

Thermodynamische Stabilisierung von Proteinen
durch *in vitro* Evolution und
biophysikalische Analyse ihrer molekularen Grundlagen

DISSERTATION
ZUR ERLANGUNG DES GRADES
-DOKTOR DER NATURWISSENSCHAFTEN-
DER FAKULTÄT FÜR BIOLOGIE, CHEMIE
UND GEOWISSENSCHAFTEN
DER UNIVERSITÄT BAYREUTH

vorgelegt von
Diplom-Biochemiker

Michael Wunderlich

Bayreuth 2006

Vollständiger Abdruck der von der Fakultät für Biologie, Chemie und Geowissenschaften der Universität Bayreuth zur Erlangung des akademischen Grades –Doktor der Naturwissenschaften- genehmigten Dissertation.

Diese Arbeit wurde von April 2002 bis April 2006 am Lehrstuhl für Biochemie der Universität Bayreuth unter der Leitung von Prof. Dr. Franz X. Schmid angefertigt.

Promotionsgesuch eingereicht am: 28.06.2006

Tag des wissenschaftlichen Kolloquiums: 22.11.2006

Prüfungsausschuß:

Prof. Dr. Franz Xaver Schmid (Erster Gutachter)

Prof. Dr. Matthias Ullmann (Zweiter Gutachter)

Prof. Dr. Wolfgang Schumann

Prof. Dr. Carlo Unverzagt (Vorsitzender)

Meinen Eltern

Kurzfassung der Arbeit	1
Abstract	3
1. Einleitung	5
1.1 Stabilität von Proteinen	5
1.2 Strukturelle Grundlagen für die Stabilität von Proteinen.....	6
1.3 Methoden zur Stabilisierung von Proteinen.....	7
1.4 Das Selektionssystem <i>Proside</i>	9
1.5 Modellproteine zur Untersuchung von Faltung und Stabilität von Proteinen.....	10
1.5.1 Das Kälteschockprotein CspB aus <i>Bacillus subtilis</i>	10
1.5.2 Die Domäne $\beta 1$ des Streptokokkenproteins G (G $\beta 1$)	12
1.6 Problemstellung.....	13
2. Zusammenfassung und Diskussion der Ergebnisse	15
2.1 Evolutive Stabilisierung des Kälteschockproteins CspB	15
2.1.1 Ortsgerichtete Optimierung von oberflächenexponierten Positionen	15
2.1.2 Ortsunspezifische Optimierung von <i>Bs</i> -CspB.....	18
2.1.3 Abhängigkeit der Effekte eines Ladungsaustausches von der Umgebung	21
2.2 Korrelation zwischen Proteininstabilität und Dipolmoment	24
2.3 <i>In vitro</i> Evolution im Vergleich zu rechnergestütztem <i>protein design</i>	27
2.3.1 <i>In vitro</i> Evolution der <i>boundary positions</i> von G $\beta 1$	27
2.3.2 Vergleich von rechnergestützter und evolutiver Stabilisierung	29
2.3.3 Strukturelle und thermodynamische Analyse der Unterschiede zwischen <i>computational design</i> und <i>in vitro</i> Evolution.....	31
2.4 Experimentelle Optimierung von G $\beta 1$	35
2.4.1 Identifizierung von aussichtsreichen Positionen durch fehlerhafte PCR.....	35
2.4.2 Ortsgerichtete Sättigungsmutagenese der gefundenen Positionen.....	36
2.4.3 Analyse der Ursachen für die erzielte Stabilisierung von G $\beta 1$	39
2.5 Fazit der Arbeit.....	45
3. Abkürzungen	47
4. Literaturverzeichnis.....	49
5. Publikationsliste	55
6. Darstellung des Eigenanteils	56
7. Teilarbeiten	57
7.1 Teilarbeit A	57
7.2 Teilarbeit B.....	59
7.3 Teilarbeit C.....	61
7.4 Teilarbeit D	63
7.5 Teilarbeit E.....	83
7.6 Teilarbeit F	85

Kurzfassung der Arbeit

Proteine sind zentrale Bausteine des Lebens und ihre dreidimensionale Struktur ist für die Funktion von entscheidender Bedeutung. Deswegen ist das Verständnis der konformationellen Stabilität von Proteinen von großem wissenschaftlichen, aber auch kommerziellem Interesse. Aufgrund der Komplexität von Proteinen stößt man beim *rational design* zur Steigerung der Proteinstabilität sehr schnell an die Grenzen der Methoden. Eine Alternative stellt die *in vitro* Evolution im Labor dar, denn sie bietet die Möglichkeit, eine Vielzahl von Proteinvarianten parallel zu untersuchen.

In der vorliegenden Arbeit wurde das Selektionssystem *Proside* zur Stabilisierung von Proteinen genutzt. *Proside* basiert auf der *phage display* Technologie und verknüpft die thermodynamische Stabilität von Proteinen mit der Infektiosität von filamentösen Phagen. Dabei dient die negative Korrelation zwischen der thermodynamischen Stabilität und der Sensitivität von Proteinen gegenüber Proteasen als Selektionsprinzip.

Die bakteriellen Kälteschockproteine (Csp) werden sowohl von experimenteller, als auch theoretischer Seite als Modellproteine zur Untersuchung der Stabilität von Proteinen genutzt. Es ist bekannt, dass bei dieser Proteinfamilie elektrostatische Wechselwirkungen auf der Oberfläche von sehr großer Bedeutung für ihre Stabilität sind. Deswegen sollte versucht werden, alternative Ladungsnetzwerke auf der Oberfläche des Kälteschockproteins *Bs-CspB* aus *Bacillus subtilis* unter Verwendung von *in vitro* Evolution zu etablieren. Zu diesem Zweck wurden neun Reste anhand von Sequenz- und Strukturvergleichen mit homologen Proteinen ausgewählt und diese Positionen partiell randomisiert. In einem zweiten Ansatz wurden unter Verwendung von fehlerhafter PCR und unter Ausnutzung der natürlichen Mutageneserate von *Escherichia coli* zufällig Mutationen in das Gen von *Bs-CspB* eingeführt. Nach erfolgter *in vitro* Evolution der jeweiligen Bibliotheken konnten an den oberflächenexponierten Positionen 1, 3, 46, 48, 65 und 66 stabilisierende Mutationen identifiziert werden. Die Beiträge dieser Mutationen zur thermodynamischen Stabilität von *Bs-CspB* wurden einzeln und in unterschiedlichen Kombinationen analysiert. Die beste Kombination (M1R/E3K/K65I) war, mit einem Schmelzpunkt von 83,7 °C im Vergleich zu 53,8 °C vom Wildtypprotein, sogar stabiler war als das hyperthermophile Homologe *Tm-Csp* aus *Thermotoga maritima*.

Einen beachtlichen Beitrag zu dieser deutlichen Stabilisierung liefern dabei verbesserte Coulomb'sche Wechselwirkungen. Die Beiträge sind dabei jedoch sehr stark von der Umgebung abhängig. Ein gutes Beispiel stellt die Mutation E3R dar, welche im Wildtypprotein stark stabilisierend wirkt ($\Delta\Delta G_D = 11,1 \text{ kJ mol}^{-1}$), in der Variante A46K/S48R/E66L hingegen destabilisierend ($\Delta\Delta G_D = -4,0 \text{ kJ mol}^{-1}$). Vergleiche der experimentellen Daten mit vorhandenen theoretischen Betrachtungen offenbarten die Schwierigkeiten, Coulomb'sche Wechselwirkungen auf der Proteinoberfläche korrekt zu berechnen und deren Einfluss auf die thermodynamische Proteinstabilität zu bestimmen.

Ein theoretischer Ansatz, welcher unter Verwendung eines quasi-elektrischen Dipolmomentes die Einflüsse von Veränderungen der Ladungsverteilung auf die Stabilität vorhersagte, konnte experimentell nicht verifiziert werden. Ebenso konnten keine Korrelationen zwischen Veränderungen der Stabilität und der Nettoladung des Proteins festgestellt werden und auch keine Korrelation mit der Bildung von Ionenpaaren hergestellt werden. Die korrekte Bestimmung von elektrostatischen Wechselwirkungen auf der Oberfläche von Proteinen und die Berechnung ihrer Beiträge zur freien Enthalpie der Entfaltung von Proteinen bleiben eine Herausforderung.

Ebenso wie die Kälteschockproteine stellt auch die $\beta 1$ Domäne des Streptokokkenproteins G ($G\beta 1$) ein Modellprotein zur Untersuchung von thermodynamischer Stabilität dar. In der Gruppe von Stephen Mayo wurde dieses Protein unter Verwendung von Sequenzoptimierung durch *computational design* untersucht. Dabei zeigte sich, dass die vier teilweise exponierten Reste 16, 18, 25 und 29 ein sehr großes Stabilisierungspotential besitzen. Eine *in vitro* Evolution dieser vier Positionen durch das Selektionssystem *Proside* bot somit die Möglichkeit, die Leistungsfähigkeit der beiden Methoden direkt zu vergleichen.

Die Selektionen lieferten eine Vielzahl von deutlich stabilisierten Proteinvarianten. Für die stabilste Variante mit den Mutationen T16I/T18I/T25E/V29F war der Mittelpunkt des thermischen Überganges um 24,7 °C erhöht, was einer Stabilisierung um 14,9 kJ mol⁻¹ entsprach. Die beste Variante aus den theoretischen Berechnungen in der Gruppe von Stephen Mayo lag von allen untersuchten Varianten auf dem dritten Platz. Sieben der zehn berechneten Varianten, welche in den Kalkulationen alle annähernd gleich stabil waren, enthielten einzeln oder in Kombination die Aminosäurereste Leu18 und Lys29. Diese beiden Reste erwiesen sich bei der experimentellen Charakterisierung als sehr ungünstig und wirkten jeweils um etwa 4,5 kJ mol⁻¹ destabilisierend, relativ zu den besten Resten an diesen Positionen.

Um die Ursachen für diese Unterschiede zu ergründen, wurde von der stabilsten Variante aus der Selektion und einer laut Berechnungen nahezu gleichstabilen Variante die dreidimensionale Struktur mittels Röntgenkristallographie bestimmt. Die Rückgratstruktur dieser Variante war im Vergleich zum Wildtypprotein praktisch unverändert. Zusätzlich wurden die in den Varianten enthaltenen Mutationen einzeln und in unterschiedlicher Kombination thermodynamisch charakterisiert. Dabei erwiesen sich die Effekte der Mutationen an den vier untersuchten Positionen als additiv und können trotz ihres geringen Abstandes als unabhängig betrachtet werden. Außerdem zeigten sich bei der Analyse die Schwierigkeiten, strukturelles Verhalten in experimentelle energetische Terme zu übersetzen, was der Grund für die unterschiedlichen Ergebnisse bei der Stabilisierung durch *in vitro* Evolution und *computational design* sein dürfte.

In der ersten Stabilisierung von $G\beta 1$ wurden bei der Auswahl von interessanten Positionen die Ergebnisse aus dem *computational design* genutzt. In einem zweiten Ansatz sollten für beide Schritte auf dem Weg zur Stabilisierung, der Auswahl von Positionen und dem Finden der besten Aminosäurereste für diese Positionen, ein experimentelles Herangehen genutzt werden. Durch Selektion von Bibliotheken mit zufälligen Mutationen, erstellt mittels fehlerhafter PCR, konnten viel versprechende Positionen identifiziert werden. Diese wurden dann einer Sättigungsmutagenese mit anschließender Selektion unterworfen und die gefundenen Mutationen der beiden stabilsten Varianten wurden abschließend manuell kombiniert. Die erhaltene Variante E15V/T16L/T18I/N37L besaß einen um 35,1 °C erhöhten Übergangsmittelpunkt, was einer Stabilisierung um 28,5 kJ mol⁻¹ entsprach. Bei der Analyse der Ursachen zeigte sich, dass in diesem Fall starke hydrophobe Wechselwirkungen zwischen den durch diese Mutationen eingeführten Seitenketten wirken. Die Wechselwirkungen werden durch Tyr33 vermittelt und tragen einen großen Teil zur Stabilisierung bei. Außerdem war für die selektierten Varianten die Kooperativität der Entfaltung deutlich erhöht, was ebenfalls zur erhöhten Stabilität beiträgt. Dies zeigt, dass zur Stabilisierung von Proteinen vielfältige Wege beschritten werden können.

Abstract

Proteins are central building blocks of life and their three dimensional structure is of utmost importance for the function. Therefore the understanding of conformational protein stability is of high scientific as well as commercial interest. Due to the complexity of proteins rational engineering of protein stability rapidly reaches its limit. An alternative is *in vitro* evolution in the laboratory, because this method offers the chance to test large libraries of protein variants in parallel.

In this work the selection system *Proside* was used to stabilize proteins. *Proside* is based on the phage display technology linking the thermodynamic stability of proteins with the infectivity of filamentous phages. Therefore, the negative correlation between the thermodynamic stability and the sensitivity of proteins against the cleavage by proteases is used for the selection.

The bacterial cold shock proteins (Csp) were investigated by both experimentalists and theoreticians as a model system for protein stability. It is known that electrostatic interactions at the surface of the cold shock proteins are important for their stability. The first aim was to establish alternative charge networks at the surface of *Bs*-CspB from *Bacillus subtilis* to stabilize the protein. Nine residues were chosen by sequence and structure alignments of homologous proteins and these positions were partially randomized. In a second approach, random mutations were introduced into the gene of *Bs*-CspB by error-prone PCR or by using the spontaneous mutation rate of *Escherichia coli*. By *in vitro* evolution of the corresponding phage libraries stabilizing mutations were found at the surface positions 1, 3, 46, 48, 65, and 66. The contributions of these mutations to the thermodynamic stability were characterized by analyzing them individually and in combination. The best combination (M1R/E3K/K65I) increased the midpoint of thermal unfolding of *Bs*-CspB from 53.8 °C to 83.7 °C, which implies that the corresponding mutant more stable than the homologous cold shock protein from the hyperthermophilic bacterium *Thermotoga maritima*.

A significant contribution to this stabilization is provided by improved coulombic interactions. But the effects of most mutations are strongly context dependent. A good example is provided by the E3R mutation. It is strongly stabilizing ($\Delta\Delta G_D = 11.1 \text{ kJ mol}^{-1}$) in the wild-type protein, but destabilizing ($\Delta\Delta G_D = -4.0 \text{ kJ mol}^{-1}$) in the A46K/S48R/E66L variant. Comparisons of the experimental data and existing theoretical investigations show the difficulty to calculate the coulombic interactions at the surface of proteins and to determine their influence on the thermodynamic stability of proteins.

The quasi-electric dipole moment, which is used in a theoretical approach to predict the influence of a change in charge distribution to protein stability, could not be validated by experiment. The stabilizations by charge mutations did not correlate well with the corresponding changes in the protein net charge, and they could also not be ascribed to the formation of ion pairs. Therefore it remains a challenging task to correctly predict electrostatic interactions at the surface of proteins and to calculate their contribution to the Gibbs free energy of unfolding.

Similar to the cold shock protein, the $\beta 1$ domain of the streptococcal protein G (G $\beta 1$) is a widely used model protein for investigations of the thermodynamic stability of proteins. In the group of Stephen Mayo this protein was used for sequence optimization by computational design. It could be shown that four partially exposed (“boundary”) positions 16, 18, 25, and

29 carry a large potential for stabilization. The *in-vitro* evolution of the same four positions by the selection system *Proside* offered the possibility to compare the efficiency of both methods directly.

Many G β 1 variants with strongly increased thermal stabilities were found in the selections. For the most stable variant T16I/T18I/T25E/V29F the midpoint of the thermal transition was increased by 24.7 °C, which corresponds to a stabilization by 14.9 kJ mol⁻¹. The best variant from the calculations by Stephen Mayo ranked third within the panel of the selected variants. Seven of the ten nearly isoenergetic variants from the calculations contained the residues Leu18 and Lys29 individually or in combination. These two residues turned out to be less favorable in the experimental characterization, and they are both destabilizing by about 4.5 kJ mol⁻¹ in comparison to the best replacements at these positions.

To elucidate the reasons for these differences, the threedimensional structures of the most stable selected variant and of a nearly isoenergetic variant from the calculations were determined by X-ray crystallography. The peptide backbone of these variants was almost unchanged in comparison to the wildtype protein. Additionally, the mutations containing in these variants were thermodynamically characterized, individually and in different combinations. The effects of the mutations at the four positions are largely additive and in spite of the short distance between the positions they could be considered as independent. In addition the structure/stability comparisons emphasized how difficult it is to translate structural behaviour into experimental energetic terms. This is probably the reason for the differences between the results of *in vitro* evolution and computational design.

In the first stabilization of G β 1 for identification of interesting positions the results of the computational design were used. In a second approach the experimental *in-vitro* selection system *Proside* was used for both identifying promising sites (in libraries of G β 1 mutants that were created by error-prone PCR) and obtaining the best residues (in libraries created by saturation mutagenesis at these sites). A manual combination of the two most stable selected variants resulted in the hyperstable fourfold mutant E15V/T16L/T18I/N37L, which showed an increase in the midpoint of the thermal transition of 35.1 °C and an increase in $\Delta\Delta G_D$ of 28.5 kJ mol⁻¹. The analysis of the reasons showed, that strong hydrophobic interactions are acting between the aminoacids introduced by these mutations. These interactions were mediated by Tyr33 and were partially responsible for the observed stabilization. Additionally, the cooperativity of unfolding for the selected variants is strongly increased, which also contributes to the increased stability. This shows, that there are many different approaches to stabilize a protein.

1. Einleitung

1.1 Stabilität von Proteinen

Die „Stabilität“ von Proteinen ist sowohl von großem wissenschaftlichen, als auch kommerziellen Interesse. Dabei besitzt der Begriff „Stabilität“ durchaus sehr unterschiedliche Bedeutung.

Aus thermodynamischer Sichtweise stellt die Stabilität von Proteinen den Unterschied der freien Enthalpien, bezeichnet als ΔG_D , zwischen dem gefalteten Zustand N und dem entfalteten Zustand U dar. Voraussetzung für eine solche Betrachtung ist, dass der gefaltete Zustand N und der entfalteten Zustand U in einem reversiblen Gleichgewicht vorliegen (Gleichung 1) (Privalov, 1979).



Mit Hilfe der Gleichgewichtskonstanten K_D für diese Reaktion lässt sich die Stabilität von Proteinen quantifizieren (Gleichung 2).

$$\Delta G_D = -RT \cdot \ln K_D = -RT \cdot \ln \frac{[U]}{[N]} \quad (2)$$

Die Gleichgewichtskonstante kann unter physiologischen Bedingungen im Normalfall nicht bestimmt werden, da der Anteil an entfaltetem Protein in diesem Fall äußerst gering ist. Aber das Gleichgewicht lässt sich durch die Änderung von physikalischen Parametern (z. B. Temperatur und Druck) oder durch Zugabe von chemischen Denaturierungsmitteln beeinflussen und somit der Anteil des entfalteten Zustands U erhöhen. Der Verlauf der Verschiebung kann für sehr viele Proteine mit Hilfe von Fluoreszenz-, Circular dichroismus- oder Absorptionsspektroskopie verfolgt werden (Schmid, 2005).

Bei der Entfaltung durch chemische Agenzien kann die Änderung der Stabilität ΔG_D in Abhängigkeit von der Denaturierungsmittelkonzentration [D] meist sehr gut durch ein lineares Modell beschrieben werden (Gleichung 3) (Tanford, 1970).

$$\Delta G_D = \Delta G_D(H_2O) - m \cdot [D] \quad (3)$$

Durch die Annahme der linearen Abhängigkeit lässt sich die Stabilität von Proteinen in Abwesenheit des Denaturierungsmittels, bezeichnet als $\Delta G_D(H_2O)$, berechnen.

Die Entfaltung von Proteinen durch Erhöhung der Temperatur wird durch die Gibbs-Helmholtz-Gleichung (Gleichung 4) beschrieben. Die freie Enthalpie ΔG_D wird dabei in Beziehung zur Enthalpie ΔH_D und Entropie ΔS_D gesetzt.

$$\Delta G_D = \Delta H_D - T \cdot \Delta S_D \quad (4)$$

Die Parameter ΔH_D und ΔS_D zeigen dabei eine Temperaturabhängigkeit, welche durch die Änderung der Wärmekapazität Δc_p beschrieben werden kann (Gleichung 5).

$$\Delta c_p = \frac{\delta \Delta H_D}{\delta T} = T \cdot \frac{\delta \Delta S_D}{\delta T} \quad (5)$$

Es konnte gezeigt werden, dass Δc_p in erster Näherung linear von der Anzahl der Aminosäurereste eines Proteins abhängt (Myers *et al.*, 1995) und dass Δc_p als proportional zur Änderung der zugänglichen Oberfläche während der Entfaltung beschrieben werden kann (Murphy und Freire, 1992).

Betrachtet man die thermodynamische Stabilität als Funktion der Temperatur $\Delta G_D(T)$ so erhält man einen parabolischen Verlauf und $\Delta G_D(T)$ weist zwei Nullstellen [$\Delta G_D(T) = 0$] auf. Es tritt also sowohl bei hohen als auch bei niedrigen Temperaturen Entfaltung auf (Privalov, 1990).

Die freie Reaktionsenthalpie ΔG_D ist unter physiologischen Bedingungen für Proteine meist sehr niedrig ($\Delta G_D \leq 60 \text{ kJ mol}^{-1}$) (Jaenicke, 1996). Eine Ursache für die geringe Stabilität von Proteinen liegt in der Evolution begründet. Die meisten der zufällig erworbenen Mutationen in einem Protein wirken destabilisierend und können über die Generationen akkumuliert werden, solange die Funktionalität durch die schwindende Stabilität nicht beeinflusst wird. Die Stabilität von natürlich vorkommenden Proteinen liegt somit nur knapp über einem funktionsbedingten Grenzwert (Wintrode und Arnold, 2000).

Für die meisten Proteine ist eine thermodynamische Analyse der Stabilität nicht möglich, da im entfalteten Zustand oft irreversible Prozesse (wie Aggregation) geschehen. In diesem Fall kann die Stabilität durch kinetische Experimente untersucht werden. So wird die beobachtete Entfaltungsrates bei einer bestimmten Temperatur oder die Temperatur, bei der das Protein in einer definierten Zeit zur Hälfte inaktiviert wird, für Vergleiche herangezogen.

1.2 Strukturelle Grundlagen für die Stabilität von Proteinen

In Proteinen treten eine Vielzahl von stabilisierenden und destabilisierenden Wechselwirkungen auf, die sich zum Großteil gegenseitig aufheben. Daher kann die Stabilität von Proteinen durch kleine Veränderungen erheblich beeinflusst werden.

Wasserlösliche Proteine besitzen eine kompakte Struktur bei der die hydrophoben Seitenketten der Aminosäurereste im Inneren des Proteins verborgen sind und die polaren Reste exponiert auf der Oberfläche auftreten. Der so genannte hydrophobe Effekt wurde schon früh als die bestimmende Triebkraft für die Faltung und Stabilität von Proteinen erkannt (Tanford, 1978). Dabei bewirkt jede zusätzliche Methylgruppe, die im Inneren des Proteins verborgen ist, eine durchschnittliche Stabilisierung von $5,3 \text{ kJ mol}^{-1}$ (Pace, 1992).

Als Ursache für den hydrophoben Effekt von Proteinen können zwei Gründe genannt werden. Zum einen können hydrophobe Seitenketten auf der Oberfläche eines Proteins mit dem

natürlichen Lösungsmittel Wasser keine Wechselwirkung eingehen. Dies führt zu einem sehr ungünstigen entropischen Beitrag, da die Lösungsmittelmoleküle in ihren Freiheitsgraden eingeschränkt sind. Zum anderen sind hydrophobe Seitenketten im Inneren des Proteins in der Lage, enthalpisch günstige van-der-Waals-Wechselwirkungen einzugehen (Privalov und Makhatadze, 1993; Makhatadze und Privalov, 1993; Eisenberg und McLachlan, 1986).

Einen weiteren wichtigen Faktor für die Stabilität von Proteinen stellen Wasserstoffbrückenbindungen dar. Diese werden nicht nur intramolekular, z.B. durch das Peptidrückgrat, gebildet, sondern auch mit dem umgebenden wässrigen Medium. Dabei beinhaltet jede zusätzliche Wasserstoffbrückenbindung einen durchschnittlichen Energiegewinn von $2,1 \text{ kJ mol}^{-1}$ (Vogt *et al.*, 1997). Dieser Wert wird größer, wenn die beteiligten Aminosäuren im Inneren des Proteins verborgen sind. Wasserstoffbrückenbindungen sind außerdem durch ihre hohe Richtungs- und Abstandsspezifität verantwortlich für die Ausbildung der Sekundärstrukturelemente von Proteinen (Yang und Honig, 1995a; Yang und Honig, 1995b; Yang *et al.*, 1996).

