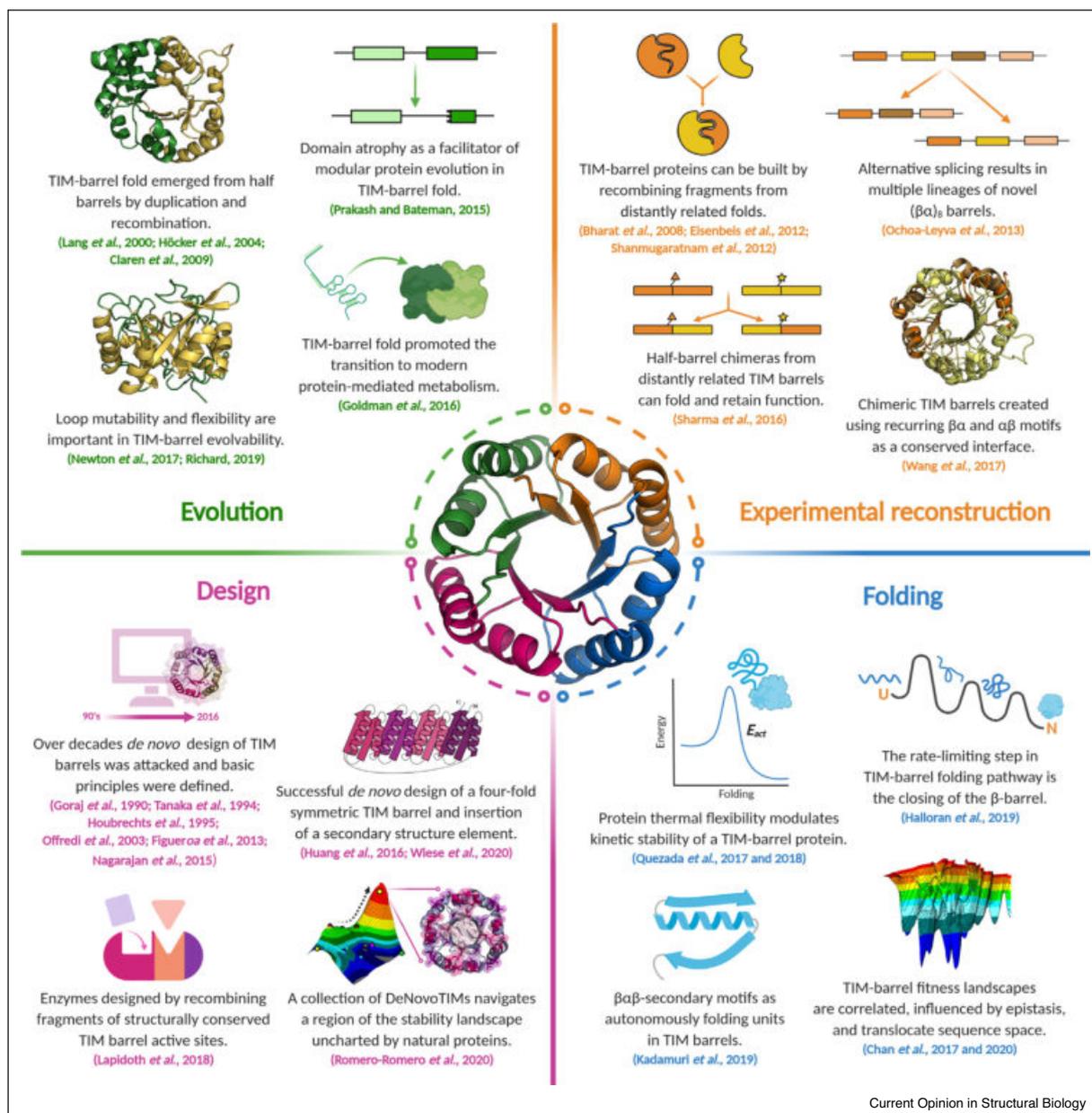
successful ancestral sequences that are to this day propagated to many modern folds.

Evolutionary events: fragments and natural TIM-barrel proteins

The previous section showed that, even after a considerable timespan, we can detect evolutionary relationships in

modern proteins. Can we decode the underlying mechanisms of conservation of subdomain fragments in natural proteins? This general question has been explored by analyzing the evolution of different protein folds, for which the TIM barrel is a model system (Figure 3). This fold is regarded to be one of the oldest and encompasses a wide variety of known protein functions [19–21]. Its

Figure 3



Summary of recent central studies that interconnect the evolution, its experimental reconstruction, folding, and design of TIM-barrel proteins as discussed in this review.

98 Sequences and topology

canonical fold consists of a central eight-stranded, parallel β -barrel surrounded by eight α -helices forming the eponymous $(\beta\alpha)_8$ -barrel structure. It has previously been shown that subdomain parts of the TIM-barrel fold present an excellent model to probe the role of subdomain events, but also explore its evolution [22].

In a recent endeavor, Kadamuri *et al.* theorized that a set of $\beta\alpha\beta$ sequences exists within the TIM-barrel fold-space, which would be autonomously folding units [23]. While there are not yet any reports of natural $\beta\alpha\beta$ motifs folding in isolation, investigating the subdomain folding regime in TIM barrels might reveal crucial steps to improve the creation of novel proteins and help elucidate the evolution of protein domains themselves.

A study by Michalska *et al.* on the structural flexibility of naturally occurring TIM barrels reported a 3D-domain swap of an $(\alpha\beta)_2$ element within a tryptophan synthase structure [24]. A similar event has recently been observed in a crystal structure of the archaeal chemotaxis protein CheY [25]. An analysis of alternative splicing events of $(\beta\alpha)_8$ barrels within the human genome also showed a considerable fraction expressing only as subdomains, and are thought to assemble to a complete barrel with their complementary partners [26]. These observations hint at a flexible subdomain composition within α/β proteins. This concept has been experimentally explored as will be discussed further in Section ‘Recreating evolutionary events in the lab: chimeragenesis and directed evolution’.

When Prakash and Bateman analyzed the variation of TIM-barrel domain boundaries, they found what they propose to be *domain atrophy* [27]. This rare event is characterized by a loss of core secondary structure features that is potentially detrimental to domain stability. While it is still not clear why such events are evolutionary fixed, a possible rescue of stability appears to be the formation of protein-protein interactions, for example, in homodimers.