Über die Bedeutung von elektrostatischen Wechselwirkungen für die thermodynamische Stabilität war man sich lange Zeit im Unklaren. Wegen der hohen Dielektrizitätskonstanten ϵ von Wasser schienen Coulomb'sche Wechselwirkung auf der Proteinoberfläche von untergeordneter Bedeutung. Im Proteininneren ist die Dielektrizitätskonstante niedrig, entsprechend sind Ladungswechselwirkungen dort stark, aber vermutlich nicht stark genug, um den hohen Aufwand der Desolvatation der geladenen Aminosäureseitenketten auszugleichen (Hendsch und Tidor, 1999). Das Interesse für die Bedeutung von elektrostatischen Wechselwirkungen wurde erst wieder durch die Erkenntnis geweckt, dass Proteine aus thermophilen Organismen meist mehr geladenen Seitengruppen auf ihrer Oberfläche besitzen als entsprechende Homologe aus mesophilen Organismen (Karshikoff und Ladenstein, 2001). Eine Reihe von theoretischen Studien legen den Schluss nahe, dass der Effekt von elektrostatischer Interaktion bei höheren Temperaturen stärker ausgeprägt ist (Elcock, 1998; de Bakker *et al.*, 1999).

Als weiterer wichtiger Parameter ist die Änderung der Kettenentropie bei der Faltung zu nennen. Die durch die Faltung der Proteine verursachte Abnahme der Kettenentropie stellt in jedem Fall den größten ungünstigen Faktor dar, der durch entsprechende andere, günstige Wechselwirkungen ausgeglichen werden muss.

1.3 Methoden zur Stabilisierung von Proteinen

Zur Erhöhung der thermodynamischen Stabilität von Proteinen bietet sich somit eine Vielzahl von Ansatzpunkten: Erhöhung der Anzahl von Wasserstoffbrückenbindungen, zusätzliche oder verbesserte elektrostatische Wechselwirkungen, stärkere hydrophobe Wechselwirkungen und verbesserte Packungsdichte, Erhöhung des α -helikalen Anteiles und Stabilisierung von α -Helices, vermehrter Einsatz von Prolin und β -verzweigten Aminosäuren in Loop-Strukturen, entropische Stabilisierung durch Versteifung des Proteinrückgrats oder die Einführung von Disulfidbrücken.

Die Vielzahl dieser Möglichkeiten zeigt die Probleme der Stabilisierung durch *rational design* auf. So können durch eingehende Analyse der dreidimensionalen Struktur manchmal aussichtsreiche Positionen identifiziert werden, entsprechende Vorhersagen müssen aber durch ortsgerichtete Mutagenese einzeln verifiziert werden.

Ähnliches gilt für Sequenzvergleiche von homologen thermophilen und mesophilen Proteinen (Fukuchi und Nishikawa, 2001; Chakravarty und Varadarajan, 2002). Diese liefern ebenfalls eine große Auswahl von viel versprechenden Positionen. Jedoch werden im Laufe der Evolution eine Vielzahl von neutralen Mutationen akkumuliert. Dies erschwert die Identifikation der entscheidenden Reste für die Stabilität erheblich (Wintrode und Arnold, 2000).

Ein weiterer Ansatz ist das *consensus design* (Steipe *et al.*, 1994). Hierbei geht man von der Annahme aus, dass konservierte Reste, abgeleitet aus Sequenzvergleichen von Proteinfamilien, mehr zur Stabilität beitragen als nicht konservierte. Entsprechend sollte der Ersatz von nicht konservierten Resten durch konservierte die Stabilität eines Proteins steigern. Allen diesen rationalen Ansätzen ist gemein, dass ein sehr großer analytischer Arbeitsaufwand bis zur Identifizierung einer einzelnen stabilisierenden Mutation nötig ist.

Durch die Vielzahl von hoch aufgelösten Proteinstrukturen und thermodynamischen Analysen konnten Computeralgorithmen geschaffen werden, die teilweise erfolgreich den Einfluss von Aminosäureaustauschen vorhersagen (Korkegian *et al.*, 2005; Dantas *et al.*, 2003; Bolon *et al.*, 2005). Besonders schwierig gestaltet sich dabei die Berechnung von Einflüssen, die sich aus Veränderungen in der Rückgratstruktur, im entfalteten Zustand oder aus entropischen Änderungen ergeben (Nauli *et al.*, 2001; Kuhlman *et al.*, 2003; Desjarlais und Handel, 1999).

Bei evolutiven Methoden wird nicht versucht, die Effekte einzelner Aminosäureaustausche vorherzusagen. Stattdessen wird zunächst eine große Anzahl verschiedener Varianten eines Proteins erzeugt und durch geeignete Selektionsmethoden Proteinvarianten mit gewünschten Eigenschaften angereichert. Alle evolutiven Methoden haben dabei gemeinsam, dass für die Amplifikation der gewünschten Proteinvarianten eine Verknüpfung von Geno- und Phänotyp nötig ist, d.h. das Protein muss mit seiner genetischen Information verbunden sein. Wenn die gewünschte Funktion eines Proteins mit dem Wachstum oder dem Überleben einer Zelle verknüpft werden kann, so ist eine Selektion *in vivo* möglich (Pelletier *et al.*, 1999; Orenica *et al.*, 2001; Hecky und Müller, 2005). Anderenfalls kann die Selektion *in vitro* erfolgen, hierbei kommen so genannte Displaymethoden zum Einsatz (Leemhuis *et al.*, 2005). Die am häufigsten eingesetzte Technik ist das Phagendisplay (Smith, 1985; Fernandez-Gacio *et al.*, 2003), bei der das zu untersuchende Protein mit einem Hüllprotein von filamentösen Phagen fusioniert und so auf deren Oberfläche präsentiert wird. Eine vergleichbare Methode ist das Zelloberflächendisplay, bei dem das Protein durch Fusion mit Membranproteinen auf der Oberfläche ganzer Zellen präsentiert wird. Im Gegensatz dazu stellen das Ribosomendisplay (Hanes und Plückthun, 1997) und das mRNA-Display (Roberts und Szostak, 1997; Nemoto *et al.*, 1997) reine *in vitro* Methoden dar. Die Verknüpfung von Geno- und Phänotyp wird dabei entweder durch einen stabilisierten Komplex mit dem Ribosom oder über eine kovalente Bindung vermittelt durch Puromycin erzeugt. In neuesten Methoden werden auch Emulsionströpfchen analog zu Kompartimenten in Zellen genutzt, um eine spezifische Verknüpfung von Geno- und Phänotyp zu erzielen (Leemhuis *et al.*, 2005). Dabei enthält zum Beispiel jedes „Kompartiment“ ein DNA-Fragment, welches sich aus der Sequenz des zu untersuchenden Proteins und der von Streptavidin zusammensetzt. Das DNA-Fragment ist gleichzeitig endständig mit Biotin modifiziert. Nach erfolgreicher *in vitro* Transkription und Translation bindet das Fusionsprotein mit Hilfe des Streptavidinteils an Biotin und somit an seine kodierende genetische Sequenz.

1.4 Das Selektionssystem *Proside*

Das von Volker Sieber (1998) entwickelte Selektionssystem *Proside* (*protein stability increased by directed evolution*) stellt eine generell anwendbare Methode zur Stabilisierung von Proteinen dar. Das System ist unabhängig von spezifischen Eigenschaften des zu untersuchenden Proteins wie Bindungsaffinität oder enzymatische Aktivität. *Proside* basiert auf dem Phagendisplay und verknüpft die erhöhte Proteaseresistenz von thermodynamisch stabilisierten Proteinvarianten mit der Infektiosität von filamentösen Phagen (Abbildung 1).

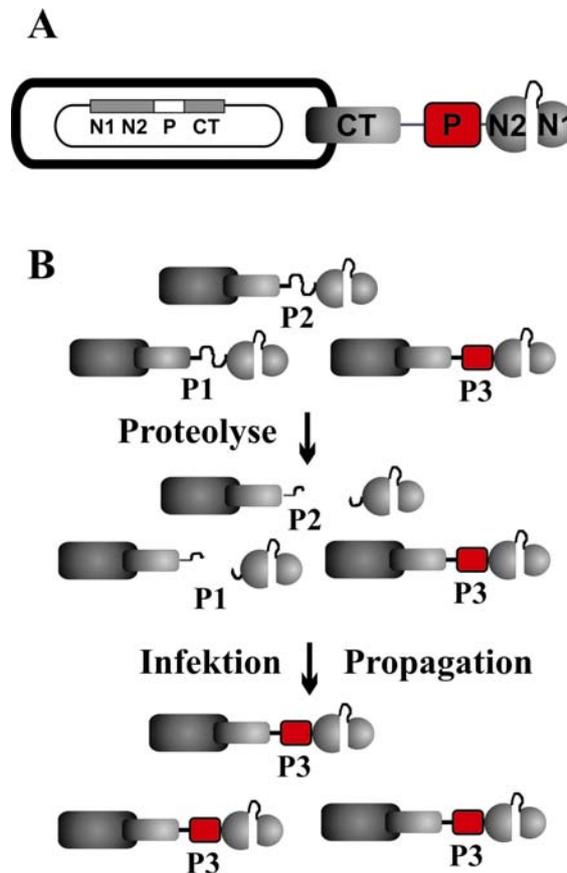


Abbildung 1. (A) Schematische Darstellung des verwendeten filamentösen Phagen fd mit Gastprotein zwischen der C-terminalen (CT) und den N-terminalen Domäne (N1 und N2) des Gen-III-Proteins. (B) Schematische Darstellung der Durchführung einer *Proside*-Selektion. Ein Phage besitzt als Insert verschiedene Varianten eines Proteins (P1, P2, P3). Durch proteolytische Spaltung werden bei den weniger stabilen Varianten (P2, P3) die N-terminalen Domänen abgetrennt und die Phagen verlieren ihre Infektiosität. Phagen mit stabilerem Insert (P1) bleiben infektiös und können durch Propagation angereichert werden.

Die Resistenz von Proteinen gegenüber der Spaltung durch Proteasen ist verknüpft mit der thermodynamischen Stabilität eines Proteins (Parsell und Sauer, 1989; Park und Marqusee, 2004; Park und Marqusee, 2005). Proteasen spalten nur die entfaltete Proteinkette und nicht den gefalteten, nativen Zustand. Chymotrypsin zum Beispiel spaltet Proteine spezifisch an aromatischen Seitenketten. Diese sind im gefalteten Zustand meist im Inneren des Proteins verborgen und für die Protease nicht zugänglich.

Der filamentöse Phage fd infiziert *Escherichia coli* mit Hilfe seines Gen-3-Proteins (G3P), das in drei bis fünf Kopien an der Spitze der Phagenhülle auftritt. G3P besteht aus drei Domänen

(Abbildung 1A). Die C-terminale Domäne CT verankert G3P in der Hülle des Phagen (Stengele *et al.*, 1990). Die beiden N-terminalen Domänen N1 und N2 sind verantwortlich für die Infektiosität des Phagen (Lubkowski *et al.*, 1998; Holliger *et al.*, 1999). Die Domäne N2 initiiert den primären Kontakt mit dem F-Pilus der Bakterienzelle (Deng und Perham, 2002) und N1 ermöglicht nach Wechselwirkung mit dem Corezeptor TolA (Lubkowski *et al.*, 1999; Karlsson *et al.*, 2003) das Einschleusen der Phagen-DNA in die Zelle. Die Domänen sind durch glycinreiche Linker von 18 bzw. 39 Aminosäuren verbunden und zum Erhalt der Infektiosität müssen diese kovalenten Verbindungen bestehen bleiben (Stengele *et al.*, 1990).

Das Selektionssystem *Proside* nützt aus, dass diese Verknüpfung der Domänen um ein Gastprotein erweitert werden kann. Bei *Proside* ist zwischen die N2 und die CT Domäne ein Repertoire von Sequenzen eingefügt, welche Varianten des zu stabilisierenden Proteins kodieren (Abbildung 1A). Die daraus resultierende Bibliothek von Phagen wird einer *in vitro* Proteolyse unterworfen. In Gegenwart einer Protease bleibt die kovalente Verbindung der beiden Domänen nur bestehen, wenn das Gastprotein nicht gespalten wird. Varianten mit erhöhter Proteaseresistenz und somit höherer thermodynamischer Stabilität können so durch wiederholte Zyklen von *in-vitro* Proteolyse, Infektion von *E. coli* und Phagenpropagation angereichert werden (Abbildung 1B). Der proteolytische Selektionsschritt wird *in vitro* durchgeführt und kann deshalb in einem weiten Rahmen variiert werden. Er ist nur durch die Stabilität des Phagens selbst und der Protease beschränkt. Die Funktionsweise des Systems konnte durch die erfolgreiche Stabilisierung von RNase T1 gezeigt werden (Sieber *et al.*, 1998) und kann auch zur Stabilisierung von G3P selbst eingesetzt werden (Martin und Schmid, 2003; Kather *et al.*, 2005).

1.5 Modellproteine zur Untersuchung von Faltung und Stabilität von Proteinen

1.5.1 Das Kälteschockprotein CspB aus *Bacillus subtilis*

Die Kälteschockproteine (Csp) sind eine Familie von sequenzhomologen Proteinen, die im Rahmen der Kälteschockantwort in vielen Bakterien exprimiert werden (Graumann und Marahiel, 1998; Graumann und Marahiel, 1999; Graumann *et al.*, 1996). Diese monomeren Proteine bestehen aus 65-70 Aminosäureresten. Ihr Faltungsmechanismus erfüllt das Kriterium der Reversibilität der Entfaltung, weshalb sie als Modellproteine für Untersuchungen zur Proteinfaltung und Stabilität verwendet werden (Schindler *et al.*, 1995; Perl *et al.*, 1998). Ihre dreidimensionale Struktur weist fünf β -Faltblattstränge auf, welche antiparallel in Form eines β -Fasses angeordnet sind (Abbildung 2) (Schindelin *et al.*, 1993; Mueller *et al.*, 2000). Sie binden mit hoher Affinität an RNA und einzelsträngige DNA (Zeeb und Balbach, 2003; Graumann und Marahiel, 1994), weshalb ihnen Funktionen als „RNA-Chaperone“ (Phadtare, 2004; Jiang *et al.*, 1997) und Transkriptionsantiterminatoren (Phadtare *et al.*, 2002; Bae *et al.*, 2000) zugeschrieben werden.

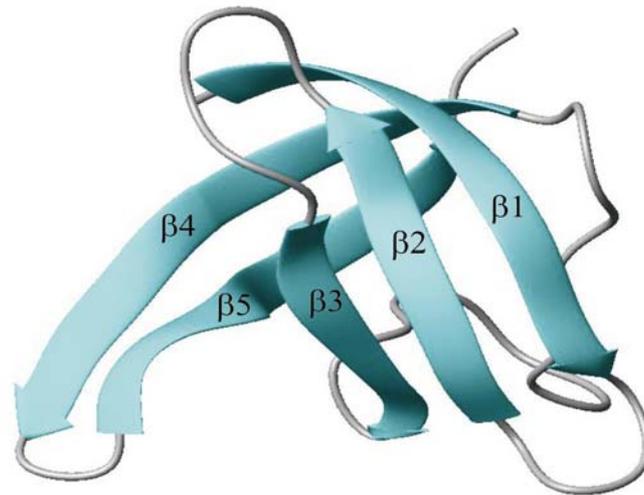


Abbildung 2. Tertiärstrukturdarstellung von Wildtyp *Bs*-CspB. Die β -Faltblattstränge sind nummeriert von $\beta 1$ bis $\beta 5$. Die Abbildung wurde erstellt unter Verwendung der Daten 1csp (Schindelin *et al.*, 1993) und dem Programm MOLMOL (Koradi *et al.*, 1996).

Die Kälteschockproteine treten in mesophilen, thermophilen und hyperthermophilen Bakterienstämmen auf. *Bs*-CspB aus dem mesophilen Organismus *Bacillus subtilis* und *Bc*-Csp aus dem thermophilen Bakterium *Bacillus caldolyticus* unterscheiden sich in lediglich zwölf oberflächenexponierten Resten. Der große Unterschied in der Stabilität ($\Delta\Delta G_D = 15,8 \text{ kJ mol}^{-1}$) kann dabei ausschließlich auf zwei Reste, Arg3 und Leu66, in *B. caldolyticus* zurückgeführt werden (Perl *et al.*, 2000). *Bs*-CspB weist an diesen beiden Positionen jeweils Glutamat auf, welches aufgrund von starker elektrostatischer Abstoßung das Protein destabilisiert (Perl und Schmid, 2001).

Zur Untersuchung des Stabilitätsunterschiedes zwischen dem mesophilen *Bs*-CspB und dem thermophilen *Bc*-Csp wurde neben ortsgerechter Analyse (Perl *et al.*, 2000) auch das *in vitro* Evolutionssystem *Proside* eingesetzt (Martin *et al.*, 2001). Dabei wurden an sechs Positionen, die bei den homologen Proteinen unterschiedliche Reste aufwiesen, alle 20 natürlich vorkommenden Aminosäurereste zugelassen und die entsprechenden Phagenbibliotheken einer *in vitro* Evolution unterworfen. Die selektierten Varianten wiesen eine deutlich erhöhte Stabilität auf, vergleichbar sogar mit der des hyperthermophilen Homologen aus *Thermotoga maritima* (*Tm*-Csp), auf. Dabei unterschied sich die stabilste Variante an allen sechs Positionen von dem thermophilen *Bc*-Csp. Die Stabilität des Proteins konnte somit auf unterschiedliche Art und Weise gesteigert werden, aber beide Wege folgten dem gleichen Prinzip, Optimierung der elektrostatischen Wechselwirkungen auf der Proteinoberfläche (Martin *et al.*, 2002).

Die Einfachheit der Kälteschockproteine und die Vielzahl experimenteller Stabilitätsdaten motivierten eine Vielzahl von theoretischen Analysen der Proteinstabilität (Zhou und Dong, 2003; Torrez *et al.*, 2003; Sanchez-Ruiz und Makhatazde, 2001; Garofoli *et al.*, 2004; Dominy *et al.*, 2002). Dabei wurden bei allen Arbeiten die Reste Glu3 und Glu66 in *Bs*-CspB als destabilisierend identifiziert. Dominy *et al.* (2002) gelang es, die Unterschiede in den Salzabhängigkeiten der Stabilität für die homologen Proteine nachzuvollziehen. Zhou und Dong (2003) setzten die Beiträge geladener Seitengruppen zur Stabilität in Beziehung zur Nettoladung des Proteins.

1.5.2 Die Domäne $\beta 1$ des Streptokokkenproteins G (G $\beta 1$)

Das Streptokokkenprotein G ist ein Multidomänenprotein und Bestandteil der Zellwand verschiedener Streptokokkenstämme. Abhängig vom Stamm besitzt es C-terminal zwei oder drei sehr homolog aufgebaute IgG-Bindungsdomänen (Sari *et al.*, 2000; Bouvet, 1994) sowie N-terminal eine Albumin-Bindungsdomäne (Johansson *et al.*, 2002; Akerstrom *et al.*, 1987). Die Antikörper werden im C-terminalen Bereich der schweren Kette, dem konstanten F_c-Teil, gebunden (Gronenborn *et al.*, 1991; Derrick und Wigley, 1992). Entsprechend stellt das Protein G, analog dem Protein A aus Staphylokokken, ein wichtiges Hilfsmittel in der Immunologie dar (Tashiro und Montelione, 1995).

Die $\beta 1$ -Domäne des Proteins G (G $\beta 1$) ist eine dieser IgG-Bindungsdomänen. Das kleine, aus 56 Aminosäureresten bestehende Protein ist aufgebaut aus einem viersträngigen β -Faltblatt und einer einzelnen α -Helix, die sich diagonal über das Faltblatt erstreckt (Abbildung 3) (Gronenborn *et al.*, 1991; Gallagher *et al.*, 1994). G $\beta 1$ ist dabei bereits in seiner Wildtyp-Form thermodynamisch sehr stabil und die Faltung ist reversibel. Aufgrund dieser Eigenschaften wurde es in einer Vielzahl von Faltungsstudien untersucht (Sheinerman und Brooks, 1998; Park *et al.*, 1999; Nauli *et al.*, 2001; McCallister *et al.*, 2000). Ein Hauptaugenmerk der Arbeiten mit G $\beta 1$ lag darauf, welche Aminosäurereste bevorzugt in welchem Sekundärstrukturelement auftreten und welchen Einfluss die Seitengruppen der Aminosäuren auf die Bildung und Stabilität von β -Faltblättern und α -Helices haben (Smith und Regan, 1995; Minor und Kim, 1994; Minor und Kim, 1996; Merkel *et al.*, 1999; Cregut *et al.*, 1999).

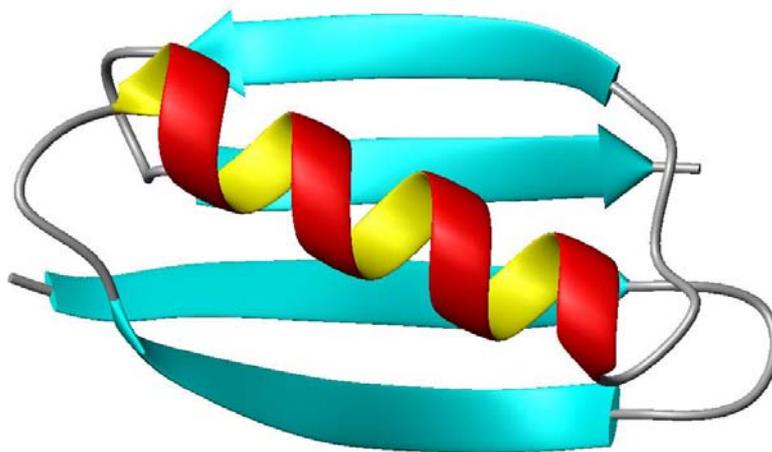


Abbildung 3. Tertiärstrukturdarstellung von Wildtyp G $\beta 1$. Die Abbildung wurde erstellt unter Verwendung der Daten 1PGB (Gallagher *et al.*, 1994) und dem Programm MOLMOL (Koradi *et al.*, 1996).

In der Gruppe von Stephen Mayo wurde der Computeralgorithmus ORBIT („Optimization of Rotamers by Iterative Techniques“) entwickelt, der zur rechnergestützten Stabilisierung von Proteinen dient (Dahiyat und Mayo, 1997a). Der Algorithmus untersucht dabei alle möglichen Sequenzen und die erlaubten Seitenkettenrotamere, um für ein strukturell fixiertes Proteinrückgrat die optimalen Aminosäureaustausche, sowie die beste Orientierung der

Seitenkette zu finden. Mit Hilfe dieser Methode wurde versucht G β 1 zu stabilisieren. In einem ersten Schritt wurden drei *core positions* (Dahiyat und Mayo, 1997b) untersucht und anschließend vier *boundary positions* (Malakauskas und Mayo, 1998). Bei den letzteren handelte es sich laut Definition um Positionen, die sich auf der Grenzfläche zwischen dem Inneren des Proteins und seiner Oberfläche befinden, d.h. teilweise exponierte Reste. Die Optimierung der *core positions* lieferte eine Dreifachmutante G β 1-c3 (Y3F/L7I/V39I), welche nur schwach stabilisiert war (Malakauskas und Mayo, 1998). Durch den anschließenden zweiten Schritt, der Optimierung von vier *boundary positions*, konnte die Stabilität deutlich erhöht werden und es ergab sich eine hyperthermophile Variante mit einem Schmelzpunkt von über 100 °C. Diese Variante beinhaltet neben den drei Optimierungen des Inneren des Proteins die Mutationen T16I/T18I/T25E/V29I, welche allein das Protein um 15,9 kJ mol⁻¹ stabilisieren (Malakauskas und Mayo, 1998). Neben dieser experimentell charakterisierten Sequenz lieferte der Computeralgorithmus neun weitere Kombinationen als Ersatz für die vier *boundary positions*, die in den Berechnungen eine annähernd gleiche Stabilisierung aufwiesen, aber experimentell von den Autoren nicht charakterisiert wurden.

1.6 Problemstellung

Ziel der hier vorliegenden Arbeit war, das Selektionssystem *Proside* (Sieber *et al.*, 1998) zur Stabilisierung verschiedener Proteine einzusetzen und die erhaltenen Proteinvarianten in Bezug auf ihre Stabilität und Faltungseigenschaften zu charakterisieren.

Bei der Selektion lag ein Augenmerk darauf durch Rekombination bedingte unerwünschte Eliminierung der Sequenz des Gastproteins aus dem Phagen genom zu verhindern bzw. zu unterdrücken. Die Rekombination in Kombination mit der für verschiedene Gastproteine deutlich erniedrigten Infektiosität des Phagen verhinderten eine erfolgreiche Selektion des humanen p19 Proteines, der RNase H aus *E. coli*, der Dihydrofolatreduktase aus *E. coli* und der Luziferase aus *Photinus pyralis*. Erfolgreich stabilisiert werden konnte im Rahmen dieser Arbeit die variable Domäne von Kamelantikörpern (Daten nicht gezeigt), die Domäne β 1 des Streptokokkenproteins G und *Bs-CspB* aus *B. subtilis*. Im Folgenden wird ausschließlich auf die beiden letzteren Bezug genommen.

Als erstes Protein zur Untersuchung wurde das Kälteschockprotein *Bs-CspB* aus *B. subtilis* gewählt. Aufgrund seiner bereits erwähnten Eigenschaften eignet es sich sehr gut für thermodynamische Untersuchungen der Proteinestabilität (Schindler *et al.*, 1995; Perl *et al.*, 1998). Durch ortgerichtete Mutagenese (Perl und Schmid, 2001; Perl *et al.*, 2000) und *in vitro* Evolution (Martin *et al.*, 2001; Martin *et al.*, 2002) konnten dabei die Gründe für die unterschiedliche Stabilität des mesophilen *Bs-CspB* und des thermophilen *Bc-Csp* analysiert werden. In beiden Fällen wurde die Bedeutung von elektrostatischen Wechselwirkungen auf der Oberfläche des Proteins klar gezeigt.