All these examples of subdomain evolutionary events in the TIM-barrel fold point to one thing: there is a propensity of some proteins to swap subdomain elements. To really gauge if this subdomain recombination played — or still plays — an important role in the diversification of proteins, more protein folds need to be examined. Understanding the common principles that govern this process could help improve our knowledge of protein stability, folding, function and evolution.

Recreating evolutionary events in the lab: chimeragenesis and directed evolution

The enormous diversity of protein structures and functions can be interpreted as the result of a massive *experiment* that has been carried out by Nature in a sustained way for millions of years, whose results are observed in the

broad number of protein sequences and structures. In the previous section, we discussed that diverse evolutionary events in natural proteins allow the expansion of the protein fold space. Now, we focus on how some of these evolutionary events can be recreated in the laboratory through chimeragenesis and directed evolution. Both approaches offer a good alternative to test evolutionary and thermodynamic hypotheses and also to generate novel proteins (Figure 3).

Newton *et al.* explored the evolution of the TIM-barrel enzyme HisA using directed evolution techniques [28**]. They follow up on the innovation-amplification-divergence model previously proposed as an explanation of how gene duplication leads to proteins with new functions [29]. They show how beneficial substitutions selected during real-time evolution can result in manifold changes in enzyme function and bacterial fitness. The results emphasize the importance of loop mutability and confirms the TIM barrel as an inherently evolvable protein scaffold.

The current evolutionary hypothesis about the emergence of the TIM-barrel fold is that it evolved from duplication and fusion events of a half barrel, that is, a $(\beta\alpha)_4$, or even smaller units [30–33]. This possible pathway has been tested computationally and experimentally by analyzing sequence, structural, and folding properties [32,34,35]. Following this idea, Sharma *et al.* engineered and characterized active and stable chimeric TIM barrels of two distantly related glycosyl hydrolases, demonstrating that half-barrel domains from different sources can assemble and adopt the pre-evolved function [36]. Likewise, Almeida *et al.* tested the idea that $(\beta\alpha)_4$ halves are self-contained evolutionary units, independent of their size and internal symmetry. They introduced mutations in the inter-half contacts of a β -glucosidase to obtain independent half barrels that unfold cooperatively [37]. Further, Wang *et al.* identified physicochemical properties from a set of non-redundant TIM-barrel proteins that strongly support the existence of recurring $\beta\alpha$ and $\alpha\beta$ motifs in this fold [38]. In addition, using a conserved $\alpha\beta\alpha$ element as a recombination site, they created a chimeric protein from two different TIM barrels, highlighting the potential of recurring motifs as naturally optimized interfaces to engineer well-folded chimeras.

Inspired by TIM-barrel modularity, Lapidoth *et al.* designed highly active and stable enzymes by creating fragments of structurally conserved sites of two unrelated TIM-barrel families and then assembled them to create a large set of combinatorial backbones [39**]. The reported computational approach mimics natural evolutionary processes such as recombinations, insertions, deletions, and mutations, but it is more radical than these individual events since all of them are applied simultaneously to modify the protein fitness.

As will be discussed in the last section (Learning from nature towards protein design), this method could be extended to create new biocatalysts by combining more distantly related families.

Apart from recombination events within a protein fold, recombination of heterologous structural motifs of unrelated folds is possible. Although difficult to detect in Nature, the idea can be tested in the laboratory and might be used to design proteins with novel biophysical properties [21]. In this context, ElGamacy *et al.* engineered an asymmetric dRP lyase fold fusing two heterologous and unrelated supersecondary structures. After interface optimization the approach generated a stable chimera with high precision to the original design [40*].

Similarly, we have used chimeragenesis in the past to elucidate evolutionary relationships of several α/β folds and design new proteins. Chimeras built combining parts of the flavodoxin-like proteins CheY or NarL with a piece of the TIM barrel HisF demonstrate that $(\beta\alpha)_8$ -barrel proteins can be constructed by recombining a large repertoire of natural protein fragments from distantly related folds [8,41–43]. This interchangeability offers a great opportunity to retrace early evolutionary steps. Following up on this, Toledo-Patiño *et al.* found sequence-based evidence that the singleton HemD-like fold emerged from the flavodoxin-like fold [44*]. To test the hypothesized path, consisting of insert-assisted segment swap, gene duplication, and fusion, these evolutionary events were experimentally reverted, yielding well-folded and stable proteins. The results strongly support the emergence of the HemD-like fold from flavodoxin-like proteins and highlight the importance of duplication and fusion as evolutionary events that allow the creation of complex proteins. These experimental reconstructions of possible evolutionary events fit well with the bioinformatic studies on protein fragments as discussed in section ‘Current views of evolutionary units’. Databases such as *Fuzzle* [11**] provide many starting points for similar evolutionary explorations and open new ways to use already existing sequences in protein design. Fragments identified in *Fuzzle* can be used directly in the tool *Protlego* (<https://hoecker-lab.github.io/protlego/>) for automated chimera design and analysis [45].

Three *f* determinants in TIM-barrel evolution: folding, function, and fitness

The evolutionary study of biophysical determinants is useful to evaluate the role evolution has on the physical properties of proteins and informs us on how changes in the amino acid sequence shaped function in a specific fold [46]. In this section, we focus on recent advances to understand the biophysical basis underlying the success of the TIM-barrel fold as one of the most robust and versatile scaffolds.

The TIM-barrel fold provides a good architecture to explore how folding mechanisms have been conserved or diverged during evolution (Figure 3). In this context, Halloran *et al.* analyzed on a molecular level the earliest events in the folding of a TIM-barrel protein [47**]. Experimental and computational approaches revealed that the kinetic intermediate commonly observed in TIM barrels is dominated by a native-like structure in the central region of the sequence. They determined the rate-limiting step in the folding pathway to be the frustration encountered by the competition between the N-terminus or C-terminus to close the internal β -barrel. Also analyzing TIM-barrel proteins, Romero-Romero *et al.* studied and compared the folding pathway of eukaryotic homologous triosephosphate isomerases. Structural and biophysical analysis suggested that interfacial water molecules and water-mediated interactions could modulate the number of equilibrium intermediates, and therefore, the folding pathway in this enzyme family [48].

TIM-barrel proteins are notable for their diversity in catalytic activities. The broad presence of this topology in different enzymes has led to the assumption that the TIM-barrel fold played a central role in early evolution of catalysis. In a bioinformatic study, Goldman *et al.* showed by comparing the functional diversity of different protein folds that TIM-barrel proteins use the broadest range of enzymatic cofactors, including some putatively ancient cofactors [49*]. This supports the idea that the TIM barrel represented an ideal scaffold to facilitate the transition from ribozymes, peptides, and abiotic catalysts to modern protein-mediated metabolism.