Im Folgenden sollten alternative Möglichkeiten für den Aufbau eines verbesserten Ladungsnetzwerkes auf der Oberfläche von *Bs-CspB* gefunden werden. Hierfür wurden anhand von Strukturanalysen und Vergleichen mit anderen homologen Proteinen aussichtsreiche Positionen identifiziert und mittels *in vitro* Evolution optimiert. Zusätzlich wurden ortsungerichtet Mutationen in das Gen von *cspB* eingeführt und stabilisierte Varianten aus den so erstellten Bibliotheken auf selektiert. Ziel war es, mit Hilfe der thermodynamischen Daten der selektierten Varianten und hochaufgelöster Strukturen die

molekularen Ursachen für die Thermostabilität zu ergründen. Ein besonderes Augenmerk galt dabei den oberflächenexponierten Resten und ihren elektrostatischen Wechselwirkungen. Außerdem sollten die experimentellen Daten mit vorhandenen theoretischen Untersuchungen und Vorhersagen für das Protein verglichen werden.

Die Domäne $\beta 1$ des Streptokokkenproteins G wurde bereits unter Einsatz des *computational design* optimiert (Malakauskas und Mayo, 1998; Dahiyat und Mayo, 1997b). Dabei zeigte sich, dass die vier *boundary positions* 16, 18, 25 und 29 ein sehr großes Stabilisierungspotential besitzen. Die Untersuchung derselben vier Positionen durch das *in vitro* Evolutionssystem *Proside* bot somit die Möglichkeit, die Leistungsfähigkeit beider Methoden direkt zu vergleichen.

In einem zweiten Ansatz war es das Ziel, G $\beta 1$ durch ein rein experimentelles Herangehen zu stabilisieren. Dabei sollten aussichtsreiche Positionen für die Stabilisierung durch fehlerhaftete PCR und *in vitro* Evolution identifiziert werden und in einem zweiten Schritt durch eine Sättigungsmutagenese die besten Aminosäureaustausche an diesen Positionen gefunden werden.

Die aus beiden Ansätzen gewonnenen Varianten von G $\beta 1$ sollten thermodynamisch charakterisiert und die molekularen Ursachen für die Stabilisierung zum Teil unter zu Hilfenahme von hochaufgelösten Strukturen ergründet werden. Außerdem war es ein Ziel die Ergebnisse beider Ansätze, eines rein experimentellen und eines gemischt theoretischen und experimentellen, miteinander zu vergleichen, um die Effizienz und die Möglichkeiten beider Wege aufzuzeigen.

2. Zusammenfassung und Diskussion der Ergebnisse

2.1 Evolutive Stabilisierung des Kälteschockproteins CspB

2.1.1 Ortsgerichtete Optimierung von oberflächenexponierten Positionen

Nach der erfolgreichen Stabilisierung von *Bs*-CspB durch ortsgerechte Mutagenese (Perl *et al.*, 2000) und Selektion (Martin *et al.*, 2001) an Positionen, die an Hand des Vergleiches mit dem thermophilen Homologen *Bc*-Csp ausgewählt worden waren, war jetzt das Ziel, das Protein durch den Aufbau alternativer Ladungsnetzwerke zu stabilisieren. Dazu wurden neun oberflächenexponierte Positionen (5, 10, 12, 19, 20, 22, 43, 46 und 48) ausgesucht, die sich entfernt von der Nukleotidbindungsstelle befanden und in homologen Proteinen von *Bs*-CspB durch polare oder geladene Aminosäuren besetzt sind. In den entsprechenden Bibliotheken waren an diesen Positionen unter Ausnutzung der Degeneration des genetischen Codes zumindest eine positiv und eine negativ geladene, eine polare und eine unpolare Aminosäure zugelassen. Eine genaue Auflistung der an den jeweiligen Positionen zugelassenen Aminosäuren ist in Tabelle 7 (Teilarbeit A) aufgeführt.

Die Selektion wurde im wesentlichen durchgeführt wie von Andreas Martin beschrieben (Martin *et al.*, 2003). Ein zweiter Proteolyseschritt mit Trypsin wurde eingeführt, um rekombinante, gegen Chymotrypsin resistente Phagen mit einem verkürzten *Bs*-CspB Insert zu unterdrücken. Außerdem wurden rekombinante Phagen nach verschiedenen Selektionsrunden entfernt, indem der gesamte Genpool amplifiziert und nur Fragmente, die dem kompletten Gen entsprachen, wieder in den Phagen eingebaut wurden.

Am Ende der Selektion konnten nur Mutationen im C-terminalen Bereich (Positionen 43, 46 und 48) gefunden werden, im N-terminalen Bereich (Positionen 5 bis 22) wiesen alle untersuchte Klone Wildtypsequenz auf. Durch Sequenzierung der Bibliothek vor Beginn der Selektion konnte gezeigt werden, dass die Codons an den Positionen 5, 10, 12, 19, 20 und 22 in der Tat randomisiert waren. Außerdem wurde eine zusätzliche Bibliothek angelegt, in der ausschließlich diese sechs N-terminalen Reste variiert waren. Nach Selektion wurden wiederum ausschließlich Wildtyp Reste an den randomisierten Positionen gefunden. Dies legt den Schluss nahe, dass diese Positionen die Stabilität des Proteins nur unbedeutend beeinflussen.

Für die Positionen 46 und 48 ergab sich am Ende der Selektion ein sehr klares Bild, es wurden fast ausschließlich positiv geladene Reste gefunden. An Position 43 wurden von den sieben zugelassenen Aminosäuren vier selektiert und dabei waren positive, negative und ungeladene Reste. Dies deutet darauf hin, dass diese Position für die Stabilität von untergeordneter Bedeutung ist. Fünf der selektierten Varianten wurden gereinigt und thermodynamisch charakterisiert. Alle waren um 11-12 kJ mol⁻¹ stabilisiert und die Mittelpunkte der thermisch induzierten Entfaltungsübergänge lagen um etwa 16 °C höher als für das Wildtypprotein. Der Übergang der Variante A46K/S48R ist in Abbildung 4 im Vergleich zu Wildtyp *Bs*-CspB gezeigt. Diese beiden Mutationen stabilisieren das Wildtypprotein allein um 11,1 kJ mol⁻¹ (Tabelle 1). Der Beitrag der Position 43 zur Stabilität der selektierten Varianten ist somit in der Tat unbedeutend.

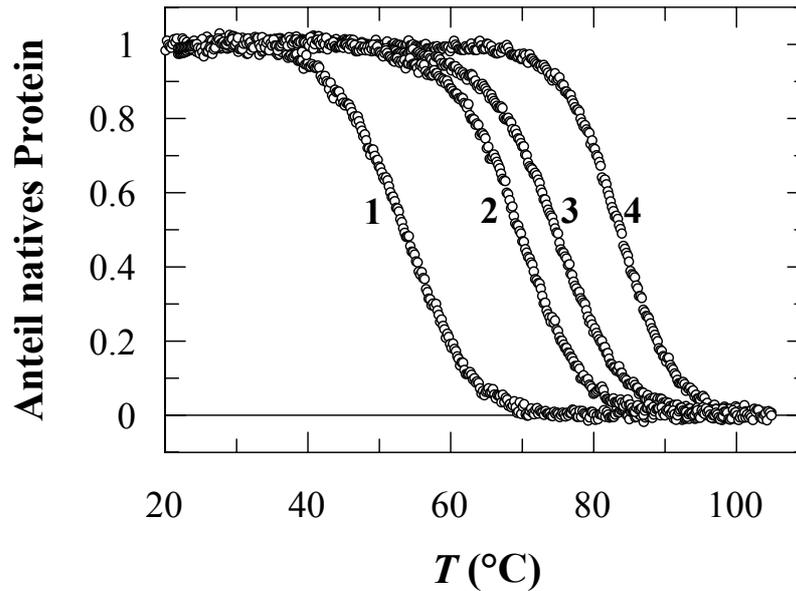


Abbildung 4. Thermische Entfaltungsübergänge von Wildtyp *Bs*-CspB (1) und den Varianten A46K/S48R (2), M1R/E3K/K65I/E66K (3) und M1R/E3K/K65I (4) in 100 mM Na-Cacodylat, pH 7,0 wurden verfolgt durch die Änderung des CD-Signals bei 222,6 nm. Die Messungen wurden durchgeführt wie in Teilarbeit A beschrieben. Dargestellt ist der Anteil an nativem Protein als Funktion der Temperatur nach Analyse der Übergänge entsprechend einem Zweizustandsmodell.

Um die individuellen Beiträge der beiden Mutationen A46K und S48R zu bestimmen, wurden sie einzeln durch ortsgerichtete Mutagenese in das Gen eingeführt. Dabei ergaben sich Stabilisierungen von 5,9 bzw. 6,6 kJ mol⁻¹ (Tabelle 1), was zeigt, dass die Effekte von Lys46 und Arg48 in etwa gleichwertig und additiv sind. Die Ursache der Stabilisierung kann man anhand der Struktur des Proteins begründen (Abbildung 5). Wie bereits erwähnt tritt im Wildtypprotein eine starke elektrostatische Abstoßung zwischen Glu3 und Glu66 auf (Perl *et al.*, 2000). Die Positionen 46 und 48 liegen flankierend zur Verbindungsachse der beiden Glutamate und die Einführung von zwei positiv geladenen Resten hebt die Abstoßung auf. In der Kristallstruktur der Variante A46K/S48R (unveröffentlichte Daten) sind die Seitenketten von Lys46 und Arg48 nur schlecht aufgelöst und besitzen sehr hohe B-Faktoren. Dies zeigt die Beweglichkeit der beiden Reste. Sie bilden keine festen Salzbrücken aus, sondern dienen dem allgemeinen Ladungsausgleich des stark negativ geladenen Proteins in einem Ladungsnetzwerk.

Tabelle 1. Thermodynamische Parameter für Wildtyp *Bs*-CspB und Varianten.

Variante	T_M (°C)	ΔH_D (kJ mol ⁻¹)	$\Delta\Delta G_D$ (kJ mol ⁻¹)		
			0 M NaCl (Total)	0-2 M NaCl (polar)	2 M NaCl (nicht-polar)
Wildtyp <i>Bs</i> -CspB	53,8	193	0	0	0
A46K/S48R	69,7	219	11,1	6,1	5,0
A46K	62,2	216	5,9	2,7	3,2
S48R	62,7	202	6,6	2,6	4,0
M1R	64,2	226	7,3	3,9	3,4
E3K	70,4	203	11,5	8,4	3,1
K65I	62,4	201	6,4	0	6,4
E66K	66,7	226	9,1	5,2	3,9
M1R/E3K	77,0	262	16,3	10,4	5,9
M1R/E3K/K65I	83,7	278	20,9	8,2	12,7
M1R/E3K/K65I/E66K	74,7	214	14,1	6,1	8,0
M1R/E3K/K65I/E66L	85,0	241	20,1	5,7	14,4

T_M ist der Mittelpunkt des thermischen Entfaltungsüberganges, ΔH_D ist die Enthalpie der Entfaltungsreaktion am Mittelpunkt, $\Delta\Delta G_D$ ist die Änderung der Gibbs'schen freien Energie der Entfaltung bei 70 °C relativ zu der des Wildtypproteins. $\Delta\Delta G_D$ gemessen bei 0 M NaCl stellt die gesamte Änderung, $\Delta\Delta G_D$ bei 2 M NaCl den hydrophoben Anteil und die Differenz der Werte zwischen 0 M und 2 M NaCl den elektrostatischen Anteil dar. Die thermodynamischen Parameter wurden aus thermischen Entfaltungsübergängen, wie in Abbildung 4 gezeigt, ermittelt. Eine Aufstellung für alle in dieser Arbeit charakterisierten Varianten findet sich in Tabelle 2 (Teilarbeit A).

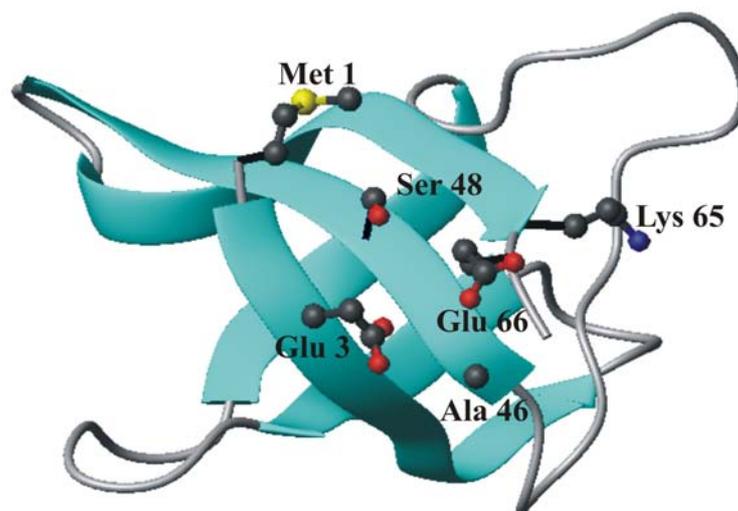


Abbildung 5. Tertiärstrukturdarstellung von Wildtyp *Bs*-CspB. Die Seitenketten, der in den Selektionen gefundenen, stabilisierenden Positionen, sind als Kugel-Stab-Modell gezeigt. Die Abbildung wurde erstellt unter Verwendung der Daten 1csp (Schindelin *et al.*, 1993) und dem Programm MOLMOL (Koradi *et al.*, 1996).

In einer theoretischen Arbeit (Torrez *et al.*, 2003) wurden publizierte Stabilitätsdaten (Perl und Schmid, 2001) unter Verwendung von Poisson-Boltzmann-Elektrostatikrechnungen reproduziert. Dabei wurde ein Korrelationskoeffizient von 0.86 zwischen berechneten und gemessenen Daten erzielt. Für die Mutation A46K wurde in diesem Rahmen ein stark destabilisierender Effekt von $-7,5 \text{ kJ mol}^{-1}$ vorhergesagt. Tatsächlich ist diese Mutation mit einer Änderung von $\Delta\Delta G_D = 5.9 \text{ kJ mol}^{-1}$ stark stabilisierend, weshalb diese Mutation häufig in den Selektionen gefunden wurde. Diese enorme Diskrepanz zeigt beispielhaft die Schwierigkeit auf, elektrostatische Einflüsse korrekt zu berechnen.

2.1.2 Ortsunspezifische Optimierung von *Bs-CspB*

Eine weitere Serie von *Proside*-Selektionen wurde durchgeführt um *Bs-CspB* in ortsungerichteter Weise zu stabilisieren. Dazu wurden Bibliotheken von *Bs-CspB*-Mutanten zum einen durch fehlerhafte PCR und zum anderen durch die natürliche Mutationsrate des Phagen während der Propagation in *E. coli* angelegt. Die entsprechenden Phagenbibliotheken wurden wiederum unter Verwendung von Chymotrypsin und Trypsin auf stabilisierte Varianten selektiert und entsprechende rekombinante Phagen, wie bereits beschrieben, entfernt.

Bei diesen Selektionen wurden die vier stabilisierenden Mutationen M1R, E3K, K65I und E66K identifiziert, alle lagen auf der Oberfläche des Proteins (Abbildung 5). Im Inneren des Proteins wurden zwar auf DNA-Ebene ebenfalls Mutationen gefunden, aber diese hatten keine Auswirkung auf die Aminosäuresequenz, sie waren still. Dies zeigt, dass die Packung des Proteinkerns bereits sehr gut ist und nur schwer durch Mutagenese verbessert werden kann. Die selektierten Proteinvarianten wurden gereinigt und thermodynamisch charakterisiert (Tabelle 1). Die beiden Mutationen E3K und E66K wirkten stark stabilisierend und erhöhten ΔG_D um 11,5 bzw. 9,1 kJ mol^{-1} . Dies bestätigt das hohe Stabilisierungspotential der beiden Positionen, welches im Vergleich zwischen dem mesophilen *Bs-CspB* und dem thermophilen *Bc-Csp* gefunden wurde (Perl *et al.*, 2000). Die Mutationen M1R und K65I erhöhen die Stabilität des Proteins um 7,3 bzw. 6,4 kJ mol^{-1} . Der Effekt von K65I scheint überraschend, denn dadurch wird die bereits hohe Nettoladung von *Bs-CspB* von -6 auf -7 erhöht.

Die thermodynamische Stabilität von Proteinen in Abhängigkeit der Salzkonzentration gibt Aufschluss über elektrostatische Wechselwirkungen auf der Proteinoberfläche oder im entfalteten Zustand. Die im Lösungsmittel vorliegenden Ionen können geladene Gruppen von exponierten Seitenketten abschirmen und somit eine Interaktion unterbinden. Dieser Effekt kann durch die Zugänglichkeit der Seitenketten beeinflusst werden. Zur Bestimmung des Coulomb'schen Anteils an der gesamten Stabilisierung dient im Folgenden die Differenz der Messwerte für $\Delta\Delta G_D$ bei 0 M und 2 M NaCl. Diese Differenz stellt eine untere Grenze für die Coulomb'schen Beiträge dar, denn nicht alle elektrostatischen Wechselwirkungen können durch 2 M NaCl abgeschirmt werden (Spencer *et al.*, 2005; Luisi *et al.*, 2003; Lee *et al.*, 2002). Es stellt jedoch das beste experimentell zugängliche Maß dar.

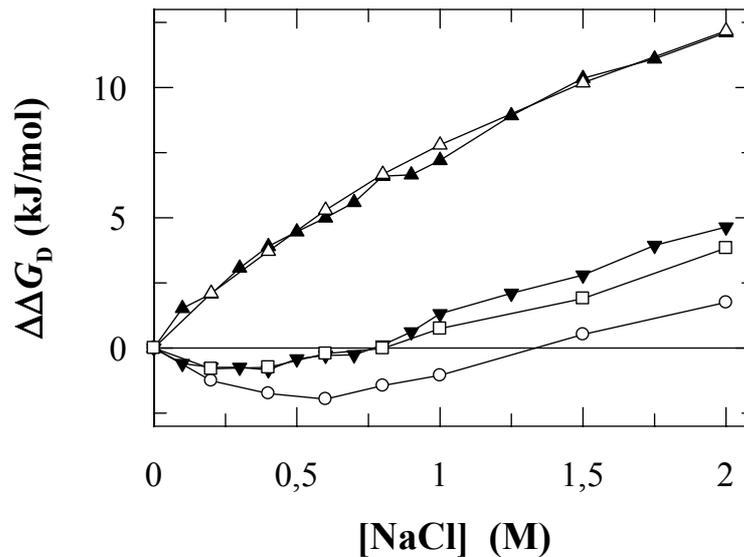


Abbildung 6. Änderung der freien Enthalpie der Entfaltung bei 70 °C in Abhängigkeit von der NaCl-Konzentration. Die Auftragung erfolgte relativ zum Wert in Abwesenheit von NaCl für die Wildtypproteine *Bs*-CspB (▲) und *Bc*-Csp (▼), sowie für die Varianten M1R/E3K (○), M1R/E3K/K65I (□) und K65I (△) von *Bs*-CspB.

Die Salzabhängigkeit der Stabilität für das Wildtypprotein *Bs*-CspB zeigt bei niedrigen Salzkonzentrationen eine sprunghafte Erhöhung (Abbildung 6). Dies liegt daran, dass die Ionen ungünstige elektrostatische Wechselwirkungen auf der Proteinoberfläche abschirmen. In Gegenwart von höheren Salzkonzentrationen steigt die Stabilität langsamer an gemäß der allgemein stabilisierenden Wirkung von NaCl (Cacace *et al.*, 1997). Im Gegensatz dazu fällt für *Bc*-Csp die Stabilität in Gegenwart von geringen Salzkonzentration, was gleichbedeutend mit einer Abschirmung von günstigen elektrostatischen Wechselwirkungen ist, und erst bei höheren Konzentrationen an NaCl kommt es zu einer Stabilisierung gemäß der Hofmeister-Reihe (Cacace *et al.*, 1997). Die Stabilität der Variante K65I zeigt dieselbe Salzabhängigkeit wie das Wildtypprotein *Bs*-CspB. Dies bedeutet, dass die Stabilisierung durch verbesserte hydrophobe Wechselwirkungen bedingt wird. Elektrostatische Einflüsse wie der Wegfall von ungünstigen Wechselwirkungen durch das Lys65 im Wildtypprotein oder die Stabilisierung des denaturierten Zustands (Pace *et al.*, 2000; Cho und Raleigh, 2005) durch die positive Ladung können somit ausgeschlossen werden.

Um Additivität der Beiträge der vier selektierten Mutationen zu untersuchen, wurde die vierfach mutierte Variante M1R/E3K/K65I/E66K durch ortsgerechte Mutagenese hergestellt. Der Mittelpunkt ihres thermischen Entfaltungsüberganges ist für um 20,9 °C erhöht (Abbildung 4), was einer Stabilisierung von 14,1 kJ mol⁻¹ entspricht. Dies ist viel geringer als die Summe der Einzelbeiträge, welche 34,4 kJ mol⁻¹ beträgt (Tabelle 1). Die Positionen 1, 3, 65 und 66 befinden sich im gefalteten Protein in lokaler Nachbarschaft (Abbildung 5). Dies erklärt, warum die Effekte nicht additiv sind.

Zur Analyse dieser wechselseitigen Abhängigkeit und um eine stärker stabilisierte Variante zu finden, wurden die vier Mutationen in unterschiedlicher Weise miteinander kombiniert. Eine vollständige Auflistung aller charakterisierten Varianten findet sich in Tabelle 2 (Teilarbeit A). Die höchste Stabilisierung wurde für die Dreifachmutante M1R/E3K/K65I erzielt. Der Mittelpunkt ihres thermischen Überganges lag bei 83,7 °C (Abbildung 4) und die freie Enthalpie der Entfaltung war um 20,9 kJ mol⁻¹ erhöht (Tabelle 1). Diese Variante mit ihren drei Mutationen war somit stabiler als das hyperthermophile Homologe *Tm*-Csp aus

Thermotoga maritima (Wassenberg *et al.*, 1999). Die Stabilität in Abhängigkeit der NaCl-Konzentration zeigte das gleiche Verhalten wie *Bc-Csp* (Abbildung 6). Dies zeigt, dass ein Teil dieser enormen Stabilisierung durch verbesserte elektrostatische Wechselwirkungen verursacht wird. Der elektrostatische Anteil beträgt dabei auf $8,2 \text{ kJ mol}^{-1}$ (Tabelle 1).

Um die molekularen Ursachen für die gesteigerte Stabilität zu ergründen, wurde die Kristallstruktur der Variante M1R/E3K/K65I bestimmt (unveröffentlichte Daten). Dabei zeigte sich vor allem für die Position 1 ein sehr klares Bild (Abbildung 7). Durch Einführung des Arginins kommt es zu der Aufspannung eines Netzwerkes verschiedener Wechselwirkungen mit teilweise sehr großen Änderungen in der Ausrichtung der beteiligten Seitengruppen. So bildet sich zwischen den Seitenketten von Arg1 und Glu50 eine Salzbrücke aus. Die Guanidylgruppe des Arginins bildet außerdem insgesamt vier Wasserstoffbrückenbindungen mit den Seitenketten der Reste Ser48 und Thr64 und der Carbonylgruppe des Rückgrats von Asn62 aus, dabei geht durch die Umorientierung eine Wasserstoffbrückenbindung zwischen Asn62 und Thr64 verloren. Die Seitengruppen der beteiligten Reste werden dabei um bis zu 120° im Vergleich zum Wildtyp gedreht. Diese zusätzlichen Wasserstoffbrücken erklären vermutlich einen wesentlichen Teil der für die Mutation M1R experimentell beobachteten Stabilisierung.

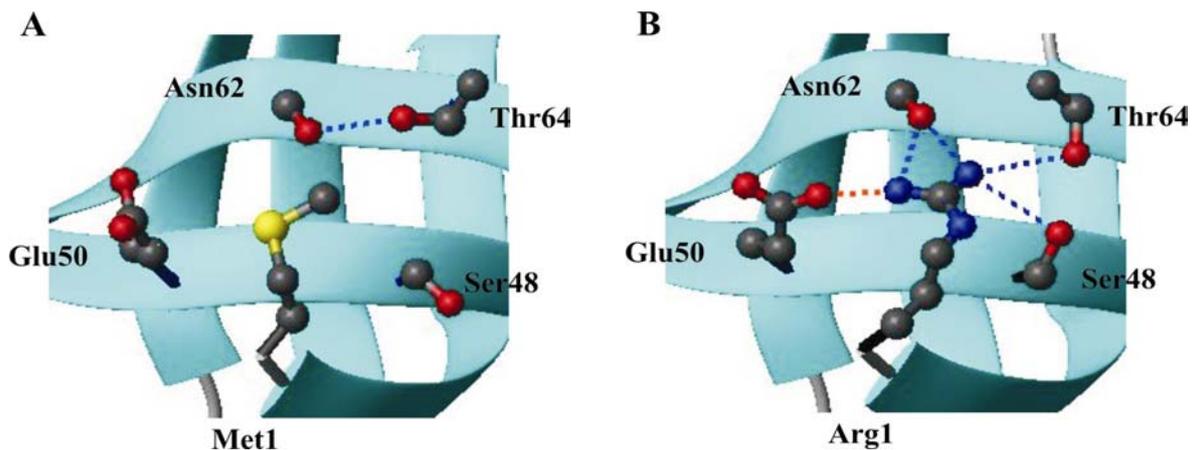


Abbildung 7. Detailausschnitt aus der Umgebung der Position 1 von A) Wildtyp *Bs-CspB* und B) der Variante M1R/E3K/K65I. Die Seitenketten der Positionen 1, 48, 50 und 64 sowie die Carbonylgruppe des Rückgrats von Position 62 sind als Kugel-Stab-Modell gezeigt. Als gestrichelte Linien sind blau Wasserstoffbrückenbindungen und rot Salzbrücken angegeben. Die Abbildung wurde erstellt unter Verwendung von 1csp (Schindelin *et al.*, 1993) für das Wildtypprotein, den von Klaas Max zur Verfügung gestellt Daten für die Variante M1R/E3K/K65I und dem Programm MOLMOL (Koradi *et al.*, 1996).