Likewise, in terms of protein flexibility and enzymatic catalysis, Richard recently discussed why the selection and optimization of protein folds with multiple flexible loops, such as the TIM-barrel topology, is favored during enzyme evolution [50*]. He proposes that in TIM barrels the exploration of many different conformations during loop movement provides a potential starting point for the evolution of a new enzyme activity and allows the conformational changes needed in floppy enzymes. Also related with protein flexibility, but in the context of stability and evolution, Quezada *et al.* analyzed the molecular basis of the kinetic stability differences of two related triosephosphate isomerases and engineered new functional TIM-barrel enzymes with fine-tuned stabilities [51,52*]. They found a correlation between thermal flexibility and kinetic stability, suggesting how evolution has reached a balance between function and stability in cell-relevant timescales.

The evolution of protein folding, function, and fitness can be seen as a walk through sequence space, in the same way as was described 50 years ago by evolutionary biologist John Maynard Smith in his seminal work about natural selection and the concept of protein space [53].

100 Sequences and topology

Generally, each of these steps can be evaluated in terms of protein fitness, a measure of the effect that a property produces on the overall fitness of an organism. Following this logic, in two subsequent works the Matthews lab performed a quantitative description of the fitness landscape of distant orthologous TIM-barrel proteins to understand their evolutionary dynamics [54^{••},55]. They detected that the fitness landscapes are correlated and influenced by long-range epistatic interactions, and that these landscapes can be translocated in sequence space as a result of TIM-barrel fold plasticity.

The three *f* determinants in evolution discussed in this section have also been analyzed in other protein folds. Examples from the last years include discussions between the Makhataдзе and Sanchez-Ruiz labs about the evolutionary validity of the minimal frustration hypothesis through the experimental characterization of ancestrally reconstructed proteins and extant homologous members of the thioredoxin family [56–58]. Also involving α/β proteins, Kukic *et al.* explored how the folding rates of Procarboxypeptidase A2 can be modulated during evolution by modification of the so-called nucleation-condensation mechanism [59]. Moreover, the Marqusee lab has made a substantial effort to understand how evolutionary pressures modify folding landscapes and tune kinetic and thermodynamic stability by characterizing one of the oldest protein folds, the RNase H-like superfamily [60–63]. Other interesting works are the analysis of the influence of folding energies on the fitness of β -lactamases [64], the study of protein folding and fitness landscapes of amidases [65], the analysis of cotranslational folding and fitness of an integral membrane protein [66], and the evolutionary history of myoglobins [67]. The information obtained both on TIM barrels and other folds has revealed unanticipated details in protein molecular evolution thereby increasing our understanding of sequence-folding-fitness relationships, which has also relevant implications for protein design.

Learning from nature towards protein design

In the previous sections we discussed the evolution of protein folds from smaller units and provided examples recreating such evolutionary events with respect to folding, function, and fitness. Same as protein engineering has been used to test evolutionary hypotheses, the gained knowledge can also be used to design new proteins. Initial protein design strategies were mostly based on parametrization of well understood folds or supersecondary structures. But in the last decades many powerful algorithms were developed to predict protein structures and design new proteins as has been recently reviewed [68].

One of the most widely used design software, namely Rosetta, uses 3-residue and 9-residue long fragments from known protein structures to sample the backbone in *ab initio* predictions [69,70]. Those fragments are a lot

smaller than the previously described evolutionary units [9,10^{••},11^{••},14], however, they still can carry information about possible conformations. Additionally, some algorithms use evolutionary mechanisms as inspiration. The SEWING algorithm for instance incorporates current understanding of protein evolution, the emergence of proteins by recombination and duplication of smaller fragments: sets of structures meeting predefined requirements are generated by recombination of small structural motifs [71]. The more recently developed program dTERMen uses the previously described TERMS by matching them to the target design and thereby determines sequence preferences [72]. Also, the approach from Lapidoth *et al.* mentioned previously is inspired by Nature and mimics evolution during the design process [39^{••}]. The fully automated method combines recombination, insertion, deletion, and mutation events in a non-sequential manner. Initially a predefined set of structures is partitioned and then assembled to combinatorial backbones, which are finally applied to a complete sequence redesign. During this process conserved sites and residues necessary for catalysis or folding can be excluded from the design. In contrast to other enzyme design approaches it has the advantage that no transition state has to be modelled which is computationally expensive. This method was applied to homologous TIM barrels but could possibly be extended to more distantly related proteins, thereby creating new biocatalysts. While this approach, that is based on existing structures, can diversify enzyme function, it will not create proteins from scratch.

The complete *de novo* design of proteins is a task that has been explored and progressed increasingly in recent years fueled by technical advances in structure determination, modelling, and computation. An increase in *de novo* designed proteins could be further observed after Koga *et al.* defined rules for the design of idealized protein topologies as recently reviewed [73]. The value of these design rules, that relate foldability of a tertiary structure to the connection between secondary structure elements [74], in combination with improvements in design algorithms can be traced in the design progression of *de novo* TIM barrels.

Several attempts were made to design a symmetric TIM barrel from scratch to understand what makes this protein fold so successful (Figure 3). In the early 1990s, first symmetric designs were created using statistical information about barrel geometries and amino acid frequencies from few known TIM-barrel structures [75–80]. However, those parameters were not sufficient to achieve designs with natural-like properties as all exhibited molten-globule like states. With an increasing number of TIM-barrel structures, geometric parameters were improved, and newly emerging algorithms were applied to sequence design and created all-atom models. In this

way, the Martial lab was able to improve previous designs and create natural-like proteins [81,82]. Later, the solubility of one of those designs was improved by directed evolution and the three-dimensional structure was determined: it differed from the intended TIM barrel and resembled a Rossmann-like fold [83]. Using the previously described rules for idealized topologies, Nagarajan *et al.* created four-fold symmetric TIM-barrel backbones [84]. Using folding simulations, they determined hydrogen bond networks and enrichment of polar residues in the pore as important features regarding the folding pathway. Those findings were applied during iterative sequence design and resulted in soluble proteins showing cooperative unfolding transitions, though structural studies indicated a molten globule.

In the meantime, Huang *et al.* also applied the rules from Koga *et al.* to design a four-fold symmetric TIM barrel [85]. Their approach sampled backbones with different secondary structure lengths using predefined geometric restrictions followed by iterative sequence design enforcing sidechain-backbone hydrogen bonds. A circular-permuted variant, sTIM11, was soluble expressed and the design was validated by solving its three-dimensional structure. Further analysis revealed a significantly lower conformational stability compared to natural TIM barrels. In a modular approach, a collection of stabilized variants (DeNovoTIMs) was designed by improving hydrophobic packing [86]. Structural and folding analysis showed that epistatic effects allow navigating an unexplored region of the stability landscape of natural proteins. One of these DeNovoTIMs was already used in a successful recombination with a *de novo* designed ferredoxin protein and engineered to bind lanthanide [87]. In another recent study, Wiese *et al.* extended sTIM11 by successfully incorporating a rationally designed small α -helix into a $\beta\alpha$ loop [88]. These works are first steps towards diversifying and ultimately functionalizing *de novo* TIM barrels.

The progression in the design of a TIM barrel reflects nicely the improvements of protein design in the last 30 years. Throughout all design approaches, a symmetric topology was targeted as despite rapidly increasing computational resources the modelling of large proteins is still time-consuming. Further, this process shows how important it is to understand a protein fold in detail and to know which interactions are essential for its stabilization. In this context, it would be interesting to analyze the design from Figueroa *et al.* [82] in detail and determine why this design acquired a different fold than intended [83]. Such analysis is important to improve our understanding and find deficiencies in current protein design strategies.

Additionally, protein design opens a door not only to increase and test our knowledge about folding, function,

and fitness, but also to compare the properties of *de novo* proteins with naturally occurring ones. In this way, studies have shown that *de novo* proteins exhibit more complex folding pathways than natural proteins, as indicated for one of the first *de novo* designed proteins Top7, a $\beta\alpha$ protein [89]. This differs from natural small proteins which show high cooperativity in folding and a smooth free energy surface. In addition, the study of another small *de novo* protein Di-III_14, an IF3-like protein, revealed a more complex folding pathway than initially assumed [90]. In-depth mutational and folding analysis revealed that electrostatic and hydrophobic networks affect the energy surface of this protein. Based on those findings, it was proposed to limit the number of charged amino acids, avoid charge segregation, and use a more diverse set of nonpolar side chains in future protein designs. Overall, these studies demonstrate that as we expand our exploration into sequence space by designing *de novo* proteins, we also expand our understanding of the molecular and physicochemical determinants that shaped and still modulate the protein-based world.

Conclusion and outlook

The study of protein evolution requires the integrated analysis of protein structure and stability, as well as folding, function, and fitness of proteins. There is clear evidence that modern diverse protein folds evolved via reuse of smaller units, which have been identified and described in recent years. Evolution of protein folds from smaller units via duplication has long been described, but also recombination is explored increasingly as an important mechanism. Understanding how protein diversity could emerge via these mechanisms is essential to learn how stable and functional proteins evolved and might be designed.

The ubiquitous TIM-barrel fold has been used in several studies to investigate its evolution, folding, and design. Explorations of the fold's evolutionary history and experiments recreating evolutionary events have revealed how recombination of recurring fragments can lead to new proteins and enzymes. These studies go hand in hand with detailed analyses of protein folding and determination of fitness landscapes of TIM barrels. Moreover, this knowledge has already been applied to the design of *de novo* TIM barrels illustrating how the connection between evolution, folding, and design closes to a cycle and how analysis of designed proteins can help us understand the biophysical properties of proteins even better. Altogether, these recent studies have significantly increased our understanding of the evolution of sequence-structure-function relationships, enabling us to access new protein space through design.

Conflict of interest statement

Nothing declared.