Für die Seitenkette von Lys3 konnte in der Struktur kein direkter Wechselwirkungspartner gefunden werden. Sie dient vermutlich der allgemeinen Kompensation der stark negativen Überschussladung von -6 im Wildtypprotein. Ein formaler Doppelmutationszyklus zwischen Arg1 und Lys3 zeigte, dass die hydrophoben Anteile der Stabilisierung beider Mutationen ungefähr additiv sind. Die Summe der elektrostatisch bedingten Stabilisierung der beiden Einzelmutationen liegt um $1,9 \text{ kJ mol}^{-1}$ höher als der in der Doppelmutante gemessene Wert. Dies zeigt, dass Arg1 und Lys3 trotz eines großen Abstandes von mehr als 10 \AA der geladenen Seitengruppen elektrostatisch interagieren.

Die Stabilisierung durch K65I ist, wie bereits erwähnt, ausschließlich auf hydrophobe Wechselwirkungen zurückzuführen, da die Salzabhängigkeit ihrer Stabilität einen identischen Verlauf wie beim Wildtypprotein zeigt (Abbildung 6). In der Struktur bildet sich ein hydrophober Cluster mit den Seitenketten der Reste Val6, Ile18 und Val63 aus. Im Wildtypprotein ist die Oberfläche von Lys65 nur zu 32,6 % für das Lösungsmittel zugänglich und der Anteil sinkt für Ile65 in der Variante auf 24,0 %. Der stark stabilisierende Effekt eines Austausches von einer polaren zu einer unpolaren Aminosäure an einer solchen nur teilweise exponierten Position, wurde auch für die Domäne $\beta 1$ des Streptokokkenproteins G beschrieben (Malakauskas und Mayo, 1998). Die Bedeutung des Lys65 im Wildtypprotein wurde auch theoretisch unter Verwendung des Tanford-Kirkwood (Sanchez-Ruiz und Makhatadze, 2001) und des Poisson-Boltzmann Modells (Dominy *et al.*, 2002) untersucht. Beide Methoden kamen zu dem Ergebnis, dass Lys65 eine starke elektrostatische Stabilisierung des Wildtypproteins bedingt. Ein Austausch von Lys65 sollte also stark destabilisierend sein. Experimentell konnten diese Vorhersagen nicht bestätigt werden und es ist kein elektrostatischer Einfluss durch Lys65 festzustellen. K65I wirkt stark stabilisierend und es ist keine Veränderung der Coulomb'schen Wechselwirkungen durch diese Mutation nachweisbar.

Eine zusätzliche Stabilisierung der Variante M1R/E3K/K65I erhält man, wenn man diese drei Mutationen mit der Mutation E66L kombiniert, die in vorangegangenen *Prosiede* Selektionen gefunden wurde (Martin *et al.*, 2001). Die daraus resultierende Variante mit vier Mutationen M1R/E3K/K65I/E66L weist einen Schmelzpunkt von 85,0 °C auf (Tabelle 1). Eine ebenfalls bereits sehr stark stabilisierte Variante von *Bs-CspB* stellt die Doppelmutante M1R/E3K dar ($\Delta\Delta G_D = 16,3 \text{ kJ mol}^{-1}$; Tabelle 1). Diese weist die höchste Stabilisierung durch elektrostatische Effekte auf, wie man anhand des stärksten Abfalls der Stabilität bei niedrigen Salzkonzentrationen erkennen kann (Abbildung 6). Die Mutation K65I verändert die elektrostatischen Wechselwirkungen nicht, wenn sie in das Wildtypprotein eingeführt wird (Nettoladung -6; vergleiche die beiden oberen Kurven in Abbildung 6). Wenn man sie jedoch in die Variante M1R/E3K (Nettoladung -3) einführt, verschiebt sich die Salzabhängigkeit nach oben, was gleichbedeutend mit einer Verschlechterung der Coulomb'schen Wechselwirkungen des Proteins ist (vergleiche die beiden unteren Kurven Abbildung 6). Es wird interessant sein, ob Computerprogramme zur Berechnung der Elektrostatik von Proteinen diesen Effekt reproduzieren können. Außerdem wurde eine größere Anzahl von weniger günstigen Kombinationen der gefundenen Mutationen geschaffen und charakterisiert (Tabelle 2, Teilarbeit A). Die experimentellen Daten dieser Varianten werden sich sehr gut dafür eignen, Elektrostatikberechnungen zu validieren.

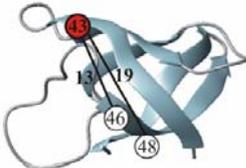
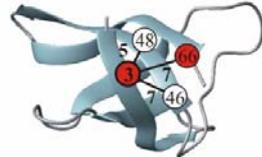
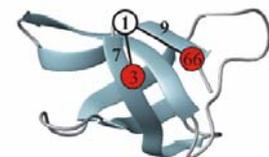
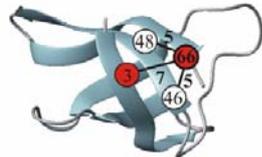
2.1.3 Abhängigkeit der Effekte eines Ladungsaustausches von der Umgebung

Die vielfältigen Daten für diese Varianten wurden im weiteren genutzt, um den Einfluß von Ladungen in der Umgebung auf einzelne Mutationen zu charakterisieren.

Durch die Mutationen E43S und E66L wird jeweils eine negative Ladung entfernt und durch eine ungeladene Seitengruppe ersetzt. Die Daten für die Mutation E43S sind in Tabelle 2a zusammengefasst. Diese Mutation verringert die Nettoladung des Wildtypproteins von -6 auf -5 und in den Varianten A46K/S46R und A46K/S48K von -4 auf -3. Der Abstand der $C\beta$ -Atome von Ser43 und den Positionen 46 und 48 beträgt 12,7 bzw. 18,6 Å. Der elektrostatische Einfluss der Mutation E43S auf die Stabilität ist in allen drei Fällen gering (0,8; 0,5

bzw. $1,7 \text{ kJ mol}^{-1}$). Dies bestätigt Ergebnisse für die Staphylokokken Nuklease (Lee *et al.*, 2002), dass Ladungsänderungen an entfernten Positionen die elektrostatischen Wechselwirkungen nicht beeinflussen.

Tabelle 2. Abhängigkeit der Mutationen E43S (a), E66L (b), M1R (c) und E3R (d) von der umgebenden Ladungsverteilung

	Hintergrund	Netto- ladung	ΔT_M (°C)	$\Delta\Delta G_D$ (kJ mol ⁻¹)		
				0 M NaCl (Total)	0-2 M NaCl (polar)	2 M NaCl (non-polar)
<i>(a) E43S in verschiedenen Hintergründen</i>						
	A46 S48	-5	0,9	1,2	0,8	0,4
	K46 R48	-3	0,2	0,2	0,5	-0,3
	K46 K48	-3	1,8	1,1	1,7	-0,6
<i>(b) E66L in verschiedenen Hintergründen</i>						
	E3 A46 S48	-5	12,8	8,8	2,8	6,0
	E3 K46 R48	-3	7,3	4,6	1,1	3,5
	R3 A46 S48	-3	5,0	3,1	-0,9	4,0
	R3 K46 R48	-1	-3,7	-2,4	-4,5	2,1
<i>(c) M1R in verschiedenen Hintergründen</i>						
	E3 E66	-5	10,4	7,3	3,9	3,4
	K3 E66	-3	6,6	4,8	2,0	2,8
	E3 K66	-3	2,8	1,9	1,2	0,7
	K3 K66	-1	-3,5	-1,5	-2,7	1,2
<i>(d) E3R in verschiedenen Hintergründen</i>						
	A46 S48 E66	-4	16,0	11,1	7,9	3,2
	A46 S48 L66	-3	8,2	5,6	4,2	1,2
	K46 R48 E66	-2	4,8	4,3	4,3	0
	K46 R48 L66	-1	-6,2	-4,0	-2,5	-1,5

Der erste Eintrag für jedes Beispiel stellt die Reste im Wildtyp Protein dar. Die Werte ΔT_M und $\Delta\Delta G_D$ geben die Änderung der Stabilität an, welche durch die Mutation in verschiedenen Hintergründen verursacht wird. Sie wurden berechnet aus Tabelle 4 (Teilarbeit A). $\Delta\Delta G_D$ bei 0 M NaCl ist die Gesamtänderung, $\Delta\Delta G_D$ bei 2 M NaCl stellt den hydrophoben Beitrag dar und die Differenz zwischen 0 M und 2 M NaCl den elektrostatischen Anteil von $\Delta\Delta G_D$. Negativ geladenen Reste sind rot gefärbt und positive blau. In der Tertiärstrukturdarstellung ist die Position der Mutation durch einen dicken Kreis markiert und die Nummern an den verbindenden Linien geben den Abstand zwischen den C^βAtomen der entsprechenden Reste in Å.

Der Effekt des Austausches E66L hängt sehr stark vom Hintergrund ab (Tabelle 2b). Im Wildtypprotein wirkt er stabilisierend ($\Delta\Delta G_D = 8,8 \text{ kJ mol}^{-1}$) mit einem günstigen elektrostatischen Beitrag von $2,8 \text{ kJ mol}^{-1}$ (Perl *et al.*, 2000). Dieser Beitrag verringert sich auf $1,1 \text{ kJ mol}^{-1}$ in der Variante A46K/S48R und auf $-0,9 \text{ kJ mol}^{-1}$ in der Variante E3R. In Gegenwart aller drei Mutationen ist der Austausch E66L destabilisierend ($\Delta\Delta G_D = -2,4 \text{ kJ mol}^{-1}$), weil der Coulomb'sche Beitrag mit $-4,5 \text{ kJ mol}^{-1}$ jetzt sehr ungünstig ist. Die Mutation E66L erhöht den Mittelpunkt des thermischen Überganges um $12,8 \text{ }^\circ\text{C}$, wenn sie in den Wildtyp eingeführt wird. Aber sie senkt ihn um $3,7 \text{ }^\circ\text{C}$ bei Einführung in die Variante E3R/A46K/S48R. Ein solches Verhalten kann nicht durch die Veränderung der Nettoladung erklärt werden, welche von theoretischer Seite häufig zur Erklärung von elektrostatischen Effekten herangezogen wird (Zhou und Dong, 2003). Diese ist für alle Varianten negativ. Offensichtlich dominiert hier die starke Abstoßung zwischen benachbarten Seitengruppen. Die variierten Positionen befinden sich in unmittelbarer Nähe, die Entfernungen der C^β -Atome von Position 66 nach 3, 46 und 48 betragen 6,6; 4,6 bzw. 4,8 Å. Ein sehr ähnliches Verhalten ergibt sich für den Austausch M1R (Tabelle 2c).

Am stärksten ausgeprägt ist die Umgebungsabhängigkeit für den Austausch E3R (Tabelle 2d). Diese Mutation trägt den größten Anteil zur Stabilisierung von *Bc*-Csp bei (Perl *et al.*, 2000). Im Wildtypprotein *Bs*-CspB verdoppelt der Austausch E3R die Stabilität des Proteins nahezu ($\Delta\Delta G_D = 11,1 \text{ kJ mol}^{-1}$), hauptsächlich durch günstige elektrostatische Wechselwirkungen von $7,9 \text{ kJ mol}^{-1}$. Dieser Beitrag wird deutlich gesenkt, wenn eine negative Ladung in der Umgebung durch den Austausch E66L entfernt wird oder durch die Einführung von zwei positiven Ladungen an den Positionen 46 und 48 (auf $4,2 \text{ kJ mol}^{-1}$ bzw. $4,3 \text{ kJ mol}^{-1}$). In Gegenwart aller drei Mutationen A46K/S48R/E66L ist der Austausch E3R destabilisierend und der Coulomb'sche Beitrag wird ungünstig ($\Delta\Delta G_D = -2,5 \text{ kJ mol}^{-1}$). Die gleiche Mutation E3R erhöht einmal den Mittelpunkt des thermischen Übergangs um $16 \text{ }^\circ\text{C}$ und einmal senkt sie ihn um $6,2 \text{ }^\circ\text{C}$. Diese Änderung entspricht einem $\Delta\Delta G_D$ von $15,1 \text{ kJ mol}^{-1}$, was mehr als die freie Enthalpie des Wildtypproteins *Bs*-CspB ($11,3 \text{ kJ mol}^{-1}$ bei pH 7,0 und $25 \text{ }^\circ\text{C}$) ist (Perl *et al.*, 1998). Die größte Änderung zeigt dabei der elektrostatische Beitrag zur Stabilität, welcher sich von $+7,9 \text{ kJ mol}^{-1}$ auf $-2,5 \text{ kJ mol}^{-1}$ verschlechtert.

Die Ergebnisse aus Tabelle 2 zeigen klar, dass Veränderungen in den lokalen Ladungsnetzwerken die Änderungen der Proteinstabilität eher erklären als die Nettoladung. Dies ist besonders offensichtlich für den Bereich, der die Positionen 3, 46, 48, 64 und 66 umfasst. In Tabelle 6 (Teilarbeit A) sind jeweils Varianten mit identischer lokaler Ladung für die fünf Positionen zusammengefasst. Es zeigte sich, dass fünf Varianten mit einer identischen lokalen Ladung von 0 um etwa 11 kJ mol^{-1} stabilisiert waren im Vergleich zum Wildtyp (lokale Ladung für diese Positionen hier -2). Dies war unabhängig von der exakten Verteilung der Ladung. Eine lokale Ladung von +1 bzw. +2 war jedoch mit $15,7 \text{ kJ mol}^{-1}$ bzw. $14,1 \text{ kJ mol}^{-1}$ noch günstiger. Der elektrostatische Anteil an dieser Stabilität war sogar bei einer lokalen Ladung von +2 mit $9,2 \text{ kJ mol}^{-1}$ am höchsten. Erst eine Variante mit einer lokalen Ladung von +3 war in etwa genauso stabil wie die Varianten mit lokaler Ladung 0. Dies zeigt, dass auch eine lokale Nettoladung ein sehr schwieriges Kriterium für die Vorhersage von Ladungseinflüssen ist. Es wird interessant zu sehen, ob Elektrostatikrechnungen all diese Effekte nachvollziehen können.

2.2 Korrelation zwischen Proteinstabilität und Dipolmoment

Um nicht nur die Nettoladung eines Proteins, sondern auch die räumliche Verteilung der Ladungen in die Vorhersage von elektrostatischen Effekten mit einzubeziehen, stellten Mozo-Villarias *et al.* (2003) ein sehr einfaches Kriterium auf. Sie berechneten ein Quasidipolmoment P zwischen dem Zentrum der negativen Ladungen und dem Zentrum der positiven Ladungen eines Proteins. Eine gleichmäßigere Verteilung der geladenen Gruppen (z.B. durch eine Mutation) sollte mit einer Erhöhung der Stabilität und einer Verringerung von P einhergehen. Umgekehrt sollte eine Erhöhung von P eine Destabilisierung des Proteins voraussagen. Sie testeten ihre These an homologen Paaren von jeweils einem mesophilen und einem thermophilen Protein sowie an Wildtypproteinen und in der Literatur beschriebenen Varianten davon. Sie fanden in der Tat eine inverse Korrelation zwischen Dipolmoment P und der Stabilität der untersuchten Proteine.

Für die beiden Kälteschockproteine *Bs-CspB* ($P = 9,7$ D) und *Bc-Csp* ($P = 6,6$ D) erhält man ebenfalls eine inverse Beziehung. Die Erniedrigung des Dipolmoments um 3,1 Debye geht einher mit einer Steigerung der Stabilität um $16,5 \text{ kJ mol}^{-1}$ (Tabelle I, Teilarbeit B). Für die beiden homologen Proteine nützten sie ihr Kriterium zur Vorhersage von Stabilitätsänderungen durch Ladungsaustausche. Sie berechneten die jeweilige Änderung für das Dipolmoment P , wenn man an jeweils einer Position der Proteinkette entweder eine positive oder eine negative Ladung einführt. Geladene Reste wurden durch eine entgegengesetzt geladene Aminosäure oder eine ungeladene ersetzt (Mozo-Villarias *et al.*, 2003).

Für die in den Selektionen gefundenen Austausche war keine signifikante Korrelation von Stabilität und Dipolmoment zu erkennen (Abbildung 5, Teilarbeit A). Um die Vorhersagekraft der Methode zu testen, wurden jeweils die vier Varianten von *Bc-Csp* mit der niedrigsten und die mit der höchsten berechneten Änderung des Dipolmoments erstellt und thermodynamisch charakterisiert (Abbildung 8). Für die Austausche wurde Glutamat gewählt, wenn eine negative Ladung erforderlich war, und Lysin als positiv geladener Ersatz. Eine komplette Aufstellung der thermodynamischen Parameter der Varianten findet sich in Tabelle I (Teilarbeit B).

Die Mutationen E12K, E36K, E50K und N55K bewirkten den größten Anstieg des Quasidipolmoments und sollten deshalb stark destabilisierend sein. Die Werte für P lagen im Bereich von 10,1 bis 12,5 D und somit höher als für das mesophile Homologe *Bs-CspB*. Die Varianten E21K, H29E, E46K und R56E lagen am anderen Ende der Skala in einem Bereich von 4,9 bis 2,0 D und wiesen das niedrigste Quasidipolmoment auf. Entsprechend der Vorhersage sollten sie stark stabilisierend sein. Alle untersuchten Positionen befinden sich auf der Oberfläche des Proteins und sind zugänglich für das Lösungsmittel (Abbildung 8).

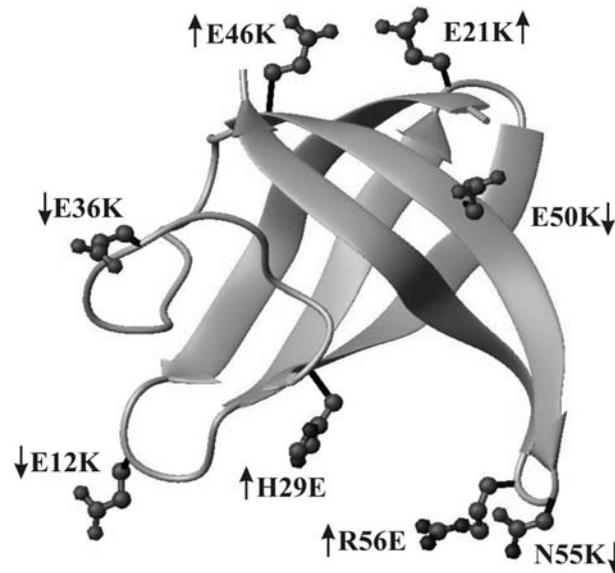


Abbildung 8. Tertiärstrukturdarstellung von Wildtyp *Bc-Csp*. Die Seitenketten der untersuchten Positionen sind als Kugel-Stab-Modell gezeigt. Mutationen, die als stabilisierend vorhergesagt sind, sind mit einem Pfeil nach oben gekennzeichnet und entsprechend destabilisierende mit einem Pfeil nach unten. Die Abbildung wurde erstellt unter Verwendung der Daten 1c9o (Mueller *et al.*, 2000) und dem Programm MOLMOL (Koradi *et al.*, 1996).

Die vier Varianten mit sehr hohem Dipolmoment zeigten eine nahezu unveränderte Stabilität. Die Mittelpunkte ihrer thermischen Übergänge lagen im Bereich von 73,5 bis 77,7 °C, was in etwa mit dem des Wildtyps *Bc-Csp* von 76,9 °C übereinstimmt und ungefähr 20 °C höher ist als der von *Bs-CspB*. Besonders auffällig war der Austausch N55K, der das Dipolmoment von 6,6 auf 10,1 D und gleichzeitig (entgegen der Vorhersage) auch die Stabilität erhöht. Ein ähnliches Bild ergab sich auch für die vier Varianten mit einem sehr niedrigen Dipolmoment. Der Mittelpunkt der thermischen Übergänge lag hier in einem Bereich von 73,5 bis 80,0 °C und somit ebenfalls nahezu unverändert im Vergleich zum Wildtypprotein. Einzig der Austausch R56E erhöhte entsprechend der Vorhersage die Stabilität des Proteins. Diese Variante besaß mit einem Wert von 2,0 das niedrigste Quasidipolmoment.

Trägt man den Mittelpunkt des thermischen Entfaltungsüberganges der Varianten gegen das Quasidipolmoment auf (Abbildung 9a), so erkennt man keinerlei Korrelation. Die Varianten decken dabei einen großen Bereich für das Dipolmoment von 2,0 bis 12,5 D ab. Gleiches gilt für die Freie Enthalpie der Entfaltung $\Delta\Delta G_D$ (Abbildung 9b). Einzig für die Wildtypproteine, das thermophile *Bc-Csp* und das mesophile *Bs-CspB*, kann man eine Korrelation erkennen. Eine Erhöhung des Dipolmoments von 6,6 auf 9,7 führt zu einer Destabilisierung von 16,5 kJ mol⁻¹. Eine Änderung der Stabilität, die durch eine Änderung des Dipolmomentes bedingt wird, sollte von Coulomb'scher Natur sein. Eine Auftragung des elektrostatischen Anteils von $\Delta\Delta G_D$ (Abbildung 9c), bestimmt als Differenz der Werte bei 0 M und 2 M NaCl, zeigt allerdings auch keine Korrelation.

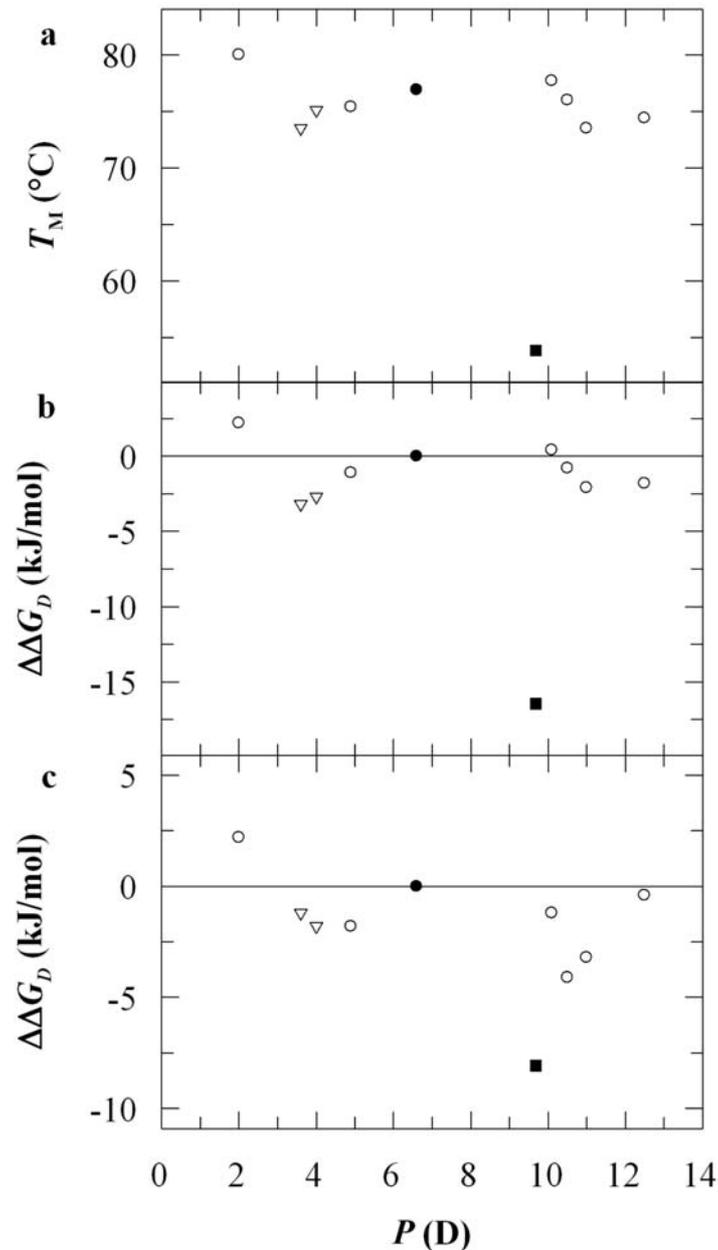


Abbildung 9. Korrelation zwischen Quasidipolmoment und dem experimentell gemessenen (a) Schmelzpunkt (T_M), (b) der gesamt Änderung der freien Enthalpie der Entfaltung relativ zum Wildtypprotein ($\Delta\Delta G_D$) und (c) dem elektrostatischen Anteil dieser Änderung bestimmt als Differenz der Messwerte bei 0 M und 2 M NaCl für Wildtyp *Bs-CspB* (■) und *Bc-Csp* (●), sowie die untersuchten Varianten gemessen unter Verwendung des Circular dichroismus (○) oder DSC („*differential scanning calorimetry*“) (▽). Die experimentellen Daten sind zusammengefasst in Tabelle 1 (Teilarbeit B).

Der Ansatz, das Quasidipolmoment zur Vorhersage von Stabilitätsänderungen bei Ladungsaustauschen zu nützen, ist für dieses Modellsystem ungeeignet. Die Mittelpunkte der thermischen Entfaltungsübergänge variieren in einem Bereich von ± 3 °C um den des Wildtypproteins. Die Positionen mit den größten Änderungen für das Dipolmoment liegen alle auf der Oberfläche des Proteins, meist in sehr exponierten Schleifenbereichen (Abbildung 8). Durch den großen Abstand zum Massenschwerpunkt des Proteins ist hier der Einfluss auf das Dipolmoment durch Variationen am größten. Die Reste 12, 36, 55 und 56

sind zu mehr als 60% zugänglich für das Lösungsmittel, was die höchsten Werte in *Bc-Csp* sind. Dies erklärt, warum Ladungsmutationen an diesen Positionen die Stabilität des Proteins kaum beeinflussen. Diese Reste sind zudem sehr beweglich, nicht in lokale Ladungsnetzwerke eingebunden und die Dielektrizitätskonstante ihrer (wässrigen) Umgebung ist hoch. Deshalb beeinflussen sie die elektrostatischen Wechselwirkungen kaum (Lazaridis und Karplus, 2003; Kumar und Nussinov, 2002). Ein gutes Beispiel stellen die Positionen 55 und 56 dar. Beide liegen an der Spitze einer langen exponierten β Schleife (Abbildung 8). N55K erhöht das Dipolmoment auf 10,1 D, R56E senkt es auf 2,0 D. Beide Mutationen wirken jedoch im Experiment geringfügig stabilisierend. Insgesamt gesehen konnte keine Korrelation zwischen dem Quasidipolmoment und der Proteinstabilität gefunden werden, deshalb ist dieses Kriterium zur Vorhersage von elektrostatischen Effekten ungeeignet.

2.3 *In vitro* Evolution im Vergleich zu rechnergestütztem *protein design*

2.3.1 *In vitro* Evolution der *boundary positions* von G β 1

Zur Stabilisierung von Proteinen werden momentan hauptsächlich zwei Methoden genutzt, *computational design* und *in vitro* Evolution. In der Gruppe von Stephen Mayo wurde die Domäne β 1 des Streptokokkenproteins G unter Verwendung des rechnergestützten Protein-design optimiert (Malakauskas und Mayo, 1998; Dahiyat und Mayo, 1997b). In einem ersten Schritt wurden drei Positionen im Inneren des Proteins untersucht, doch die Austausche (Y3F/L7I/V39I) erwiesen sich nur als geringfügig stabilisierend. Anschließend wurden vier teilweise exponierte („*boundary*“) Reste (16, 18, 25 und 29) betrachtet, welche in den Berechnung zehn nahezu gleich stabile Sequenzen lieferten. Die beste Variante T16I/T18I/T25E/V29I (bezeichnet als G β 1-c3b4) wurde auch experimentell charakterisiert und wies eine deutliche Stabilisierung von 15,9 kJ mol⁻¹ auf (Malakauskas und Mayo, 1998).

Im Folgenden sollten für diese vier *boundary* Positionen mit Hilfe des *in vitro* Evolutionssystems *Proside* optimal stabilisierende Aminosäurekombinationen selektiert werden und die Ergebnisse beider Methoden verglichen werden. Es wurden fünf unabhängige Bibliotheken von G β 1 angelegt, in denen die Positionen 16, 18, 25 und 29 randomisiert waren. An diesen Positionen waren alle 20 natürlichen Aminosäuren zugelassen. G β 1 ist bereits in seiner Wildtypform thermodynamisch sehr stabil. Um eine Selektion zu ermöglichen fehlten die drei *core* Mutationen (Y3F, L7I, V39I) und es wurden zusätzlich zwei destabilisierende Mutationen an den Positionen 6 und 53 eingeführt (Smith *et al.*, 1994; Smith und Regan, 1995; Merkel *et al.*, 1999). Diese beiden Positionen befinden sich auf der entgegen gesetzten Seite von G β 1 und sollten deshalb die vier *boundary* Positionen nicht beeinflussen. Die Mutation T53A wurde in den Bibliotheken 1-3 kombiniert mit I6A und in den Bibliotheken 4-5 mit I6V.

Mit jeder der Bibliotheken wurden bis zu drei unabhängige Selektionen durchgeführt. Die Starttemperatur betrug dabei im Allgemeinen 40 °C und wurde innerhalb von ungefähr zehn Selektionsrunden schrittweise auf 57,5 °C erhöht. Bei der Selektion wurde zusätzlich die Konzentration an Chymotrypsin von 0,25 μ M auf 25 μ M gesteigert. Im Verlauf der Selektion wurden ungefähr 100 Phagenvarianten sequenziert. Die gefundenen Codons und die entsprechenden Aminosäurereste an den Positionen 16, 18, 25 und 29 sind zusammengefasst in Tabelle 1 (Teilarbeit C).

Der Verlauf der Selektionen war vergleichbar und bestimmte Arten von Aminosäuren wurden bereits sehr früh in der Selektion ausgeschlossen. Es wurden keine polaren oder geladenen Reste an Position 16, 18 oder 29 gefunden und an Position 25 keine positiv geladenen oder aromatischen Aminosäuren. Nach diesem anfänglichen Ausschluss blieb die Diversität der Bibliotheken über viele Selektionsrunden sehr groß. Letztendlich wurden an Position 16 hauptsächlich aliphatische Reste (Ile, Leu und Val) identifiziert, an Position 18 aromatische (Phe, Tyr) oder aliphatische Reste (Ile, Val), aber interessanterweise nie Leu. An Position 25 wurden am häufigsten Glu, Ser und Thr selektiert und die Position 29 war fast ausschließlich durch aromatische Aminosäuren besetzt. An drei der randomisierten Positionen wurden aromatische oder große aliphatische Aminosäuren angereicht. Dies ist bemerkenswert, denn diese Reste stellen für Chymotrypsin, welches zur Selektion genutzt wurde, das Substrat dar. Dies bestätigt, dass die Selektion durch thermodynamische Stabilität bedingt wird und nicht durch den Verlust an Spaltstellen für die Protease.

Für 21 Phagen mit einer hohen Proteaseresistenz wurden die gefundenen Aminosäurereste in eine Variante von G β 1 eingeführt, welche die drei stabilisierenden *core* Mutationen Y3F, L7I und V39I beinhaltet. Dies ermöglichte einen direkten Vergleich der Varianten mit denen aus dem rechnergestützten *protein design*, welche diese Mutationen ebenfalls enthielten. Die entsprechenden Proteinvarianten wurden gereinigt und thermodynamisch charakterisiert. Das Referenzprotein mit nur den drei *core* Mutationen G β 1-c3 zeigt bei pH 5,5 einen Schmelzpunkt von 91 °C (Malakauskas und Mayo, 1998). Die meisten der selektierten Varianten ebenso wie G β 1-c3b4 zeigten unter diesen Bedingungen einen Übergangsmittelpunkt von über 100 °C. Deshalb wurden die Varianten in Gegenwart von 2,5 M GdmCl bei pH 7,0 verglichen. Unter diesen Bedingungen zeigten G β 1-c3 und G β 1-c3b4 einen Übergangsmittelpunkt von 49,3 °C bzw. 72,1 °C (Abbildung 10).

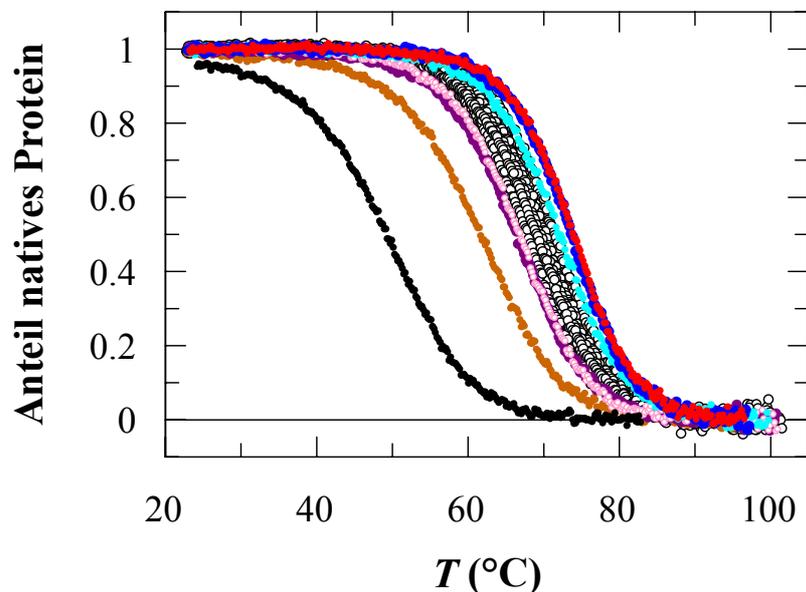


Abbildung 10. Thermische Entfaltungsgänge der selektierten Varianten. (●), (●) Die beiden stabilsten Varianten 1 bzw. 2 aus den Selektionen. (○) Varianten 3-10 (für die Nummerierung vergleiche Tabelle 2, Teilarbeit C). (●) G β 1-c3, (●) G β 1-c3b4, (●) R1, (●) R3, (●) R4. Die Übergänge von 4 μ M Protein in 100 mM Na-Cacodylat/HCl (pH 7,0) wurden verfolgt durch die CD-Änderung bei 222,6 nm in Gegenwart von 2,5 M GdmCl. Dargestellt ist der Anteil an nativem Protein als Funktion der Temperatur nach Analyse der Übergänge entsprechend einem Zweizustandsmodell. Die erhaltenen thermodynamischen Parameter sind teilweise zusammengefasst in Tabelle 3 und eine komplette Aufstellung findet sich in Tabelle 2 (Teilarbeit C).

Alle 21 gereinigten Varianten waren thermodynamisch deutlich stabilisiert. Eine Aufstellung der Daten aller Varianten findet sich in Tabelle 2 (Teilarbeit C). Die Mittelpunkte der thermisch induzierten Entfaltungsübergänge waren um 15-25 °C erhöht, was einer Stabilisierung von 9-15 kJ mol⁻¹ entsprach. Für die beiden besten Varianten betrug die Stabilisierung 14,9 bzw. 14,8 kJ mol⁻¹ (Varianten 1 und 2, Tabelle 3) und war somit höher als für die beste Variante aus dem *computational design* ($\Delta\Delta G_D = 13,6$ kJ mol⁻¹). Für einige stark stabilisierte Varianten wurden zusätzlich chemisch induzierte Entfaltungsübergänge gemessen (Abbildung 3a, Teilarbeit C). Die Mittelpunkte der Übergänge ($[G_{dmCl}]_m$, Tabelle 3, Teilarbeit C) wiesen die gleiche Rangordnung wie die der thermisch induzierten Entfaltungsübergänge auf. Für diese Varianten wurden auch Faltungskinetiken gemessen, die zeigten, dass die Stabilisierung hauptsächlich durch eine Verlangsamung der Entfaltung bedingt wird (Abbildung 3b, Teilarbeit C). Dies bedeutet, dass die meisten der stabilisierenden Wechselwirkungen der Reste an Position 16, 18, 25 und 29 erst spät im Faltungsprozess ausgebildet werden, d.h. nach Durchschreiten des Übergangszustandes. Dies stimmt sehr gut mit Analysen zur Aufklärung des Faltungsmechanismus von G β 1 überein (Nauli *et al.*, 2001; McCallister *et al.*, 2000).

Tabelle 3. Thermodynamische Stabilitätsdaten für Varianten von G β 1 zum Vergleich von *in vitro* Evolution und *computational design*

Variante	Rest an Position				T_M^a (°C)	ΔH_D^b (kJ mol ⁻¹)	$\Delta G_D(70\text{ °C})^c$ (kJ mol ⁻¹)	$\Delta\Delta G_D^d$ (kJ mol ⁻¹)
	16	18	25	29				
1	I	I	E	F	74,0	240	2,6	14,9
2	I	I	E	Y	73,7	241	2,5	14,8
3	I	I	T	W	71,5	232	1,0	13,3
4	L	I	E	Y	71,2	228	0,8	13,1
16	L	V	T	F	67,0	206	-1,9	10,4
G β 1-c3b4	I	I	E	I	72,1	232	1,3	13,6
G β 1-c3	T	T	T	V	49,3	151	-12,3	0
R1	I	L	E	F	66,6	201	-2,1	10,2
R2	I	I	Q	F	71,7	231	1,1	13,4
R3	I	I	E	K	66,8	210	-2,0	10,3
R4	I	L	E	K	61,7	172	-4,7	7,6

^a T_M ist der Mittelpunkt des thermischen Entfaltungsüberganges, ^b ΔH_D ist die Enthalpie der Entfaltungsreaktion am Mittelpunkt, ^c ΔG_D ist die Änderung der Gibbs'schen freien Energie der Entfaltung bei 70 °C, ^d $\Delta\Delta G_D$ ist die Änderung der Gibbs'schen freien Energie der Entfaltung bei 70 °C relativ zu der des Referenzproteins G β 1-c3. Die thermodynamischen Parameter wurden aus thermischen Entfaltungsübergängen, wie in Abbildung 10 gezeigt, ermittelt. Die Varianten sind nummeriert gemäß sinkender Stabilität. Die Varianten R1-R4 wurden erstellt, um die Unterschiede zwischen *computational design* und *in vitro* Evolution zu erklären. Eine Aufstellung für alle in dieser Arbeit charakterisierten Varianten findet sich in Tabelle 2 (Teilarbeit C).

2.3.2 Vergleich von rechnergestützter und evolutiver Stabilisierung

Die rechnergestützte Optimierung von G β 1 ergab zehn Varianten. Deren berechnete Stabilitäten lagen in einem Bereich von -444,6 bis -446,6 kcal mol⁻¹ und waren somit etwa 28 kcal mol⁻¹ stabiler als das Referenzprotein G β 1-c3 (Malakauskas und Mayo, 1998). Die beste Variante G β 1-c3b4 ist dabei im Experiment geringfügig weniger stabil als die beiden

besten Varianten aus den Selektionen (Tabelle 3). Keine der berechneten Varianten wurde in den Selektionen gefunden. Um beide Methoden miteinander vergleichen zu können, sollen jeweils die zehn besten Varianten herangezogen werden. Eine Aufstellung der gefundenen Aminosäurereste an den jeweiligen Positionen für beide Methoden, ist in Tabelle 4 zu finden. Die Variante 2 (Tabelle 3) aus der *in-vitro* Evolution entspricht dabei der Konsensus-Sequenz der Selektion, was darauf hindeutet, dass die Effekte der einzelnen Reste in erster Näherung additiv sind.

Tabelle 4. Vergleich der zehn stabilsten Varianten aus der rechnergestützten Stabilisierung und den *Proside* Selektionen

	Auftreten der Reste an Position			
	16	18	25	29
Rechnergestützte Stabilisierung	9 × Ile 1 × Tyr	6 × Leu 4 × Ile	6 × Glu 4 × Gln	5 × Ile 3 × Lys 2 × Tyr
<i>Proside</i> Selektionen	6 × Ile 4 × Leu	7 × Ile 1 × Val 1 × Phe 1 × Tyr	5 × Glu 4 × Ser 1 × Thr	5 × Tyr 3 × Trp 2 × Phe

An Position 16 trat in keiner der berechneten Varianten ein Leu auf, aber vier der selektierten Varianten zeigten diesen Rest. Vergleicht man Variante 2 (mit Ile16) und Variante 4 (mit Leu16, Tabelle 3), so zeigt sich, dass Leu annähernd gleich stabilisierend wirkt wie Ile an dieser Position. Dies erklärt, warum diese beiden Reste in den Selektionen identifiziert wurden.

An Position 18 findet man in den Berechnungen am häufigsten Leu, aber Leu18 trat nie bei der *in vitro* Evolution auf. Dabei wird Leu durch sechs verschiedene Codons codiert und sollte deshalb in den Bibliotheken überrepräsentiert sein. Um die Ursache für diesen Unterschied zu erschließen, wurde Leu18 ortsgerichtet in die beste selektierte Variante eingeführt. Dies erniedrigte den Mittelpunkt der thermischen Entfaltung um 7,6 °C (vergleiche Variante 1 und R1, Tabelle 3) und erklärt, warum dieser Rest nie selektiert wurde.

An Position 25 wurde durch beide Methoden am häufigsten Glu als bester Ersatz identifiziert, gefolgt von Ser und Thr in den Selektionen, sowie Gln in den Berechnungen. Um den Einfluss von Gln zu ergründen, wurde in der stabilsten Variante Glu25 zu Gln verändert. Diese Mutation war geringfügig destabilisierend um 2,3 °C (vergleiche Variante 1 und R2, Tabelle 3).

An Position 29 wurde in den Berechnungen am häufigsten Ile gefunden, in den Selektionen hingegen ausschließlich aromatische Aminosäuren (Tabelle 4). Vergleicht man die Varianten 1 und 2 mit Gβ1-c3b4 (Tabelle 3), so zeigt sich, dass Phe29 und Tyr29 tatsächlich stärker stabilisierend wirken als Ile an dieser Position. Drei Varianten des *computational design* besaßen Lys29, welches nie selektiert wurde. Um auch diese Diskrepanz aufzuklären, wurde Lys29 in die beste Variante eingeführt. Diese Mutation wirkte stark destabilisierend und senkte den Mittelpunkt des thermischen Überganges um 7,2 °C (vergleiche Variante 1 und R3, Tabelle 3).

Kombiniert man die beiden Mutationen I18L und F29K in der stabilsten Variante, so erhält man eine Destabilisierung um 12,3 °C oder entsprechend 7,3 kJ mol⁻¹ (vergleiche Variante 1 und R4, Tabelle 3). Die entsprechende Variante R4 ist identisch mit Gβ1-mc4 (Malakauskas und Mayo, 1998), welche auf dem vierten Platz des *computational design* liegt und laut Berechnungen nahezu genauso stabil sein sollte wie die beste Variante Gβ1-c3b4. Tatsächlich ist sie jedoch signifikant destabilisiert und instabiler als alle 21 Varianten aus den Selektionen. Sieben der zehn berechneten Sequenzen beinhalten mindestens einen der beiden sehr ungünstigen Reste Leu18 bzw. Lys29 und die Einführung allein einer der beiden Mutationen senkt den Schmelzpunkt unter 67 °C, was in den Selektionen nur Platz 16 entsprechen würde (vergleiche R1 bzw. R3 und Variante 16, Tabelle 3). Dies zeigt, dass die Mehrheit der Varianten aus dem *computational design* erheblich weniger stabil sind als die meisten Varianten aus der *in vitro* Evolution.

Zusammenfassend kann man sagen, dass das rechnergestützte *protein design* in der Lage ist interessante Positionen mit einem Stabilisierungspotential zu identifizieren. Es bleibt aber schwierig, exakte Beiträge zur Stabilisierung durch einzelne Aminosäurereste zu berechnen. Vor allem ist es kaum möglich, Effekte, die durch Änderung des nativen Rückgrats des Proteins (Kuhlman *et al.*, 2003; Desjarlais und Handel, 1999) oder im entfalteten Zustand (Pace *et al.*, 2000; Kuhlman *et al.*, 1999) auftreten, in die Berechnungen einfließen zu lassen. *In vitro* Evolution ist dagegen eingeschränkt in der Bibliotheksgröße und deshalb können nur wenige Reste gleichzeitig randomisiert werden. Daher ergänzen sich die beiden Methoden hervorragend (Voigt *et al.*, 2001; Arnold, 2001). Das *computational design* kann zur Identifikation von viel versprechenden Positionen eingesetzt werden. Diese Positionen können einer Sättigungsmutagenese unterworfen werden und die entsprechenden Bibliotheken durch *in vitro* Evolution untersucht werden, um den besten Aminosäureersatz zu identifizieren.

2.3.3 Strukturelle und thermodynamische Analyse der Unterschiede zwischen *computational design* und *in vitro* Evolution

Um einen besseren Einblick in die Prinzipien der Proteinstabilität zu erhalten und um die Unterschiede der durch *in vitro* Evolution und *computational design* erhaltenen Varianten zu erklären, wurden von der stabilsten selektierten und der Variante R4 (mit den ungünstigen Aminosäureresten Leu18 und Lys29, Tabelle 3) Kristallstrukturen bestimmt (Abbildung 11). Im Folgenden werden die beiden Varianten als Gβ1c3-IIEF und Gβ1c3-ILEK bzw. in Abwesenheit der drei *core* Mutationen als Gβ1-IIEF und Gβ1-ILEK bezeichnet. Dabei stehen die letzten vier Buchstaben für die jeweiligen Aminosäuren an Position 16, 18, 25 und 29. Die Parameter der Kristallstrukturen für Gβ1c3-IIEF und Gβ1c3-ILEK sind in Tabelle 1, Teilarbeit D zusammengefasst. Zusätzlich wurden, um das Verständnis zu verbessern, die in beiden Varianten aufgetretenen Mutationen einzeln thermodynamisch charakterisiert und die Wechselwirkungen der Reste durch Doppelmutationszyklen (Horovitz und Fersht, 1992; Horovitz, 1996) bestimmt. Dabei wurden die thermodynamischen Effekte der einzelnen Mutationen in Abwesenheit der drei stabilisierenden Variationen im Inneren des Proteins analysiert. Dies ermöglichte die Messung von thermisch induzierten Entfaltungsübergängen ohne den Zusatz von GdmCl. Durch den Verzicht auf das chaotrope Salz GdmCl bei den Messungen war es außerdem möglich, die Art der stabilisierenden Effekte, ob hydrophobe oder elektrostatische Wechselwirkungen, zu bestimmen.

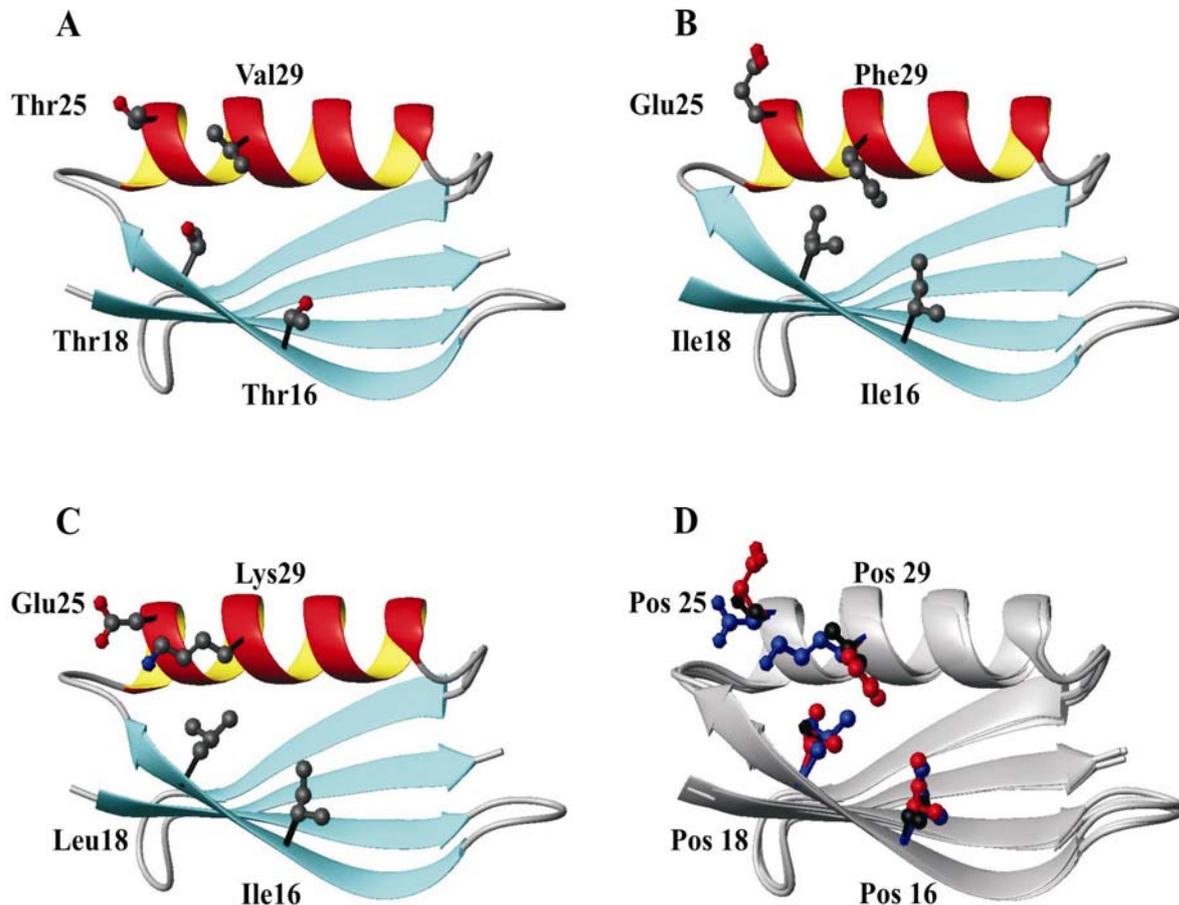


Abbildung 11. Tertiärstrukturdarstellung von A) Wildtyp Gβ1, B) Gβ1c3-IIEF, C) Gβ1c3-ILEK und D) Überlagerung von Wildtyp Gβ1, Gβ1c3-IIEF und Gβ1c3-ILEK. Die Seitenketten der Positionen 16, 18, 25 und 29 sind als Kugel-Stab-Modell gezeigt. In Teilabbildung D sind die Seitenketten von Wildtyp Gβ1 schwarz, von Gβ1c3-IIEF rot und von Gβ1c3-ILEK blau dargestellt. Die Abbildung wurde erstellt unter Verwendung der Daten 1pgb (Gallagher *et al.*, 1994) für Wildtyp Gβ1, den in Teilarbeit D gewonnenen Datensätzen für Gβ1c3-IIEF und Gβ1c3-ILEK und dem Programm MOLMOL (Koradi *et al.*, 1996).