102 Sequences and topology

Acknowledgements

We gratefully acknowledge financial support from the Foundations Alexander von Humboldt and Bayer Science & Education (Humboldt-Bayer Research Fellowship for Postdoctoral Researchers to S.R.R.), from the European Research Council (ERC Consolidator grant 647548 'Protein Lego'), and the Volkswagenstiftung (grant 94747). Figures were created with [BioRender.com](https://www.biorender.com) and Pymol.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Eick RV, Dayhoff MO: **Evolution of the structure of ferredoxin based on living relics of primitive amino acid sequences.** *Science* 1966, **152**:363-366.
 2. Fetrow JS, Godzik A: **Function driven protein evolution. A possible proto-protein for the RNA-binding proteins.** *Pac Symp Biocomput* 1998.
 3. Andreeva A, Howorth D, Chothia C, Kulesha E, Murzin AG: **SCOP2 prototype: a new approach to protein structure mining.** *Nucleic Acids Res* 2014, **42**:310-314.
 4. Sillitoe I, Dawson N, Lewis TE, Das S, Lees JG, Ashford P, Tolulope A, Scholes HM, Senatorov I, Bujan A *et al.*: **CATH: expanding the horizons of structure-based functional annotations for genome sequences.** *Nucleic Acids Res* 2019, **47**:D280-D284.
 5. Cheng H, Schaeffer RD, Liao Y, Kinch LN, Pei J, Shi S, Kim BH, Grishin NV: **ECOD: an evolutionary classification of protein domains.** *PLoS Comput Biol* 2014, **10**:e1003926.
 6. Ohta T: **Mechanisms of molecular evolution.** *Philos Trans R Soc B Biol Sci* 2000, **355**:1623-1626.
 7. Sikosek T, Chan HS: **Biophysics of protein evolution and evolutionary protein biophysics.** *J R Soc Interface* 2014, **11**:20140419.
 8. Fariás-Rico JA, Schmidt S, Höcker B: **Evolutionary relationship of two ancient protein superfolds.** *Nat Chem Biol* 2014, **10**:710-715.
 9. Alva V, Söding J, Lupas AN: **A vocabulary of ancient peptides at the origin of folded proteins.** *eLife* 2015, **4**:e09410.
 10. Nepomnyachiy S, Ben-Tal N, Kolodny R: **Complex evolutionary footprints revealed in an analysis of reused protein segments of diverse lengths.** *Proc Natl Acad Sci U S A* 2017, **114**:11703-11708.
- In this bioinformatic analysis the reuse of segments, which are similar in sequence and structure, is described. These segments, called themes, are subdomain elements with reuse traced to the amino acid position, highlighting the impact of protein evolution.
11. Ferruz N, Lobos F, Lemm D, Toledo-Patino S, Fariás-Rico JA, Schmidt S, Höcker B: **Identification and analysis of natural building blocks for evolution-guided fragment-based protein design.** *J Mol Biol* 2020, **432**:3898-3914.
- Following an all-vs-all comparison of protein domain sequences, reused protein fragments were compiled into a network. This customizable network of fragments not only carries evolutionary significance, but can also function as a starting point for protein design by recombination.
12. Alva V, Remmert M, Biegert A, Lupas AN, Söding J: **A galaxy of folds.** *Protein Sci* 2010, **19**:124-130.
 13. Nepomnyachiy S, Ben-Tal N, Kolodny R: **Global view of the protein universe.** *Proc Natl Acad Sci U S A* 2014, **111**:11691-11696.
 14. Berezovsky IN, Guarnera E, Zheng Z: **Basic units of protein structure, folding, and function.** *Prog Biophys Mol Biol* 2017, **128**:85-99.
 15. Berezovsky IN: **Towards descriptor of elementary functions for protein design.** *Curr Opin Struct Biol* 2019, **58**:159-165.
 16. Zheng Z, Goncarenco A, Berezovsky IN: **Nucleotide binding database NBDB - a collection of sequence motifs with specific protein-ligand interactions.** *Nucleic Acids Res* 2016, **44**:D301-D307.
 17. Romero Romero ML, Yang F, Lin Y-R, Toth-Petroczy A, Berezovsky IN, Goncarenco A, Yang W, Wellner A, Kumar-Deshmukh F, Sharon M *et al.*: **Simple yet functional phosphate-loop proteins.** *Proc Natl Acad Sci U S A* 2018, **115**:E11943-E11950.
 18. MacKenzie CO, Zhou J, Grigoryan G: **Tertiary alphabet for the observable protein structural universe.** *Proc Natl Acad Sci U S A* 2016, **113**:E7438-E7447.
 19. Banner DW, Bloomer AC, Petsko GA, Phillips DC, Pogson CI, Wilson IA, Corran PH, Furth AJ, Milman JD, Offord RE *et al.*: **Structure of chicken muscle triose phosphate isomerase determined crystallographically at 2.5 Å resolution: using amino acid sequence data.** *Nature* 1975, **255**:609-614.
 20. Nagano N, Orenge CA, Thornton JM: **One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions.** *J Mol Biol* 2002, **321**:741-765.
 21. Sterner R, Höcker B: **Catalytic versatility, stability, and evolution of the (β_α)₈-barrel enzyme fold.** *Chem Rev* 2005, **105**:4038-4055.
 22. Höcker B: **Design of proteins from smaller fragments-learning from evolution.** *Curr Opin Struct Biol* 2014, **27**:56-62.
 23. Kadamuri RV, Irukuvajjala SS, Vadrevu R: **β_αβ super-secondary motifs: sequence, structural overview, and pursuit of potential autonomously folding β_αβ sequences from (β_α)₈/TIM barrels.** *Methods in Molecular Biology.* Humana Press Inc; 2019:221-236.
 24. Michalska K, Kowiel M, Bigelow L, Endres M, Gilski M, Jaskolski M, Joachimiak A: **3D domain swapping in the TIM barrel of the α subunit of Streptococcus pneumoniae tryptophan synthase.** *Acta Crystallogr Sect D Struct Biol* 2020, **76**:166-175.
 25. Paithankar KS, Enderle M, Wirthensohn DC, Miller A, Schlesner M, Pfeiffer F, Rittner A, Gringer M, Oesterheld D: **Structure of the archaeal chemotaxis protein CheY in a domain-swapped dimeric conformation.** *Acta Crystallogr Sect F Struct Biol Commun* 2019, **75**:576-585.
 26. Ochoa-Leyva A, Montero-Morán G, Saab-Rincón G, Briebe LG, Soberón X: **Alternative splice variants in TIM barrel proteins from human genome correlate with the structural and evolutionary modularity of this versatile protein fold.** *PLoS One* 2013, **8**:e70582.
 27. Prakash A, Bateman A: **Domain atrophy creates rare cases of functional partial protein domains.** *Genome Biol* 2015, **16**:88.
 28. Newton MS, Guo X, Söderholm A, Näsval J, Lundström P, Andersson DI, Selmer M, Patrick WM: **Structural and functional innovations in the real-time evolution of new (β_α)₈ barrel enzymes.** *Proc Natl Acad Sci U S A* 2017, **114**:4727-4732.
- In this work, a real-time evolution analysis is performed to understand how new TIM-barrel enzymes lead to phenotype and organismal fitness changes. The study details the structural and functional innovations of the HisA enzyme towards generalist or specialist activities providing clues about evolution from atomic to whole-organism levels.
29. Näsval J, Sun L, Roth JR, Andersson DI: **Real-time evolution of new genes by innovation, amplification, and divergence.** *Science* 2012, **338**:384-387.
 30. Lang D, Thoma R, Henn-Sax M, Sterner R, Wilmanns M: **Structural evidence for evolution of the β/α barrel scaffold by gene duplication and fusion.** *Science* 2000, **289**:1546-1550.
 31. Gerlt JA, Raushel FM: **Evolution of function in (β/α)₈-barrel enzymes.** *Curr Opin Chem Biol* 2003, **7**:252-264.
 32. Höcker B, Claren J, Sterner R: **Mimicking enzyme evolution by generating new (betaalpha)₈-barrels from (betaalpha)₄-half-barrels.** *Proc Natl Acad Sci U S A* 2004, **101**:16448-16453.
 33. Claren J, Malisi C, Höcker B, Sterner R: **Establishing wild-type levels of catalytic activity on natural and artificial (β_α)₈-barrel protein scaffolds.** *Proc Natl Acad Sci U S A* 2009, **106**:3704-3709.