Die T_M -Werte von Gβ1-IIEF und Gβ1-ILEK sind um 17,3 und 10,2 °C im Vergleich zum Wildtypprotein erhöht (Tabelle 2, Teilarbeit D). Dies entspricht einer Stabilisierung von 14,2 bzw. 8,1 kJ mol⁻¹, was sehr gut mit den Werten in Gegenwart der drei *core* Mutationen übereinstimmt (Tabelle 3). Um die einzelnen Beiträge der Mutationen an den vier Positionen zu bestimmen wurden die Einzelvarianten erstellt und deren thermodynamische Stabilität in Abwesenheit und in Gegenwart von 2 M NaCl bestimmt (Tabelle 2, Teilarbeit D). Der durch Salz abschirmbare Anteil an ΔG_D war dabei im Allgemeinen geringer als 0,8 kJ mol⁻¹ ist. Dies legt nahe, dass elektrostatische Effekte nur in geringem Ausmaß an den beobachteten Stabilitätsänderungen beteiligt sind.

Positionen 16 und 25

Ile16 und Glu25 wurden durch beide Methoden, *computational design* und *in vitro* Evolution, am häufigsten als der jeweils beste Ersatz identifiziert. Als einzelne Mutation stabilisiert T25E das Wildtypprotein um 2,4 °C bzw. 2,0 kJ mol⁻¹ (Tabelle 2, Teilarbeit D). Dabei stammt

ein kleiner Anteil von $0,8 \text{ kJ mol}^{-1}$ aus durch Salz abschirmbaren elektrostatischen Wechselwirkungen. Die Position 25 ist im N-terminalen Bereich der α -Helix lokalisiert. Die Stabilisierung durch die Mutation T25E wird wahrscheinlich teilweise durch vorteilhafte Wechselwirkungen von Glu25 mit dem Dipol der Helix bedingt (Serrano und Fersht, 1989). Die Orientierung der Seitenkette von Glu25 hängt dabei davon ab, ob die Position 29 durch Phe oder Lys besetzt ist. Im Falle von Lys an Position 29 zeigt die Seitenkette von Glu25 in die Richtung dieses Restes und der Abstand beträgt $4,3 \text{ \AA}$, was eine Salzbrücke zwischen diesen Resten nahe legt. In Gegenwart von Phe29 ist Glu25 stärker lösungsmittlexponiert und die Carboxylgruppe bildet eine Salzbrücke mit der Aminogruppe von Lys28.

Die Mutation T16I zeigt den größten Stabilisierungseffekt von allen Einzelvarianten mit einer Erhöhung des T_M -Wertes um $6,7 \text{ }^\circ\text{C}$ bzw. einer Stabilisierung von $5,6 \text{ kJ mol}^{-1}$ (Tabelle 2, Teilarbeit D). Dabei ist die Stabilisierung unabhängig von der Konzentration an NaCl, was bedeutet, dass sie hauptsächlich durch verbesserte hydrophobe Wechselwirkungen von Ile16 bedingt wird. Die Position 16 ist in der Mitte von Strang β_2 des Faltblattes lokalisiert. Die Stabilisierung durch die Mutation T16I kann dabei nicht durch eine erleichterte β -Faltblattbildung erklärt werden, da Thr und Ile eine nahezu gleiche Tendenz zur Faltblattbildung (*β sheet propensity*) haben (Smith *et al.*, 1994; Smith und Regan, 1997). Die Position liegt außerdem in lokaler Nachbarschaft zu den aromatischen Seitenketten der Reste Phe30 und Tyr33. Im Wildtypprotein bildet Thr16 dabei Kontaktflächen (Sobolev *et al.*, 1999) von 6 \AA^2 bzw. 35 \AA^2 zu diesen Resten. Für Ile16 in der Variante G β 1c3-IIEF sind diese Werte auf 23 \AA^2 für die Kontaktfläche mit Phe30 und 54 \AA^2 mit Tyr33 erhöht. Außerdem ist Ile16 in der Variante G β 1c3-IIEF im Vergleich zu Thr16 im Wildtypprotein stärker vom Lösungsmittel abgeschirmt. Der vom Lösungsmittel abgeschirmte Oberflächenanteil steigt dabei von $82,2 \%$ auf $86,4 \%$ oder in absoluten Zahlen von 202 \AA^2 für Thr16 auf 239 \AA^2 für Ile16. Dies zusammen genommen erklärt die durch verbesserte hydrophobe Wechselwirkungen bedingte starke Stabilisierung durch die Mutation T16I.

Positionen 18 und 29

Die größten Unterschiede zwischen den zehn berechneten und den zehn besten selektierten Varianten wurden für die Positionen 18 und 29 beobachtet (Tabelle 4). Die Position 18 war bei sechs von den zehn berechneten Varianten durch Leu besetzt, welches jedoch bei keiner der selektierten Varianten gefunden wurde. Durch *in vitro* Evolution wurde an dieser Position Ile als bester Ersatz identifiziert. Als einzelne Mutation stabilisiert T18I das Wildtypprotein um $3,8 \text{ kJ mol}^{-1}$, T18L hingegen nur um $1,4 \text{ kJ mol}^{-1}$ (Tabelle 2, Teilarbeit D). In beiden Fällen ist die Stabilisierung unabhängig von der Salzkonzentration, was mit den Erwartungen übereinstimmt, dass diese Mutationen das Protein durch verbesserte hydrophobe Wechselwirkungen stabilisieren. Dies drückt auch die im Proteininneren verborgene Oberfläche der Reste aus. Diese beträgt für Thr18 im Wildtypprotein 215 \AA^2 und wächst für Leu18 auf 239 \AA^2 und für Ile18 sogar auf 252 \AA^2 an. Dies ist auch ein Grund für die geringere Stabilisierung von Leu18 im Vergleich zu Ile. Eine weitere Erklärung ist die deutlich höhere Tendenz zur Ausbildung von β -Faltblattsträngen von Ile im Vergleich zu Leu (Smith *et al.*, 1994; Smith und Regan, 1997).

Tauscht man Ile18 in der stabilsten selektierten Variante gegen ein Leu aus, so erhält man eine Destabilisierung um $4,7 \text{ kJ mol}^{-1}$ (vergleiche Variante 1 und R1, Tabelle 3). Dies entspricht in etwa dem doppelten Wert der Einzelmutationen im Wildtypprotein und zeigt,

dass die Beiträge zur Stabilität von Ile und Leu an Position 18 geringfügig umgebungsabhängig sind.

An Position 29 wurden bei der Selektion ausschließlich aromatische Reste gefunden und in den Berechnungen hauptsächlich Ile, welches geringfügig schwächer stabilisierend als aromatische Reste ist (vergleiche G β 1-c3b4 mit Variante 1 und 2, Tabelle 3). In drei der berechneten Varianten wurde jedoch Lys29 als Ersatz identifiziert, welches bei den *in vitro* Selektionen nie gefunden wurde. Die Einzelmutation V29F stabilisiert das Wildtypprotein um 1,7 kJ mol⁻¹, wohingegen V29K das Protein um 2,2 kJ mol⁻¹ destabilisiert (Tabelle 2, Teilarbeit D). Der Effekt von Phe29 ist unabhängig von der Salzkonzentration, die Destabilisierung durch Lys29 ist jedoch in Gegenwart von 2 M NaCl auf 1,5 kJ mol⁻¹ verringert. Dies bedeutet, dass ungefähr ein Drittel des destabilisierenden Effektes von Lys29 auf ungünstigen elektrostatischen Wechselwirkungen dieses Restes beruht. Phe29 und Lys29 unterscheiden sich stark in der Orientierung ihrer Seitenketten. Lys29 deutet in der Variante G β 1c3-ILEK wie bereits erwähnt in Richtung von Glu25 und der Abstand von 4,3 Å legt die Bildung einer Salzbrücke nahe. In Gegenwart von Lys29 wird jedoch die Wechselwirkung von Glu25 mit Lys28 unterbunden und die Seitenkette von Glu25 zum Proteininneren gedreht, was aufgrund der höheren Desolvatisierung ungünstig ist. Phe29 deutet hingegen in Richtung von Ile18 und Tyr33 und bildet mit beiden eine Kontaktfläche von jeweils ungefähr 47 Å².

Wechselwirkungen zwischen den Positionen 16, 18 und 29

Die Positionen 16, 18, 25 und 29 sind in räumlicher Nähe und auf der gleichen Seite von G β 1 lokalisiert (Abbildung 11). Um paarweise Wechselwirkungen zwischen diesen Resten zu identifizieren, wurden Doppelmutationszyklen (Horovitz und Fersht, 1992; Horovitz, 1996) für sowohl die berechneten, als auch die selektierten Reste durchgeführt. Die beiden am nächsten benachbarten Positionen sind 18 und 29 mit einem Abstand der C β -Atome im Wildtypprotein von 5,6 Å. In der selektierten Variante sind diese beiden Positionen durch Ile18 und Phe29 besetzt. Die Doppelmutante T18I/V29F ist um 5,5 kJ mol⁻¹ stabilisiert, was genau der Summe der Effekte der beiden Einzelmutationen entspricht (Abbildung 4, Teilarbeit D). Auch in der berechneten Variante ergab sich nur eine unbedeutende Wechselwirkung von 0,2 kJ mol⁻¹ zwischen Leu18 und Lys29 (Abbildung 4, Teilarbeit D), was bedeutet, dass die Positionen 18 und 29 in beiden Fällen als unabhängig zu betrachten sind. Die Varianten von G β 1 unterscheiden sich stark in der Kontaktfläche zwischen den Resten 18 und 29. Im Wildtypprotein beträgt diese nur 17 Å². Sie steigt für die berechnete Variante (mit Leu18 und Lys29) auf 40 Å² und für die selektierte Variante (mit Ile18 und Phe29) sogar auf 61 Å². Interessanterweise konnte experimentell weder zwischen Leu18 und Lys29 noch zwischen Ile18 und Phe29 eine beachtenswerte Wechselwirkung festgestellt werden. Dies zeigt die Schwierigkeit Kontakte in Raumstrukturen in thermodynamische Effekte zu übersetzen und umgekehrt.

Als weiteres wurden die Positionen 16 und 18 auf Wechselwirkungen untersucht. Zwischen diesen Positionen beträgt der Abstand der C β -Atome im Wildtypprotein 6,8 Å. In der selektierten Variante sind beide Positionen durch Ile besetzt und die Doppelmutante T16I/T18I ist um 13,0 °C bzw. 10,3 kJ mol⁻¹ stabilisiert (Tabelle 2, Teilarbeit D). Die Summe der Stabilitätsbeiträge der Einzelmutationen T16I und T18I beträgt 9,4 kJ mol⁻¹, was bedeutet, dass in diesem Fall eine zusätzliche stabilisierende Wechselwirkung von 0,9 kJ mol⁻¹

zwischen Ile16 und Ile18 auftritt (Abbildung 4, Teilarbeit D). Der Doppelmutationszyklus für die Reste Ile16 und Leu18 ergab eine nahezu gleiche Wechselwirkung von $1,1 \text{ kJ mol}^{-1}$ (Abbildung 4, Teilarbeit D).

Addiert man die Beiträge der einzelnen Mutationen und die durch die Mutationszyklen gefundenen Wechselwirkungen, so ergibt sich für G β 1-IIIEF ein Wert von $14,0 \text{ kJ mol}^{-1}$ und für G β 1-ILEK von $8,1 \text{ kJ mol}^{-1}$. Diese Werte stimmen sehr gut mit den gemessenen Werten für die beiden Vierfachvarianten von $14,2 \text{ kJ mol}^{-1}$ und $8,1 \text{ kJ mol}^{-1}$ überein (Tabelle 2, Teilarbeit D). Dies bedeutet, dass lediglich zwischen den Positionen 16 und 18 eine schwache Wechselwirkung stattfindet und ansonsten die Reste 16, 18, 25 und 29 als unabhängig betrachtet werden können.

2.4 Experimentelle Optimierung von G β 1

2.4.1 Identifizierung von aussichtsreichen Positionen durch fehlerhafte PCR

In der vorangegangenen Stabilisierung von G β 1 wurde ein gemischtes Verfahren gewählt. Die Identifikation der Positionen erfolgte unter Verwendung von *computational design* und die besten Aminosäureaustausche an diesen Positionen wurden durch *in-vitro* Evolution gefunden. Dadurch konnte das Problem der beschränkten Bibliotheksgrößen bei Selektionen umgangen werden. Im nächsten Ansatz wurde für beide Schritte auf dem Weg zur Stabilisierung, die Identifikation von Positionen und das Finden der besten Ersatz, ein experimentelles Vorgehen gewählt. Zur Identifikation der Positionen wurde dabei fehlerhafte PCR und *in vitro* Evolution verwendet und die dabei gefundenen, viel versprechenden Positionen wurden in einem zweiten Schritt einer Sättigungsmutagenese mit anschließender Selektion der entsprechenden Bibliotheken unterworfen.

Ausgehend von G β 1, welches, um eine Selektion zu ermöglichen, die bereits beschriebenen zwei destabilisierenden Mutationen I6V und T53A enthielt, wurden fünf unabhängige Bibliotheken unter Verwendung von fehlerhafter PCR erstellt. Diese Bibliotheken wurden bis zu 17 Selektionsrunden unterworfen. Die Startbedingungen betragen dabei $40 \text{ }^\circ\text{C}$ und $0,25 \text{ }\mu\text{M}$ Chymotrypsin, sie wurden schrittweise auf $55 \text{ }^\circ\text{C}$ und $2,5 \text{ }\mu\text{M}$ erhöht. Zu verschiedenen Zeitpunkten der Selektionen wurde der G β 1-Anteil von Phagen mit erhöhter Proteaseresistenz sequenziert. Eine Aufstellung aller gefundenen Aminosäurereste und deren zugehörige Codons ist in Tabelle 1 (Teilarbeit E) zu finden. Insgesamt wurden Mutationen an elf verschiedenen Positionen (1, 8, 9, 11, 13, 15, 16, 18, 24, 37 und 56), verteilt über das gesamte Protein, identifiziert (Abbildung 1, Teilarbeit E). Mit Ausnahme der Mutationen G9R und N37M stellten alle anderen Ein-Basen-Austausche dar. Dabei handelte es sich fast ausschließlich um C \rightarrow T Transversionen und A \rightarrow T Transitionen, welche bevorzugt bei fehlerhafter PCR auftreten (Firth und Patrick, 2005; Drummond *et al.*, 2005).

Um den Einfluss der gefundenen Mutationen auf die Stabilität zu bestimmen, wurden diese in das Wildtypprotein G β 1 übertragen, d.h. ohne die beiden destabilisierenden Mutationen, die genutzt wurden um die Selektion zu ermöglichen. Die entsprechenden Varianten wurden gereinigt und thermodynamisch charakterisiert (Tabelle 1, Teilarbeit E). Das am häufigsten selektierte Motiv N8Y/G9V wirkte dabei destabilisierend im Wildtypprotein und senkte den Übergangsmittelpunkt von $77,0 \text{ }^\circ\text{C}$ auf $71,2 \text{ }^\circ\text{C}$. Der Grund für dieses unerwartete Ergebnis wurde klar, als die beiden Mutationen N8Y/G9V in die für die Selektionen verwendete Variante von G β 1 eingeführt wurden. Diese Variante, welche die beiden destabilisierenden

Mutationen I6V und T53A enthielt, wurde durch das selektierte Motiv um 4,2 °C stabilisiert. Dies bestätigt die korrekte Funktionsweise des Selektionssystems *Proside*. Die Positionen 8 und 9 sind in lokaler Nachbarschaft zu den für die Selektion verwendeten, destabilisierenden Mutationen an den Positionen 6 und 53. Im Wildtypprotein ist das C^β-Atom von Asn8 nur 6,3 bzw. 6,7 Å von den C^β-Atomen der Reste 6 und 53 entfernt. Es ist möglich, dass die aromatische Seitenkette von Tyr8 die durch die destabilisierenden Mutationen geschaffene „Lücke“ schließt. Deshalb wirkt das selektierte Motiv N8Y/G9V in Gegenwart der Mutationen I6V und T53A stabilisierend, im Wildtypprotein aufgrund von sterischen Hindernissen hingegen destabilisierend.

Die in Bibliothek E2 gefundenen Mutationen T16I und T18I wurden wie vorher beschrieben bereits unter Einsatz von *computational design* identifiziert und beinhalten ein sehr großes Stabilisierungspotential (Malakauskas und Mayo, 1998). Die Bedeutung der Position 1 ist schwieriger zu analysieren. In den Selektionen war diese Position ebenso wie im gesamten Streptokokkenprotein G durch Asp besetzt. Für die Expression der Domäne in bakteriellen Systemen ist Met an dieser Stelle essentiell. Aus vorangegangenen Studien ist bekannt, dass die Entfernung von Met1 das Wildtypprotein Gβ1 um 7,1 kJ mol⁻¹ destabilisiert (Smith *et al.*, 1994). Diese Präferenz für hydrophobe Aminosäurereste erklärt vielleicht, warum die Mutation D1V selektiert wurde. Jedoch sollte das Expressionssystem beibehalten werden und deshalb wurde die Position 1 von der weiteren Betrachtung ausgeschlossen. Die Mutation T11I ging in späteren Selektionsrunden verloren und die Mutation K13I hatte keinen Einfluss auf die Stabilität des Proteins (vergleiche Variante 13/1 und 13/3 von Bibliothek E1; Tabelle 1, Teilarbeit E). E56V als Einzelmutation wirkte im Wildtypprotein leicht destabilisierend. Ob diese Mutation ebenso wie das Motiv N8Y/G9V in Gegenwart der Mutationen I6V und T53A stabilisierend wirkt, wurde nicht untersucht.

2.4.2 Ortsgerichtete Sättigungsmutagenese der gefundenen Positionen

Aufgrund der Beschränkung durch die Bibliotheksgröße wurden nur fünf Positionen (15, 16, 18, 24 und 37) für eine gleichzeitige Sättigungsmutagenese ausgewählt. Das Potential der Positionen 16 und 18 war bereits aus Teilarbeit C und D bekannt. Die drei Mutationen E15V, A24V, und N37M stabilisieren das Wildtypprotein um insgesamt 3,8 °C (vergleiche Variante E1-13/1 und E3-17/1; Teilarbeit E, Tabelle 1). Die restlichen Positionen 1, 8, 9, 11, 13 und 56 wurden aus den bereits beschriebenen Gründen ausgeschlossen. Dabei ist die Position 13 ein gutes Beispiel für die Schwierigkeit der Auswahl. Die Mutation K13I beeinflusst die Stabilität des Proteins nicht und tritt nur in Gegenwart von anderen Mutationen auf. Der Ausschluss solcher „stiller“ Mutationen bedarf des thermodynamischen Vergleichs einer Vielzahl von Proteinvarianten.

Es wurden drei unabhängige Bibliotheken angelegt, in denen die oben genannten fünf Positionen randomisiert und dabei alle 20 natürlichen Aminosäuren zugelassen waren. Als Ausgangspunkt wurde wiederum die Variante mit den beiden destabilisierenden Mutationen I6V und T53A gewählt. Diese Bibliotheken wurden 10 bis 13 Selektionsrunden unterworfen. Dabei wurden die Bedingungen ausgehend von 40 °C und 0,25 μM Chymotrypsin schrittweise bis auf 57,5 °C und 25 μM erhöht. Am Ende wurden aus jeder Bibliothek zehn Varianten mit deutlich erhöhter Proteaseresistenz sequenziert. Die drei Bibliotheken zeigten ein vergleichbares Selektionsbild (Abbildung 12). An Position 16 und 37 wurde in allen 30 Klonen Leu gefunden. An Position 15 und 18 wurde beide Male Val als häufigster Ersatz

identifiziert, ergänzt das eine mal durch polare Reste (Glu, Thr, Ser, Gln) und das andere mal durch andere unpolare Reste (Val, Ile, Leu, Met). An Position 24 wurden insgesamt neun verschiedene Aminosäurereste gefunden. Dabei reichte die Größe von Gly bis Trp und es waren aliphatische, polare, geladene und aromatische Reste vertreten. Diese Diversität legt nahe, dass die Position 24 für die Stabilität des Proteins von untergeordneter Bedeutung ist. Für die Aminosäurereste wurden dabei auf DNA-Ebene unterschiedliche Codons gefunden (Abbildung 12). An Position 16 und 37 wurden von den sechs möglichen Leu Codons fünf identifiziert. Val an Position 18 war durch alle vier möglichen Basentriplets kodiert, sowie drei der vier an Position 15. Außerdem wurden einige Varianten mit identischer Aminosäureabfolge durch unterschiedliche Nukleotidsequenzen kodiert. Dies zeigt, dass die Bibliotheken zu Beginn ausreichend divers waren.

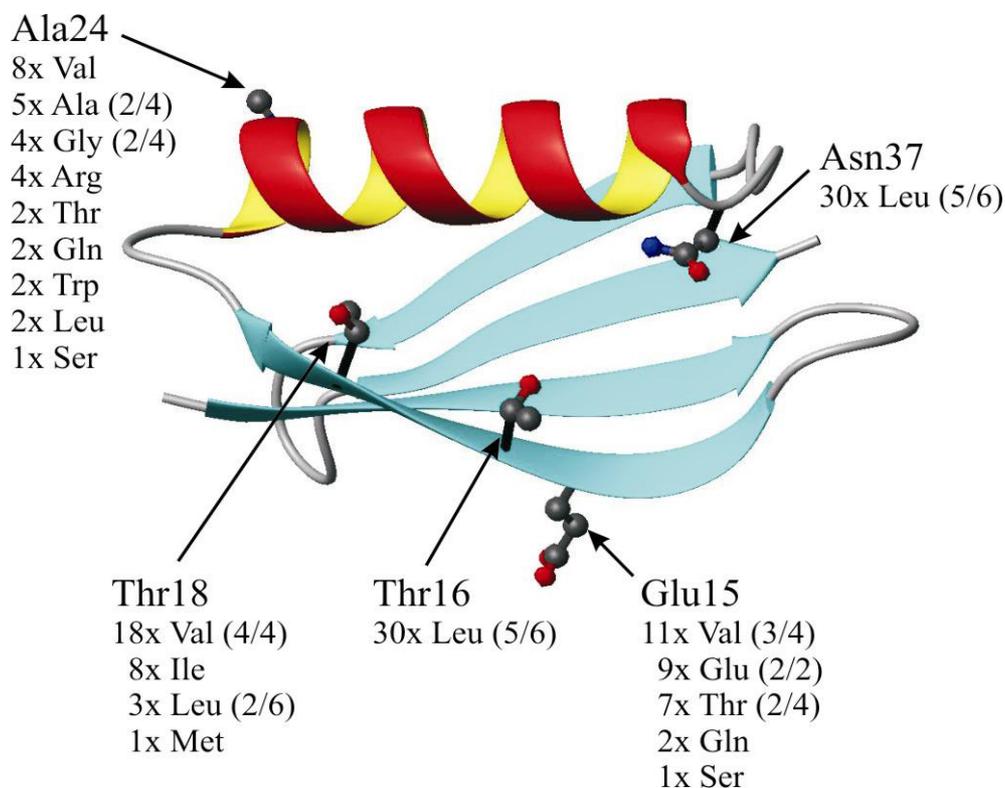


Abbildung 12. Tertiärstrukturdarstellung von Wildtyp G β 1. Die Seitenketten, die in den Selektionen variierten Positionen, sind als Kugel-Stab-Modell gezeigt. Außerdem sind die selektierten Reste sowie deren Häufigkeit angegeben und in den Klammern das Verhältnis von selektierten Codons zu möglichen Codons für diese Aminosäure (falls mehr als ein Codon selektiert wurde). Die Abbildung wurde erstellt unter Verwendung der Daten 1PGB (Gallagher *et al.*, 1994) und dem Programm MOLMOL (Koradi *et al.*, 1996).

Zehn der selektierten Varianten wurden als Proteine im Wildtyphintergrund, d.h. in Abwesenheit der beiden zur Selektion verwendeten destabilisierenden Mutationen I6V und T53A, gereinigt und thermodynamisch charakterisiert. Aufgrund der hohen Stabilität der Varianten wurden die thermischen Übergänge in Gegenwart von 1,5 M GdmCl gemessen. Für alle Varianten war der Mittelpunkt des Überganges um mindestens 22 °C erhöht und für die beiden besten sogar um über 30 °C (Abbildung 13). Dies entspricht einer Stabilisierung von bis zu 24 kJ mol⁻¹ (Tabelle 4).