34. Höcker B, Beismann-Driemeyer S, Hettwer S, Lustig A, Sterner R: **Dissection of a (β/α)₈-barrel enzyme into two folded halves.** *Nat Struct Biol* 2001, **8**:32-36.
35. Seitz T, Bocola M, Claren J, Sterner R: **Stabilisation of a (β/α)₈-barrel protein designed from identical half barrels.** *J Mol Biol* 2007, **372**:114-129.
36. Sharma P, Kaila P, Guptasarma P: **Creation of active TIM barrel enzymes through genetic fusion of half-barrel domain constructs derived from two distantly related glycosyl hydrolases.** *FEBS J* 2016, **283**:4340-4356.
37. Almeida VM, Frutuoso MA, Marana SR: **Search for independent (β/α)₄ subdomains in a (β/α)₈ barrel β -glucosidase.** *PLoS One* 2018, **13**:e0191282.
38. Wang JJ, Zhang T, Liu R, Song M, Wang JJ, Hong J, Chen Q, Liu H: **Recurring sequence-structure motifs in (β/α)₈-barrel proteins and experimental optimization of a chimeric protein designed based on such motifs.** *Biochim Biophys Acta - Proteins Proteomics* 2017, **1865**:165-175.
39. Lapidtho G, Khersonsky O, Lipsh R, Dym O, Albeck S, Rogotner S, Fleishman SJ: **Highly active enzymes by automated combinatorial backbone assembly and sequence design.** *Nat Commun* 2018, **9**:2780
- A design approach is reported that uses protein fragments to create combinatorial backbones by applying evolutionary mechanisms during the assembly. This is applied to two homologous TIM-barrel families to build enzymes, which either restore or even increase the parental activities.
40. ElGamacy M, Coles M, Lupas A: **Asymmetric protein design from conserved supersecondary structures.** *J Struct Biol* 2018, **204**:380-387
- The authors explored the combination of heterologous structural motifs as a new potential mechanism in protein fold evolution. Replacement of an α -hairpin motif from an unrelated family followed by interface optimization led to a stable, well-folded dRP-lyase protein verified by NMR.
41. Bharat TAM, Eisenbeis S, Zeth K, Höcker B: **A β α -barrel built by the combination of fragments from different folds.** *Proc Natl Acad Sci U S A* 2008, **105**:9942-9947.
42. Eisenbeis S, Proffitt W, Coles M, Truffault V, Shanmugaratnam S, Meiler J, Höcker B: **Potential of fragment recombination for rational design of proteins.** *J Am Chem Soc* 2012, **134**:4019-4022.
43. Shanmugaratnam S, Eisenbeis S, Höcker B: **A highly stable protein chimera built from fragments of different folds.** *Protein Eng Des Sel* 2012, **25**:699-703.
44. Toledo-Patiño S, Chaubey M, Coles M, Höcker B: **Reconstructing the remote origins of a fold singleton from a flavodoxin-like ancestor.** *Biochemistry* 2019, **58**:4790-4793
- A hypothetical evolutionary pathway is reconstructed supporting the emergence of the HemD-like fold from the flavodoxin-like fold. This work exemplifies how protein fold evolution can be reconstructed in the lab using modern day proteins.
45. Ferruz N, Noske J, Höcker B: **Protlego: a python package for the analysis and design of chimeric proteins.** *bioRxiv* 2020 <http://dx.doi.org/10.1101/2020.10.04.325555>.
46. Harms MJ, Thornton JW: **Evolutionary biochemistry: revealing the historical and physical causes of protein properties.** *Nat Rev Genet* 2013, **14**:559-571.
47. Halloran KT, Wang Y, Arora K, Chakravarthy S, Irving TC, Bilsel O, Brooks CL, Robert Matthews C: **Frustration and folding of a TIM barrel protein.** *Proc Natl Acad Sci U S A* 2019, **116**:16378-16383
- This work combines computational simulations with experiments to monitor the folding pathway of indole-3-glycerol phosphate synthase. A common structurally-stable intermediate is formed early in folding and it can be concluded that the rate-limiting step in the folding pathway is the closing of the β -barrel.
48. Romero-Romero S, Becerril-Sesin LA, Costas M, Rodríguez-Romero A, Fernández-Velasco DA: **Structure and conformational stability of the triosephosphate isomerase from *Zea mays*. Comparison with the chemical unfolding pathways of other eukaryotic TIMs.** *Arch Biochem Biophys* 2018, **658**:66-76.
49. Goldman AD, Beatty JT, Landweber LF: **The TIM barrel architecture facilitated the early evolution of protein-mediated metabolism.** *J Mol Evol* 2016, **82**:17-26
- This bioinformatic work studies the role of the TIM-barrel fold in early evolution. Analysis of function, cofactor usage, and metabolic pathways suggested that TIM-barrel proteins participated in the transition from non-peptidic catalysis to protein-mediated metabolism.
50. Richard JP: **Protein flexibility and stiffness enable efficient enzymatic catalysis.** *J Am Chem Soc* 2019, **141**:3320-3331
- The author discusses how protein flexibility is an important factor in TIM-barrel enzymes to find a balance between substrate-binding energy and catalysis. It is suggested that selection of flexible loops provides a starting point for the evolution and divergence of new enzyme activities.
51. Quezada AG, Díaz-Salazar AJ, Cabrera N, Pérez-Montfort R, Piñeiro Á, Costas M: **Interplay between protein thermal flexibility and kinetic stability.** *Structure* 2017, **25**:167-179.
52. Quezada AG, Cabrera N, Piñeiro Á, Díaz-Salazar AJ, Díaz-Mazariegos S, Romero-Romero S, Pérez-Montfort R, Costas M: **A strategy based on thermal flexibility to design triosephosphate isomerase proteins with increased or decreased kinetic stability.** *Biochem Biophys Res Commun* 2018, **503**:3017-3022
- Two closely-related triosephosphate isomerases are used to explore the correlation between protein thermal flexibility and kinetic stability. Based on MD simulations and DSC experiments, the authors designed new functional TIM-barrel enzymes with fine-tuned kinetic stabilities.
53. Smith JM: **Natural selection and the concept of a protein space.** *Nature* 1970, **225**:563-564.
54. Chan YH, Venev SV, Zeldovich KB, Matthews CR: **Correlation of fitness landscapes from three orthologous TIM barrels originates from sequence and structure constraints.** *Nat Commun* 2017, **8**:1-12
- A mutational scanning approach is used to analyze the conservation of the fitness landscapes in three orthologous indole-3-glycerol phosphate synthases. It was found that their fitness landscapes are correlated, influenced by epistasis, and translocate sequence space due to the plasticity of the TIM-barrel fold.
55. Chan YH, Zeldovich KB, Matthews CR: **An allosteric pathway explains beneficial fitness in yeast for long-range mutations in an essential TIM barrel enzyme.** *Protein Sci* 2020, **29**:1911-1923.
56. Tzul FO, Vasilchuk D, Makhatadze GI: **Evidence for the principle of minimal frustration in the evolution of protein folding landscapes.** *Proc Natl Acad Sci U S A* 2017, **114**:E1627-E1632.
57. Candel AM, Romero-Romero ML, Gamiz-Arco G, Ibarra-Molero B, Sanchez-Ruiz JM: **Fast folding and slow unfolding of a resurrected Precambrian protein.** *Proc Natl Acad Sci U S A* 2017, **114**:E4122-E4123.
58. Gamiz-Arco G, Rizzo VA, Candel AM, Inglés-Prieto A, Romero-Romero ML, Gaucher EA, Gavira JA, Ibarra-Molero B, Sanchez-Ruiz JM: **Non-conservation of folding rates in the thioredoxin family reveals degradation of ancestral unassisted-folding.** *Biochem J* 2019, **476**:3631-3647.
59. Kukic P, Pustovalova Y, Camilloni C, Gianni S, Korzhnev DM, Vendruscolo M: **Structural characterization of the early events in the nucleation-condensation mechanism in a protein folding process.** *J Am Chem Soc* 2017, **139**:6899-6910.
60. Hart KM, Harms MJ, Schmidt BH, Ely C, Thornton JW: **Thermodynamic system drift in protein evolution.** *PLoS Biol* 2014, **12**:1001994.
61. Lim, Shion A, Marqusee S: **The burst-phase folding intermediate of ribonuclease H changes conformation over evolutionary history.** *Biopolymers* 2018, **109**:e23086.
62. Lim SA, Hart KM, Harms MJ, Marqusee S: **Evolutionary trend toward kinetic stability in the folding trajectory of RNases H.** *Proc Natl Acad Sci U S A* 2016, **113**:13045-13050.
63. Lim SA, Bolin ER, Marqusee S: **Tracing a protein's folding pathway over evolutionary time using ancestral sequence reconstruction and hydrogen exchange.** *eLife* 2018, **7**:e38369.
64. Yang J, Naik N, Patel JS, Wylie CS, Gu W, Huang J, Marty Ytreberg F, Naik MT, Weinreich DM, Rubenstein BM: **Predicting the viability of beta-lactamase: how folding and binding free**