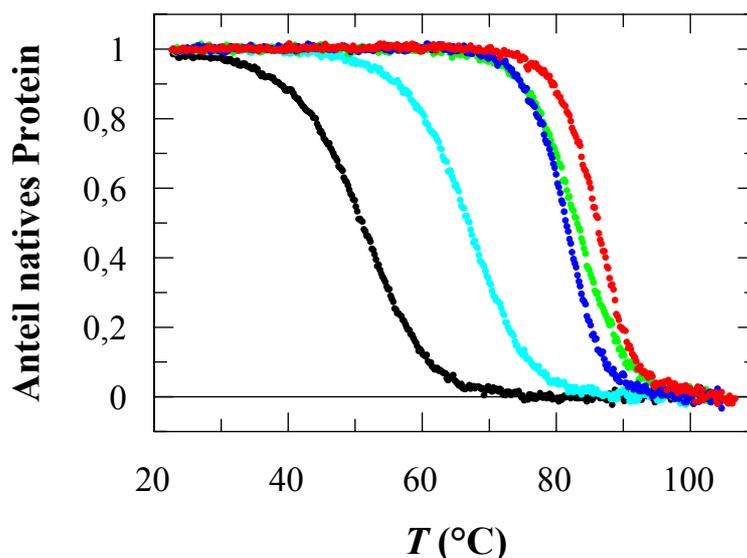


Abbildung 13. Thermische Entfaltungübergänge von Wildtyp Gβ1 (●), der stabilsten Variante aus der fehlerhaften PCR Gβ1 T16I/T18I (●), der stabilsten Variante 1 aus der Sättigungsmutagenese (●) und Variante M2 (●) erhalten aus manueller Kombination der beiden besten Varianten, siebenfache Mutante Gβ1c3-IIIEF (●) erhalten durch Optimierung von drei *core* Positionen Y3F/L7I/V39I und vier *boundary positions* T16I/T18I/T25E/V29F. Die Übergänge von 4 μM Protein in 100 mM Na-Cacodylat/HCl (pH 7,0) wurden verfolgt durch die CD-Änderung bei 222,6 nm in Gegenwart von 1,5 M GdmCl. Dargestellt ist der Anteil an nativem Protein als Funktion der Temperatur nach Analyse der Übergänge entsprechend einem Zweizustandsmodell. Die erhaltenen thermodynamischen Parameter sind zusammengefasst in Tabelle 5 (Teilarbeit E).

Tabelle 4. Stabilitätsdaten der selektierten Varianten aus der Sättigungsmutagenese der Positionen 15, 16, 18, 24 und 37 und der ortsgerichtet erstellten Varianten zur Identifikation der besten Kombination.

Variante	Rest an Position					T_M^a (°C)	ΔH_D^b (kJ mol ⁻¹)	$\Delta G_D(70\text{ °C})^c$ (kJ mol ⁻¹)	$\Delta\Delta G_D^d$ (kJ mol ⁻¹)
	15	16	18	24	37				
wt	Glu	Thr	Thr	Ala	Asn	51,2	186	-12,9	0
1	Val	Leu	Val	Ala	Leu	81,9	351	10,9	23,8
2	Glu	Leu	Ile	Val	Leu	81,5	369	11,2	24,1
M1	Val	Leu	Ile	Val	Leu	85,4	371	14,5	27,4
M2	Val	Leu	Ile	Ala	Leu	86,3	378	15,6	28,5
M3	Val	Ile	Ile	Ala	Leu	84,5	362	13,5	25,9

^a T_M ist der Mittelpunkt des thermischen Entfaltungüberganges, ^b ΔH_D ist die Enthalpie der Entfaltungsreaktion am Mittelpunkt, ^c ΔG_D ist die Änderung der Gibbs'schen freien Energie der Entfaltung bei 70 °C, ^d $\Delta\Delta G_D$ ist die Änderung der Gibbs'schen freien Energie der Entfaltung bei 70 °C relativ zu der des Wildtypproteins. Die thermodynamischen Parameter wurden aus thermischen Entfaltungübergängen, wie in Abbildung 13 gezeigt, ermittelt. Die Varianten sind nummeriert gemäß sinkender Stabilität. Die Varianten M1-M3 stellen die manuelle Kombination der besten Varianten aus den Selektionen dar. Eine Aufstellung für alle in dieser Arbeit charakterisierten Varianten findet sich in Tabelle 3 (Teilarbeit E).

Die stabilste Variante 1 besitzt an Position 24 ein Ala ebenso wie das Wildtypprotein. Drei im Vergleich zu Variante 1 geringfügig destabilisierte, mit Ausnahme von Position 24 aber

identische Varianten, haben ein Arg, Thr, oder Ser an dieser Stelle (vergleiche Variante 1 mit 3, 4 und 5, Tabelle 3, Teilarbeit E). Dies bestätigt die Annahme, dass diese Position unbedeutend für die Stabilität des Proteins ist. Die beiden stabilsten Varianten 1 und 2 teilen außer dem Motiv Leu16/Leu37, welches alle selektierten Varianten beinhalten, keine weiteren Gemeinsamkeiten (Tabelle 4). Deshalb sollte im Folgenden untersucht werden, ob es eine bessere Kombination der Mutationen an den Positionen 15, 18 und 24 gibt als die bisher selektierten.

Die Variante 2 besitzt nicht das an Position 15 am häufigsten selektierte Val, welches in vier der fünf besten Varianten zu finden ist. Deshalb wurde Glu15 in der Variante 2 gegen Val ausgetauscht und man erhält die Variante M1. Dieser Ersatz erhöhte die Stabilität um 3,9 °C bzw. 3,3 kJ mol⁻¹ (Tabelle 4). Als nächstes wurde Val24 gegen Ala24 ausgetauscht, wie es in der besten selektierten Variante zu finden ist. Diese Mutation bewirkt eine weitere, aber eher geringfügige Stabilisierung um 0,9 °C bzw. 1,1 kJ mol⁻¹ (vergleiche Variante M1 und M2, Tabelle 4). Offensichtlich stellt das im Wildtypprotein auftretende Ala24 bereits einen adäquaten Rest an dieser Position dar. Die Variante M2 ist die stabilste Kombination und ihr T_M -Wert ist um 35,1 °C erhöht (Abbildung 13), was einer Stabilisierung von 28,5 kJ mol⁻¹ entspricht (Tabelle 4).

In der vorangegangenen Optimierung der *boundary positions* wurde Ile an Position 16 als der beste Ersatz identifiziert. Es war um 1,7 kJ mol⁻¹ stärker stabilisierend als Leu an entsprechender Stelle (vergleiche Variante 2 und 4, Tabelle 3). In den jetzigen Selektionen wurde allerdings nie ein Ile an der Position 16 gefunden. Um diesen Unterschied aufzuklären, wurde Ile16 durch ortsgerichtete Mutagenese in die Variante M2 eingeführt. Diese Mutation destabilisierte das Protein um 1,8 °C bzw. 2,1 kJ mol⁻¹ (vergleiche Variante M2 und M3, Tabelle 4). Dies zeigt, dass die Effekte der Mutationen durch deren lokale Umgebung beeinflusst werden und deshalb sind allgemeine Lösungen für die Stabilität eines Proteins schwer zu finden.

Die Variante Gβ1-M2 mit nur vier Mutationen und einer Stabilisierung von 35,1 °C bzw. 28,5 kJ mol⁻¹ war deutlich stabiler als die beste Variante Gβ1c3-IIIEF aus dem gemischt auf Berechnungen und *in vitro* Evolution beruhenden Ansatz aus Teilarbeit C und D. Diese siebenfache Mutante zeigt einen um 31,9 °C erhöhten T_M -Wert (Abbildung 13) und ist um 22,7 kJ mol⁻¹ stabilisiert. Dabei kann man deutlich erkennen, dass für die aus der Selektion stammenden Varianten 1 und M2 die Kooperativität der Entfaltung im Vergleich zu Gβ1c3-IIIEF deutlich erhöht ist.

2.4.3 Analyse der Ursachen für die erzielte Stabilisierung von Gβ1

Im folgenden wurden die Ursachen für die große Stabilisierung um 35,1 °C bzw. 28,5 kJ mol⁻¹ der durch *in vitro* Evolution geschaffenen Variante Gβ1-M2 (E15V/T16L/T18I/N37L) im Vergleich zum Wildtypprotein durch ortsgerichtete Mutagenese untersucht. Dabei wurden zwei Ansätze gewählt, zum einen wurden jeweils einzeln die vier Mutationen von Gβ1-M2 in das Wildtypprotein eingebaut und zum anderen ausgehend von Gβ1-M2 jeweils eine der Positionen 15, 16, 18 und 37 zum entsprechenden Rest des Wildtypproteins zurückmutiert. Die entsprechenden Proteine wurden gereinigt und wegen der hohen Stabilität in Gegenwart von 1,5 M GdmCl thermodynamisch charakterisiert (Tabelle 1, Teilarbeit F). Die aufgetretenen Effekte werden im Folgenden anhand der Kristallstruktur des Wildtypproteins diskutiert. Einflüsse auf den entfaltenen Zustand werden dabei vernachlässigt,

obwohl diese prinzipiell auch zur Stabilität von Proteinen beitragen können (Pace *et al.*, 2000; Cho und Raleigh, 2005).

Position 15

Position 15 ist auf der Außenseite des β -Faltblattes und hier im Faltblattstrang $\beta 2$ lokalisiert (Abbildung 14). Das im Wildtypprotein zu findende Glu15 an dieser Stelle ist lösungsmittelzugänglich. In den Selektionen wurden sowohl Val als auch hydrophile Aminosäuren als geeigneter Ersatz identifiziert (Abbildung 12), was mit den strukturellen Gegebenheiten der Position übereinstimmt. Tauscht man Glu15 gegen Val aus, so erhöht sich der Schmelzpunkt des Proteins um 2,4 °C (vergleiche Variante 5 und wt, Tabelle 1, Teilarbeit F), was einer Stabilisierung von $1,6 \text{ kJ mol}^{-1}$ entspricht (Abbildung 14). Der Effekt der Mutationen ist mit einer Destabilisierung von $3,8 \text{ kJ mol}^{-1}$ noch etwas ausgeprägter, wenn man von G $\beta 1$ -M2 ausgeht und hier die Umkehrmutation V15E einführt (Abbildung 14). Dieses Verhalten ist in einer schwachen Wechselwirkung des Val mit den Aminosäureketten der anderen drei Mutationen begründet. Der stabilisierende Effekt von E15V ist zum Teil auf den unterschiedlichen Hang der beiden Resten zur Bildung von β -Faltblattstrukturen zurückzuführen. Dieses Bestreben ist sehr niedrig für Glu, jedoch sehr hoch für Val (Smith *et al.*, 1994; Smith und Regan, 1997).

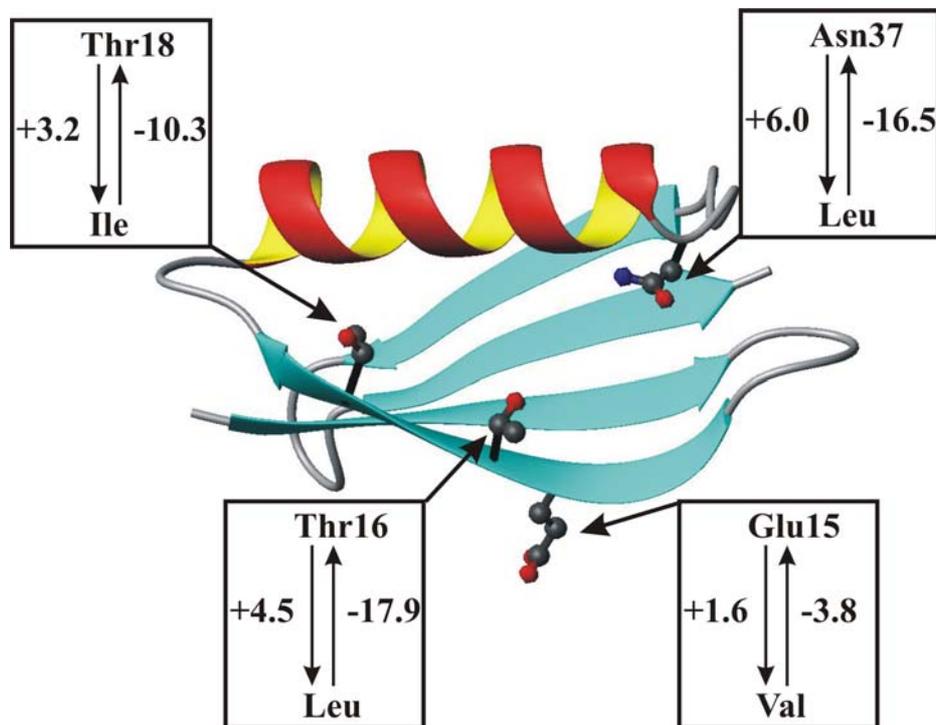


Abbildung 14. Tertiärstrukturdarstellung von Wildtyp G $\beta 1$. Die Seitenketten der Positionen 15, 16, 18 und 37 sind als Kugel-Stab-Modell gezeigt. Außerdem sind der stabilisierende Einfluss von jeder Einzelmutation im Wildtyphintergrund und der destabilisierende Effekt des Austausches einer Aminosäure in der stabilsten Variante zurück zum Wildtyprest aufgeführt (jeweils in kJ mol^{-1}). Die thermodynamischen Daten sind aus Tabelle 1, Teilarbeit F entnommen. Die Abbildung wurde erstellt unter Verwendung der Daten 1PGB (Gallagher *et al.*, 1994) und dem Programm MOLMOL (Koradi *et al.*, 1996).

Position 18

Position 18 ist am Ende von Strang $\beta 2$ des β -Faltblattes lokalisiert. Die Bedeutung der Position für die Stabilität von G $\beta 1$ wurde bereits durch Einsatz von *computational design* identifiziert und wurde durch *in-vitro* Evolution (Teilarbeit C) und ortsgerichtete Mutagenese (Teilarbeit D) bestätigt. Ile18 stabilisiert das Wildtypprotein um 4,7 °C (vergleiche Variante 7 und wt, Tabelle 1, Teilarbeit F) bzw. 3,2 kJ mol⁻¹ (Abbildung 14). In Abwesenheit von GdmCl ist der Effekt mit 3,8 kJ mol⁻¹ (Tabelle 2, Teilarbeit D) etwas stärker ausgeprägt. Im Gegensatz dazu destabilisiert die Umkehrmutation I18T ausgehend von G $\beta 1$ -M2 das Protein um 10,2 kJ mol⁻¹ (Abbildung 14). Dieser deutliche Unterschied der Effekte der beiden Mutationen zeigt, dass die Umgebung einen großen Einfluss auf die Wirkung der Position 18 besitzt und Wechselwirkungen zwischen den Resten auftreten. Deshalb können die Austausche in G $\beta 1$ -M2 nicht als unabhängig betrachtet werden. Die Seitenkette von Thr18 ist dem aromatischen Ring von Phe30 benachbart (Abstand 3,9 Å) und die beiden Reste bilden im Wildtypprotein eine Kontaktfläche (Sobolev *et al.*, 1999) von 31,4 Å². Offensichtlich sind hydrophobe Wechselwirkungen zwischen Ile18 und Phe30 für den stabilisierenden Effekt der Mutation T18I verantwortlich.

Position 16

Position 16 ist ebenfalls im Faltblattstrang $\beta 2$ lokalisiert und in der Variante G $\beta 1$ -M2 durch Leu besetzt. Interessanterweise wurde bei der Optimierung der *boundary* Positionen 16, 18, 25 und 29 sowohl durch das *computational design* (Malakauskas und Mayo, 1998), als auch durch *in vitro* Evolution (Teilarbeit C) Ile als der beste Ersatz für Thr16 identifiziert. Die Mutation T16L stabilisiert das Wildtypprotein um 5,6 °C (vergleiche Variante 6 und wt, Tabelle 1, Teilarbeit F), was 4,5 kJ mol⁻¹ entspricht (Abbildung 14). Die umgekehrte Mutation L16T ausgehend von G $\beta 1$ -M2 destabilisiert das Protein sogar um 20,0 °C (vergleiche Variante 6 und wt, Tabelle 1, Teilarbeit F) bzw. 17,9 kJ mol⁻¹ (Abbildung 14). Dieser beträchtliche Unterschied (13,4 kJ mol⁻¹, Abbildung 12) zwischen der Einführung von Leu16 in das Wildtypprotein und der Entfernung aus G $\beta 1$ -M2 ist annähernd gleich groß wie die gesamte Stabilisierung von G $\beta 1$ durch die Optimierung der vier *boundary* Positionen (Tabelle 3). Dieses Verhalten deutet wiederum auf starke, zusätzlich stabilisierende Wechselwirkungen zwischen den veränderten Resten in G $\beta 1$ -M2 hin. Die Seitenkette von Thr16 liegt in lokaler Nachbarschaft zum aromatischen Ring von Tyr33 (Abstand 4,0 Å) und der hydrophoben Seitenkette von Leu5 (Abstand 4,1 Å). Kontaktflächen (Sobolev *et al.*, 1999) von 34,0 Å² bzw. 42,1 Å² werden im Wildtypprotein zwischen Thr16 und diesen Resten gebildet. Der Austausch von Thr16 gegen Leu innerhalb dieser hydrophoben Tasche ermöglicht zusätzliche van-der-Waals-Wechselwirkungen und ist für den stabilisierenden Effekt von Leu16 verantwortlich. Die Stabilisierung durch T16L ist annähernd gleich groß wie die durchschnittliche Stabilisierung von Proteinen durch eine zusätzliche Methylgruppe, die während der Faltung im Inneren eines Proteins verborgen wird (Pace, 1992).

Position 37

Position 37 liegt am Beginn der Schleife zwischen der α -Helix und dem β -Faltblattstrang $\beta 3$ lokalisiert (Abbildung 14). Bei der Sättigungsmutagenese mit anschließender *in vitro* Evolution wurde ausschließlich Leu als Ersatz für das im Wildtypprotein hier vorkommende

Asn gefunden. Die Mutation N37L stabilisiert das Wildtypprotein um $6,0 \text{ kJ mol}^{-1}$ (Abbildung 14), was der höchste Wert für alle in Teilarbeit D und F untersuchten Einzelvarianten ist. Interessanterweise wurde dieses sehr große Potential der Position 37 durch das *computational design* nicht erkannt, obwohl diese Position ebenfalls als „*boundary*“ klassifiziert worden war (Malakauskas und Mayo, 1998). Der Austausch von Leu37 zu Asn in der Variante G β 1-M2 bewirkt mit $16,5 \text{ kJ mol}^{-1}$ (Abbildung 14) wiederum eine deutliche größere Destabilisierung als die umgekehrte Einzelvariante im Wildtypprotein. Die Seitenkette von Asn37 ist von großen hydrophoben oder aromatischen Resten umgeben. Der Abstand im Wildtypprotein beträgt zum Phenylring von Tyr33 $3,6 \text{ \AA}$ und zur Seitenkette von Leu12 $3,7 \text{ \AA}$. Zu diesen beiden Resten werden Kontaktflächen (Sobolev *et al.*, 1999) von $48,9 \text{ \AA}^2$ bzw. $30,4 \text{ \AA}^2$ gebildet. Die Verkürzung der Seitenkette von Asn37 durch die Mutation N37A führte zu einer leichten Stabilisierung von $0,7 \text{ kJ mol}^{-1}$ (McCallister *et al.*, 2000). Dieses Ergebnis zeigt deutlich, dass die polare Seitenkette von Asn37 ungünstige Wechselwirkung wahrscheinlich vor allem zum aromatischen Ring von Tyr33 ausbildet. Durch den Austausch mit dem isosteren, aber hydrophoben Leu37 werden diese ungünstigen Wechselwirkungen unterbunden und zusätzlich stabilisierende van-der-Waals-Wechselwirkungen in das Protein eingebracht.

Analyse der Wechselwirkungen zwischen den Positionen 16, 18 und 37

Wie die Analyse der Einzelmutationen gezeigt hat, ist der destabilisierende Effekt bei der Entfernung einer Mutation aus der Variante G β 1-M2 besonders für die Positionen 16, 18 und 37 deutlich größer als für die Einführung einer einzelnen Mutation im Hintergrund des Wildtypproteins. Ein solches Verhalten lässt sich nur durch Wechselwirkungen der einzelnen mutierten Aminosäuren miteinander erklären. Für drei Positionen kann man solche Wechselwirkungsenergien mit Hilfe eines Dreifach-Mutationszyklus quantifizieren (Perl und Schmid, 2001; Horovitz und Fersht, 1992; Horovitz, 1996). Eine solche Analyse wurde für die Position 16, 18 und 37 durchgeführt (Abbildung 15).

Die Positionen 16 und 18 liegen direkt benachbart auf der zum Inneren des Proteins gewandten Seite des β -Faltblattstranges β_2 und im Wildtypprotein beträgt der Abstand der C $^\beta$ -Atome $6,8 \text{ \AA}$. In der Variante G β 1-M2 sind diese beiden Positionen durch Leu bzw. Ile belegt und die Doppelvariante ist um $10,3 \text{ kJ mol}^{-1}$ stabilisiert (vergleiche Variante 9, Tabelle 1, Teilarbeit F). Hingegen beträgt die Summe der stabilisierenden Effekte der beiden Einzelmutationen $7,7 \text{ kJ mol}^{-1}$. Dies bedeutet, dass zwischen Leu16 und Ile18 eine zusätzliche Wechselwirkung von $2,6 \text{ kJ mol}^{-1}$ besteht (vergleiche Zyklus oben links, Abbildung 15). Im Wildtypprotein wird von Thr16 und Thr18 keine Kontaktfläche (Sobolev *et al.*, 1999) gebildet, jedoch ermöglicht der Ersatz durch die sterisch größeren Aminosäuren Leu16 und Ile18 offenbar eine Wechselwirkung.

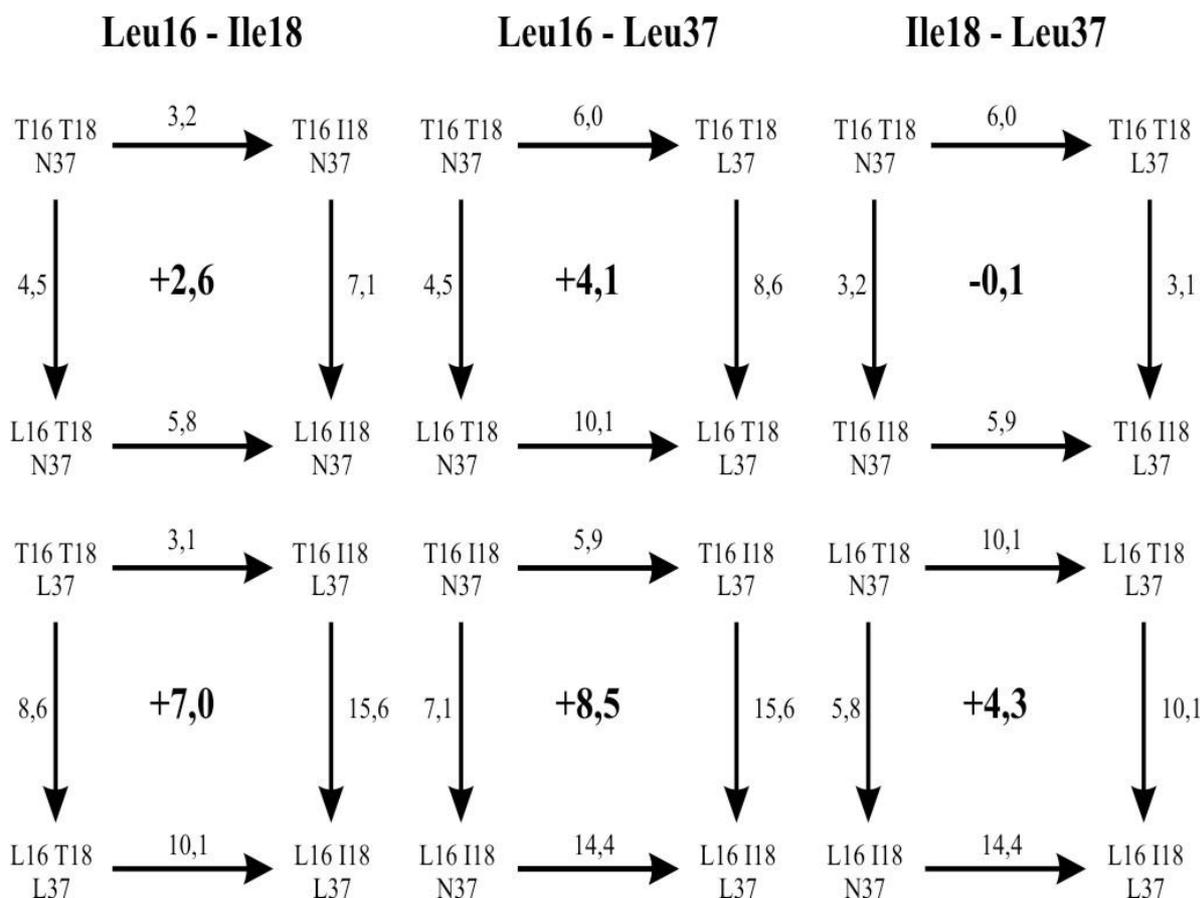


Abbildung 15. Dreifach-Mutationszyklus zur Analyse der stabilisierenden Wechselwirkungen zwischen Leu16, Ile18 und Leu37 in der Variante Gβ1-M2. Die experimentellen Daten sind aus Tabelle 1, Teilarbeit F entnommen. Für jede einzelne Mutation sind die $\Delta\Delta G_D$ -Werte bei 70 °C gezeigt. In der Mitte eines jeden Zyklus ist die Wechselwirkungsenergie zwischen den jeweiligen zwei Aminosäureresten angegeben (kJ mol^{-1}). In der oberen Reihe sind jeweils die Zyklen in Abwesenheit der dritten Mutation und in der unteren Reihe in Anwesenheit der dritten Mutation gezeigt. Die linke Spalte bestimmt die Wechselwirkungsenergie zwischen Leu16 und Ile18, die mittlere zwischen Leu16 und Leu37 und die rechte zwischen Ile18 und Leu37.