104 Sequences and topology

- energies correlate with beta-lactamase fitness.** *PLoS One* 2020, **15**:e0233509.
65. Faber MS, Wrenbeck EE, Azouz LR, Steiner PJ, Whitehead TA: **Impact of in vivo protein folding probability on local fitness landscapes.** *Mol Biol Evol* 2019, **36**:2764-2777.
 66. Choi H-K, Min D, Kang H, Ju Shon M, Rah S-H, Chan Kim H, Jeong H, Choi H-J, Bowie JU, Yoon T-Y: **Watching helical membrane proteins fold reveals a common N-to-C-terminal folding pathway.** *Science* 2019, **366**:1150-1156.
 67. Isogai Y, Imamura H, Nakae S, Sumi T, Takahashi KI, Nakagawa T, Tsuneshige A, Shirai T: **Tracing whale myoglobin evolution by resurrecting ancient proteins.** *Sci Rep* 2018, **8** 16883.
 68. Korendovych IV, DeGrado WF: **De novo protein design, a retrospective.** *Q Rev Biophys* 2020, **53**:e3.
 69. Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, Kaufman K, Renfrew PD, Smith CA, Sheffler W *et al.*: **Rosetta3: an object-oriented software suite for the simulation and design of macromolecules.** *Methods Enzymol* 2011, **487**:545-574.
 70. Alford RF, Leaver-Fay A, Jeliakzov JR, O'Meara MJ, DiMaio FP, Park H, Shapovalov MV, Renfrew PD, Mulligan VK, Kappel K *et al.*: **The Rosetta all-atom energy function for macromolecular modeling and design.** *J Chem Theory Comput* 2017, **13**:3031-3048.
 71. Jacobs TM, Williams B, Williams T, Xu X, Eletsky A, Federizon JF, Szyperski T, Kuhlman B: **Design of structurally distinct proteins using strategies inspired by evolution.** *Science* 2016, **352**:687-690.
 72. Zhou J, Panaitiu AE, Grigoryan G: **A general-purpose protein design framework based on mining sequence-structure relationships in known protein structures.** *Proc Natl Acad Sci U S A* 2020, **117**:1059-1068.
 73. Koga R, Koga N: **Consistency principle for protein design.** *Biophys Physicobiology* 2019, **16**:304-309.
 74. Koga N, Tatsumi-Koga R, Liu G, Xiao R, Acton TB, Montellone GT, Baker D: **Principles for designing ideal protein structures.** *Nature* 2012, **491**:222-227.
 75. Goraj K, Renard A, Martial JA: **Synthesis, purification and initial structural characterization of octarellin, a de novo polypeptide modelled on the α/β -barrel proteins.** *Protein Eng Des Sel* 1990, **3**:259-266.
 76. Tanaka T, Hayashi M, Kimura H, Oobatake M, Nakamura H: **De novo design and creation of a stable artificial protein.** *Biophys Chem* 1994, **50**:47-61.
 77. Tanaka T, Kuroda Y, Kimura H, Kidokoro SI, Nakamura H: **Cooperative deformation of a de novo designed protein.** *Protein Eng Des Sel* 1994, **7**:969-976.
 78. Tanaka T, Kimura H, Hayashi M, Fujiyoshi Y, Fukuhara KI, Nakamura H: **Characteristics of a de novo designed protein.** *Protein Sci* 1994, **3**:419-427.
 79. Houbrechts A, Moreau B, Abagyan R, Mainfroid V, Pr aux G, Lamproye A, Poncin A, Goormaghtigh E, Ruysschaert JM, Martial JA *et al.*: **Second-generation octarellins: Two new de novo (β/α)8 polypeptides designed for investigating the influence of β -residue packing on the α/β -barrel structure stability.** *Protein Eng Des Sel* 1995, **8**:249-259.
 80. Beauregard M, Goraj K, Goffin V, Heremans K, Goormaghtigh E, Ruysschaert JM, Martial JA: **Spectroscopic investigation of structure in octarellin (a de novo protein designed to adopt the α/β -barrel packing).** *Protein Eng Des Sel* 1991, **4**:745-749.
 81. Offredi F, Dubail F, Kischel P, Sarinski K, Stern AS, Van de Weerd C, Hoch JC, Prosperi C, Fran ois JM, Mayo SL *et al.*: **De novo backbone and sequence design of an idealized α/β -barrel protein: evidence of stable tertiary structure.** *J Mol Biol* 2003, **325**:163-174.
 82. Figueroa M, Oliveira N, Lejeune A, Kaufmann KW, Dorr BM, Matagne A, Martial JA, Meiler J, Van de Weerd C: **Octarellin VI: using Rosetta to design a putative artificial (β/α)8 protein.** *PLoS One* 2013, **8**:e71858.
 83. Figueroa M, Sleutel M, Vandevenne M, Parvizi G, Attout S, Jacquin O, Vandenameele J, Fischer AW, Dambon C, Goormaghtigh E *et al.*: **The unexpected structure of the designed protein Octarellin V.1 forms a challenge for protein structure prediction tools.** *J Struct Biol* 2016, **195**:19-30.
 84. Nagarajan D, Deka G, Rao M: **Design of symmetric TIM barrel proteins from first principles.** *BMC Biochem* 2015, **16**:18.
 85. Huang PS, Feldmeier K, Parmeggiani F, Fernandez Velasco DA, Hocker B, Baker D: **De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy.** *Nat Chem Biol* 2016, **12**:29-34.
- This work reports the first successful *de novo* designed TIM barrel. The design approach comprises determination of geometric restrictions, backbone generation and iterative sequence design. Experimental characterization showed well-folded proteins and the intended topology for the construct sTIM11.
86. Romero-Romero S, Costas M, Silva D-A, Kordes S, Rojas-Ortega E, Tapia C, Guerra Y, Shanmugaratnam S, Rodr guez-Romero A, Baker D *et al.*: **Epistasis on the stability landscape of de novo TIM barrels explored by a modular design approach.** *bioRxiv* 2020 <http://dx.doi.org/10.1101/2020.09.29.319103>
- A modular design approach was used to create a family of stabilized sTIM11 variants by improving hydrophobic packing. Detailed analysis showed that unexplored regions of the stability landscape are accessed. This landscape is shaped by epistatic effects arising from improved hydrophobic clusters.
87. Caldwell SJ, Haydon IC, Piperidou N, Huang P-S, Bick MJ, S jstr m HS, Hilvert D, Baker D, Zeymer C: **Tight and specific lanthanide binding in a de novo TIM barrel with a large internal cavity designed by symmetric domain fusion.** *Proc Natl Acad Sci U S A* 2020, **117**:30362-30369.
 88. Wiese G, Shanmugaratnam S, H cker B: **Extension of a de novo TIM barrel with a rationally designed secondary structure element.** *bioRxiv* 2020 <http://dx.doi.org/10.1101/2020.10.16.342774>.
 89. Watters AL, Deka P, Corrent C, Callender D, Varani G, Sosnick T, Baker D: **The highly cooperative folding of small naturally occurring proteins is likely the result of natural selection.** *Cell* 2007, **128**:613-624.
 90. Basak S, Paul Nobrega R, Tavella D, Deveau LM, Koga N, Tatsumi-Koga R, Baker D, Massi F, Robert Matthews C: **Networks of electrostatic and hydrophobic interactions modulate the complex folding free energy surface of a designed $\beta\alpha$ protein.** *Proc Natl Acad Sci U S A* 2019, **116**:6806-6811.
- This work analyses the complex folding pathway of the *de novo* protein Di-III₁₄. Electrostatic and hydrophobic networks are identified as possible modulators and their contribution specified by mutational analysis. These findings have implications for future protein design strategies.

List of publications

- I **The stability landscape of de novo TIM barrels explored by a modular design approach**
S. Romero-Romero, M. Costas, D.-A. Silva Manzano, **S. Kordes**, E. Rojas-Ortega, C. Tapia, Y. Guerra, S. Shanmugaratnam, A. Rodriguez-Romero, D. Baker, B. Höcker, D. A. Fernandez-Velasco
Journal of Molecular Biology, 2021, 433(18)
- II **A newly introduced salt bridge cluster improves structural and biophysical properties of de novo TIM barrels**
S. Kordes*, S. Romero-Romero*, L. Lutz, B. Höcker
Protein Science, 2022, 31, pp. 513-527
- III **Extension of a de novo TIM barrel with rationally designed antiparallel coiled-coils**
S. Kordes, M. Flecks, L. Lutz, B. Höcker
Manuscript
- IV **Evolution, folding, and design of TIM barrels and related proteins**
S. Romero-Romero*, **S. Kordes***, F. Michel*, B. Höcker
Current Opinion in Structural Biology, 2021, 68, pp. 94-104
- V **Recombinant protein production and purification of SiiD, SiiE and SiiF - Components of the SPI4-encoded type I secretion system from Salmonella Typhimurium**
S. Klingl*, **S. Kordes***, B. Schmid, R. Gerlach, M. Hensel, Y. A. Muller
Protein Expression and Purification, 2020, 172(105632)
This publication is not part and did not contribute to this thesis.

*equal contribution

Acknowledgements

In the first place, I would like to express my great gratitude to my supervisor Prof. Dr. Birte Höcker, for giving me the opportunity to work on these fascinating projects. I also want to say thank you for all the guidance, encouragement and advice throughout this time.

I want to thank Sergio Romero-Romero for the awesome cooperation, the advices regarding especially, but not only, stability analysis. I think we were always a good team, and I will miss it to work with you on further projects.

Special thanks goes also to Renate Gelfert for supporting me during this time, having a great time in Bayreuth with wine and kickboxing.

I want to thank Bruce Lichtenstein for his support and discussions especially in the starting time of this work.

Special thanks goes also to Sabrina Wischt for helping me in the Lab and making the work always a funnier place.

I also want to thank Sooruban Shanmugaratnam for knowing everything in the Lab and helping a lot when it comes to crystallography topics.

Special thanks goes also to Michael Weyand for excellent advice and support in crystallographic topics.

I'd like to say thank you to all past and present people of the AG Höcker for all the help, advice, the fantastic time, great BBQ events and everything else. In random order: Pascal, Florian, Francisco, Julian, Lukas, Basti, Josef, Noelia, Florian, Merve, Guto, Abhishek, Julio, Horst. Thanks to all of you!

Last but not least I would like to thank my husband, Stefan Klingl, who went through all the ups and downs of my PhD with me. Thanks for driving all these years every weekend to me, giving me new strength for the next week and always motivating me again. Without you I wouldn't have made it so far! Finally, I want to thank my parents, Katrin and Christian, and my sister Julia for always having my back, supporting me and believing in me.

(Eidesstattliche) Versicherungen und Erklärungen

(§ 8 Satz 2 Nr. 3 PromO Fakultät)

Hiermit versichere ich eidesstattlich, dass ich die Arbeit selbstständig verfasst und keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt habe (vgl. Art. 64 Abs. 1 Satz 6 BayHSchG).

(§ 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. 8 PromO Fakultät)

Hiermit erkläre ich mein Einverständnis, dass bei Verdacht wissenschaftlichen Fehlverhaltens Ermittlungen durch universitätsinterne Organe der wissenschaftlichen Selbstkontrolle stattfinden können.

(Ort, Datum)

(Unterschrift)