Der Abstand zwischen den C^β -Atomen der Position 16 und 37 beträgt 10,4 Å und aufgrund dieses Wertes ist eine direkte Wechselwirkung zwischen den beiden Positionen auszuschließen. Addiert man die Wirkung der beiden Einzelmutationen, so erhält man eine Stabilisierungsenergie von 10,5 kJ mol^{-1} . Aber im Vergleich dazu ist die Doppelvariante mit beiden Mutationen um 14,6 kJ mol^{-1} stabilisiert (Tabelle 1, Teilarbeit F). Offensichtlich bilden Leu16 und Leu37 Wechselwirkungen aus und erzielen dadurch eine zusätzliche Stabilisierung von 4,1 kJ mol^{-1} (vergleiche Zyklus Mitte oben, Abbildung 15). Eine direkte Wechselwirkung ist wie bereits erwähnt nicht möglich, jedoch bildet sowohl die Seitenkette von Position 16 als auch von Position 37 eine Kontaktfläche (Sobolev *et al.*, 1999) zu dem aromatischen Ring des Restes Tyr33. Die Abstände im Wildtypprotein betragen dabei für beide Positionen maximal 4,0 Å. Diese enge Verknüpfung der hydrophoben Seitenketten von Leu16, Tyr33 und Leu37 ist offensichtlich für die starke Stabilisierung verantwortlich und die Seitenkette von Tyr33 vermittelt die im Mutationszyklus gemessene energetische Wechselwirkung zwischen Leu16 und Leu37.

Der Abstand zwischen den C^β -Atomen der Positionen 18 und 37 beträgt 14,3 Å und die Summe der Effekte der beiden Einzelmutationen T18I und N37L ist identisch zur

Stabilisierung der Doppelvariante. Wie aufgrund der großen Entfernung erwartet ist somit keine Wechselwirkung zwischen diesen beiden Positionen festzustellen (vergleiche Zyklus oben rechts, Abbildung 15).

Zusätzlich zur Wechselwirkungsenergie zwischen zwei Aminosäureresten liefert ein Dreifachmutationszyklus auch die Wechselwirkung in Gegenwart von allen drei Mutationen. Diese Wechselwirkungsenergie kann quantifiziert werden durch einen Vergleich der jeweiligen Doppelmutationszyklen in Abwesenheit und Gegenwart der dritten Mutation. In diesem Fall beträgt die zusätzliche Stabilisierung in Gegenwart aller drei Austausche $4,4 \text{ kJ mol}^{-1}$ (vergleiche obere und untere Reihe, Abbildung 15). Diese Anhäufung der hydrophoben Aminosäuren Leu16, Ile18, Tyr33 und Leu37 und die damit einhergehende Ausbildung von van-der-Waals-Wechselwirkungen ist verantwortlich für die starke Stabilisierung der Variante G β 1-M2. Die Summe der Effekte der vier Einzelmutationen dieser Variante beträgt $15,3 \text{ kJ mol}^{-1}$ und im Vergleich dazu ist die Vierfachvariante G β 1-M2 um $28,5 \text{ kJ mol}^{-1}$ stabilisiert. Somit ist ungefähr die Hälfte der Stabilisierung auf Wechselwirkungen zwischen den variierten Seitenketten zurückzuführen. Interessanterweise konnten bei der Optimierung der vier *boundary* Positionen 16, 18, 25 und 29 nahezu keine Wechselwirkungen zwischen diesen Positionen identifiziert werden und die Stabilisierung von in diesem Fall $14,9 \text{ kJ mol}^{-1}$ beruhte auf vier additiven Einzelmutationen (Teilarbeit C und D), obwohl auch diese Reste eng benachbart sind.

Analyse der Enthalpie der Entfaltung

Varianten von G β 1, die das Motiv T16L/N37L enthielten, zeigten bei der Entfaltung eine deutlich erhöhte Kooperativität oder anders ausgedrückt der Übergangsbereich für diese Varianten war kleiner. Durch eine Erhöhung der Kooperativität wird gleichfalls wie bei einer Erhöhung des Mittelpunktes bei der thermischen Entfaltung der Beginn des Überganges und somit die Zugänglichkeit für Proteasen zu höheren Temperaturen verschoben. Dies erklärt, warum auf den Positionen 16 und 37 ausschließlich Leu selektiert wurde (Abbildung 12). Das Verhalten wird offensichtlich, wenn man die van't Hoff Enthalpien der Einzelvarianten Leu16 und Leu37 mit der der entsprechenden Doppelvariante vergleicht (vergleiche Variante 6 und 8 mit Variante 10, Tabelle 1, Teilarbeit F). Für die beiden Einzelvarianten ist die van't Hoff Enthalpie etwa gleich und beträgt ca. 200 kJ mol^{-1} . Für die Doppelmutante T16L/N37L ist ΔH_D um etwa 120 kJ mol^{-1} erhöht, obwohl der T_M -Wert nur um ca. $12 \text{ }^\circ\text{C}$ erhöht ist. Der Δc_p -Wert der Entfaltung von G β 1 beträgt etwa $4 \text{ kJ mol}^{-1} \text{ K}^{-1}$, was somit eine maximale Differenz von 50 kJ mol^{-1} begründen würde. Die gleichzeitige Anwesenheit von Leu16 und Leu37 führt somit zu einem Anstieg von ΔH_D um etwa 70 kJ mol^{-1} .

Um dieses Verhalten zu analysieren, wurden thermische Entfaltungsübergänge in der Gegenwart von verschiedenen GdmCl-Konzentrationen gemessen. Trägt man die erhaltenen van't-Hoff-Enthalpien ΔH_D gegen den jeweiligen Schmelzpunkt T_M auf, so erhält man für die Variante G β 1-M2 eine deutlich steilere Gerade als für das Wildtypprotein (Abbildung 6, Teilarbeit F). Die Steigungen der beiden Geraden repräsentieren die apparente Änderung der Wärmekapazität für die Entfaltung Δc_p . Die so erhaltenen Δc_p -Werte sind höher als die wirklichen Werte, denn GdmCl solvatisiert das entfaltete Protein (Pfeil und Privalov, 1976; Makhatadze und Privalov, 1992), aber für Varianten eines Proteins vergleichbar. In Abbildung 6 (Teilarbeit F) ist zu erkennen, dass durch die Mutationen der Δc_p -Wert um etwa

1,5 kJ mol⁻¹ K⁻¹ erhöht wird, was einer Steigerung von etwa 40 % im Vergleich zum Δc_p -Wert des Wildtypproteins entspricht.

In erster Näherung besteht eine lineare Beziehung zwischen Δc_p und der Anzahl der Aminosäurereste eines Proteins (Myers *et al.*, 1995). Auf der anderen Seite ist Δc_p verknüpft mit der Oberfläche eines Proteins, die während der Entfaltung dem Lösungsmittel zugänglich gemacht wird (Murphy und Freire, 1992). Die Anzahl der Aminosäuren ist für G β 1-M2 im Vergleich zum Wildtypprotein unverändert. Deshalb ist eine mögliche Erklärung für die Erhöhung von Δc_p eine kompaktere Packung der Variante, die bei der Entfaltung von G β 1-M2 zu einer größeren Oberfläche führt. Das Verhalten der Variante zeigt, dass bei der *in vitro* Evolution von G β 1 vom Protein zwei Möglichkeiten genutzt wurden, der Spaltung durch Proteasen zu entgehen. Zum einem ist natürlich die Hauptursache eine parallele Verschiebung des gesamten Übergangs nach rechts, d.h. zu höheren Temperaturen, aber auch eine steilere Kurvenform, denn somit wird der Beginn der Entfaltung ebenfalls zu höheren Temperaturen verschoben. Dies zeigt die vielfältigen Möglichkeiten, die ein Protein für seine thermodynamische Stabilisierung nützt und beide beschriebenen Effekte sind verantwortlich für die deutliche Erhöhung der Stabilität um 28,5 kJ mol⁻¹.

2.5 Fazit der Arbeit

Das Ziel der vorliegenden Arbeit war das Selektionssystem *Proside* zur Stabilisierung von Proteinen zu nützen. Als Modellproteine wurden hierfür das Kälteschockprotein *Bs-CspB* und die β 1-Domäne des Strepptokokkenproteins G gewählt und für beide Proteine konnten erfolgreich beachtliche Stabilisierungen erzielt werden.

Ein jeder Versuch der Stabilisierung ist dabei aufgebaut aus zwei Schritten, der Identifikation von viel versprechenden Positionen und dem Finden der besten Aminosäureaustausche für diese Positionen. Es konnte gezeigt werden, dass sich für den ersten Schritt sowohl rechnergestützte als auch experimentelle Methoden eignen. Für den zweiten Schritt ist jedoch die *in vitro* Evolution vorzuziehen. Eine Verbesserung der in den Berechnungen verwendeten Algorithmen durch die gewonnenen experimentellen Daten ermöglicht in Zukunft vielleicht den Einsatz des *computational design* für beide Schritte.

Bei der Stabilisierung von *Bs-CspB* konnte die Bedeutung geladener Gruppen auf der Oberfläche von Proteinen gezeigt werden. Vergleiche der experimentellen Ergebnisse mit vorhandenen theoretischen Untersuchungen offenbarten die immer noch vorhandenen Schwierigkeiten Coulomb'sche Wechselwirkungen auf der Oberfläche von Proteinen korrekt zu berechnen. Die geschaffenen, umfangreichen experimentellen Daten für das Kälteschockprotein werden sich sehr gut dafür eignen, Elektrostatikberechnungen zu validieren.

Für G β 1 konnte gezeigt werden, welche Schwierigkeiten bestehen Kontakte in Raumstrukturen in thermodynamische Effekte zu übersetzen und umgekehrt. So beruht die Stabilisierung bei der Optimierung der vier „*boundary*“ Positionen 16, 18, 25 und 29 trotz der lokalen Nähe auf vier nahezu unabhängigen Austausch. Im Gegensatz dazu traten bei einer Mutagenese der Positionen 16, 18 und 37 starke stabilisierende Wechselwirkungen auf.

Dies alles zeigt die vielfältigen Möglichkeiten ein Protein erfolgreich zu stabilisieren. Die Ergebnisse legen jedoch den Schluss nahe, dass es keine allgemein gültige Lösung für Stabilisierung von Proteinen gibt.

3. Abkürzungen

<i>Bc</i> -Csp	Kälteschockprotein Csp aus <i>Bacillus caldolyticus</i>
<i>Bs</i> -CspB	Kälteschockprotein CspB aus <i>Bacillus subtilis</i>
Csp	Kälteschockprotein
Δc_p	Wärmekapazitätsänderung während der Entfaltung bei konstantem Druck
CD	Circulardichroismus
CT	C-terminale Domäne des Gen-3-Proteins
D	Debye
[D]	Konzentration des Denaturierungsmittels
DSC	<i>differential scanning calorimetry</i>
ε	Dielektrizitätskonstante
ΔG_D	freie Enthalpie der Entfaltungsreaktion
$\Delta\Delta G_D$	Differenz der freien Enthalpien der Entfaltung zweier Proteinvarianten
$\Delta\Delta G_D(70\text{ °C})$	Differenz der freien Enthalpien der Entfaltung zweier Proteinvarianten bei einer bestimmten Temperatur
$\Delta G_D(\text{H}_2\text{O})$	freie Enthalpie der Entfaltungsreaktion in Abwesenheit des Denaturierungsmittels
$\Delta G_D(T)$	Temperaturabhängigkeit der freien Enthalpie der Entfaltungsreaktion
G β 1	Domäne β 1 des Streptokokkenproteins G
G β 1-c3	Variante von G β 1 mit den optimierten drei <i>core</i> Positionen Y3F/L7I/V39I
G β 1-c3b4	Variante von G β 1 mit den optimierten drei <i>core</i> Positionen Y3F/L7I/V39I und den vier <i>boundary</i> Positionen T16I/T18I/T25E/V29I, stabilste Variante aus dem <i>computational design</i>
G β 1c3-IIEF	Stabilste Variante erhalten aus der in vitro Evolution der vier <i>boundary</i> Positionen sie enthält neben den drei optimierten <i>core</i> Positionen Y3F/L7I/V39I die Mutationen T16I/T18I/T25E/V29F
G β 1c3-ILEK	Variante identisch mit der aus dem <i>computational design</i> stammenden Variante G β 1-mc4, enthält die Mutationen Y3F/L7I/V39I an den <i>core</i> Positionen und T16I/T18L/T25E/V29K an den <i>boundary</i> Positionen
G β 1-IIEF	Identisch mit G β 1c3-IIEF, jedoch ohne die drei <i>core</i> Mutationen
G β 1-ILEK	Identisch mit G β 1c3-ILEK, jedoch ohne die drei <i>core</i> Mutationen

Gβ1-M2	Stabilste Variante von Gβ1 erhalten bei der rein experimentellen Optimierung, enthält die Mutationen E15V/T16L/T18I/N37L
Gβ1-mc4	Variante von Gβ1 mit den optimierten drei <i>core</i> Positionen Y3F/L7I/V39I und den vier <i>boundary</i> Positionen T16I/T18L/T25E/V29K, Platz 4 im <i>computational design</i>
GdmCl	Guanidiniumchlorid
[GdmCl] _m	Mittelpunkt des GdmCl-induzierten Entfaltungsübergangs
G3P	Gen-3-Protein von filamentöser Phagen
ΔH_D	Enthalpie der Entfaltungsreaktion
ΔH_{vH}	van't Hoff-Enthalpie der Entfaltung
K_D	Gleichgewichtskonstante der Entfaltungsreaktion
k_N	Geschwindigkeitskonstante der Rückfaltungsreaktion
k_U	Geschwindigkeitskonstante der Entfaltungsreaktion
m	Kooperativitätsparameter bei chemisch induzierten Entfaltungsübergängen
N	nativer Zustand eines Proteins
N1	Domäne N1 des Gen-3-Proteins
N2	Domäne N2 des Gen-3-Proteins
Na-Cacodylat	Natriumsalz der Dimethylarsinsäure
P	Quasidipolmoment
ΔS_D	Entropie der Entfaltungsreaktion
T_M	Mittelpunkt des thermisch induzierten Entfaltungsüberganges
<i>Tm</i> -Csp	Kälteschockprotein Csp aus <i>Thermotoga maritima</i>
U	entfalteter Zustand eines Proteins

4. Literaturverzeichnis

- Akerstrom, B., E. Nielsen, et al. (1987). "Definition of IgG- and albumin-binding regions of streptococcal protein G." *J. Biol. Chem.* **262**(28): 13388-13391.
- Arnold, F. H. (2001). "Combinatorial and computational challenges for biocatalyst design." *Nature* **409**(6817): 253-257.
- Bae, W., B. Xia, et al. (2000). "Escherichia coli CspA-family RNA chaperones are transcription antiterminators." *Proc. Natl. Acad. Sci. USA* **97**(14): 7784-7789.
- Bolon, D. N., R. A. Grant, et al. (2005). "Specificity versus stability in computational protein design." *Proc. Natl. Acad. Sci. USA* **102**(36): 12724-12729.
- Bouvet, J. P. (1994). "Immunoglobulin Fab fragment-binding proteins." *Int. J. Immunopharmacol.* **16**(5-6): 419-424.
- Cacace, M. G., E. M. Landau, et al. (1997). "The Hofmeister series: salt and solvent effects on interfacial phenomena." *Q. Rev. Biophys.* **30**(3): 241-277.
- Chakravarty, S. and R. Varadarajan (2002). "Elucidation of factors responsible for enhanced thermal stability of proteins: a structural genomics based study." *Biochemistry* **41**(25): 8152-8161.
- Cho, J. H. and D. P. Raleigh (2005). "Mutational analysis demonstrates that specific electrostatic interactions can play a key role in the denatured state ensemble of proteins." *J. Mol. Biol.* **353**(1): 174-185.
- Cregut, D., C. Civera, et al. (1999). "A tale of two secondary structure elements: when a beta-hairpin becomes an alpha-helix." *J. Mol. Biol.* **292**(2): 389-401.
- Dahiyat, B. I. and S. L. Mayo (1997a). "De novo protein design: fully automated sequence selection." *Science* **278**(5335): 82-87.
- Dahiyat, B. I. and S. L. Mayo (1997b). "Probing the role of packing specificity in protein design." *Proc. Natl. Acad. Sci. USA* **94**(19): 10172-10177.
- Dantas, G., B. Kuhlman, et al. (2003). "A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins." *J. Mol. Biol.* **332**(2): 449-460.
- de Bakker, P. I., P. H. Hunenberger, et al. (1999). "Molecular dynamics simulations of the hyperthermophilic protein sac7d from *Sulfolobus acidocaldarius*: contribution of salt bridges to thermostability." *J. Mol. Biol.* **285**(4): 1811-1830.
- Deng, L. W. and R. N. Perham (2002). "Delineating the site of interaction on the pIII protein of filamentous bacteriophage fd with the F-pilus of *Escherichia coli*." *J. Mol. Biol.* **319**(3): 603-614.
- Derrick, J. P. and D. B. Wigley (1992). "Crystal structure of a streptococcal protein G domain bound to an Fab fragment." *Nature* **359**(6397): 752-754.
- Desjarlais, J. R. and T. M. Handel (1999). "Side-chain and backbone flexibility in protein core design." *J. Mol. Biol.* **290**(1): 305-318.
- Dominy, B. N., D. Perl, et al. (2002). "The effects of ionic strength on protein stability: the cold shock protein family." *J. Mol. Biol.* **319**(2): 541-554.
- Drummond, D. A., B. L. Iverson, et al. (2005). "Why high-error-rate random mutagenesis libraries are enriched in functional and improved proteins." *J. Mol. Biol.* **350**(4): 806-816.
- Eisenberg, D. and A. D. McLachlan (1986). "Solvation energy in protein folding and binding." *Nature* **319**(6050): 199-203.
- Elcock, A. H. (1998). "The stability of salt bridges at high temperatures: implications for hyperthermophilic proteins." *J. Mol. Biol.* **284**(2): 489-502.
- Fernandez-Gacio, A., M. Uguen, et al. (2003). "Phage display as a tool for the directed evolution of enzymes." *Trends Biotechnol.* **21**(9): 408-414.

- Firth, A. E. and W. M. Patrick (2005). "Statistics of protein library construction." *Bioinformatics* **21**(15): 3314-3315.
- Fukuchi, S. and K. Nishikawa (2001). "Protein surface amino acid compositions distinctively differ between thermophilic and mesophilic bacteria." *J. Mol. Biol.* **309**(4): 835-843.
- Gallagher, T., P. Alexander, et al. (1994). "Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR." *Biochemistry* **33**(15): 4721-4729.
- Garofoli, S., M. Falconi, et al. (2004). "Thermophilicity of wild type and mutant cold shock proteins by molecular dynamics simulation." *J. Biomol. Struct. Dyn.* **21**(6): 771-780.
- Graumann, P. and M. A. Marahiel (1994). "The major cold shock protein of *Bacillus subtilis* CspB binds with high affinity to the ATTGG- and CCAAT sequences in single stranded oligonucleotides." *FEBS Lett.* **338**(2): 157-160.
- Graumann, P., K. Schroder, et al. (1996). "Cold shock stress-induced proteins in *Bacillus subtilis*." *J. Bacteriol.* **178**(15): 4611-4619.
- Graumann, P. L. and M. A. Marahiel (1998). "A superfamily of proteins that contain the cold-shock domain." *Trends Biochem. Sci.* **23**(8): 286-290.
- Graumann, P. L. and M. A. Marahiel (1999). "Cold shock response in *Bacillus subtilis*." *J Mol Microbiol Biotechnol* **1**(2): 203-209.
- Gronenborn, A. M., D. R. Filpula, et al. (1991). "A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G." *Science* **253**(5020): 657-661.
- Hanes, J. and A. Plückthun (1997). "In vitro selection and evolution of functional proteins by using ribosome display." *Proc. Natl. Acad. Sci. USA* **94**(10): 4937-4942.
- Hecky, J. and K. M. Müller (2005). "Structural perturbation and compensation by directed evolution at physiological temperature leads to thermostabilization of beta-lactamase." *Biochemistry* **44**(38): 12640-12654.
- Hendsch, Z. S. and B. Tidor (1999). "Electrostatic interactions in the GCN4 leucine zipper: substantial contributions arise from intramolecular interactions enhanced on binding." *Protein Sci.* **8**(7): 1381-1392.
- Holliger, P., L. Riechmann, et al. (1999). "Crystal structure of the two N-terminal domains of g3p from filamentous phage fd at 1.9 Å: evidence for conformational lability." *J. Mol. Biol.* **288**(4): 649-657.
- Horovitz, A. (1996). "Double-mutant cycles: a powerful tool for analyzing protein structure and function." *Fold. Des.* **1**(6): R121-126.
- Horovitz, A. and A. R. Fersht (1992). "Co-operative interactions during protein folding." *J. Mol. Biol.* **224**(3): 733-740.
- Jaenicke, R. (1996). "How do proteins acquire their three-dimensional structure and stability?" *Naturwissenschaften* **83**(12): 544-554.
- Jiang, W., Y. Hou, et al. (1997). "CspA, the major cold-shock protein of *Escherichia coli*, is an RNA chaperone." *J. Biol. Chem.* **272**(1): 196-202.
- Johansson, M. U., I. M. Frick, et al. (2002). "Structure, specificity, and mode of interaction for bacterial albumin-binding modules." *J. Biol. Chem.* **277**(10): 8114-8120.
- Karlsson, F., C. A. Borrebaeck, et al. (2003). "The mechanism of bacterial infection by filamentous phages involves molecular interactions between TolA and phage protein 3 domains." *J. Bacteriol.* **185**(8): 2628-2634.
- Karshikoff, A. and R. Ladenstein (2001). "Ion pairs and the thermotolerance of proteins from hyperthermophiles: a "traffic rule" for hot roads." *Trends Biochem. Sci.* **26**(9): 550-556.
- Kather, I., C. A. Bippes, et al. (2005). "A stable disulfide-free gene-3-protein of phage fd generated by in vitro evolution." *J. Mol. Biol.* **354**(3): 666-678.

- Koradi, R., M. Billeter, et al. (1996). "MOLMOL: a program for display and analysis of macromolecular structures." *J. Mol. Graph.* **14**(1): 51-55.
- Korkegian, A., M. E. Black, et al. (2005). "Computational thermostabilization of an enzyme." *Science* **308**(5723): 857-860.
- Kuhlman, B., G. Dantas, et al. (2003). "Design of a novel globular protein fold with atomic-level accuracy." *Science* **302**(5649): 1364-1368.
- Kuhlman, B., D. L. Luisi, et al. (1999). "pKa values and the pH dependent stability of the N-terminal domain of L9 as probes of electrostatic interactions in the denatured state. Differentiation between local and nonlocal interactions." *Biochemistry* **38**(15): 4896-4903.
- Kumar, S. and R. Nussinov (2002). "Close-range electrostatic interactions in proteins." *ChemBiochem* **3**(7): 604-617.
- Lazaridis, T. and M. Karplus (2003). "Thermodynamics of protein folding: a microscopic view." *Biophys. Chem.* **100**(1-3): 367-395.
- Lee, K. K., C. A. Fitch, et al. (2002). "Distance dependence and salt sensitivity of pairwise, coulombic interactions in a protein." *Protein Sci.* **11**(5): 1004-1016.
- Leemhuis, H., V. Stein, et al. (2005). "New genotype-phenotype linkages for directed evolution of functional proteins." *Curr. Opin. Struct. Biol.* **15**(4): 472-478.
- Lubkowski, J., F. Hennecke, et al. (1998). "The structural basis of phage display elucidated by the crystal structure of the N-terminal domains of g3p." *Nature Struct. Biol.* **5**(2): 140-147.
- Lubkowski, J., F. Hennecke, et al. (1999). "Filamentous phage infection: crystal structure of g3p in complex with its coreceptor, the C-terminal domain of TolA." *Structure* **7**(6): 711-722.
- Luisi, D. L., C. D. Snow, et al. (2003). "Surface salt bridges, double-mutant cycles, and protein stability: an experimental and computational analysis of the interaction of the Asp 23 side chain with the N-terminus of the N-terminal domain of the ribosomal protein L9." *Biochemistry* **42**(23): 7050-7060.
- Makhatadze, G. I. and P. L. Privalov (1992). "Protein interactions with urea and guanidinium chloride. A calorimetric study." *J. Mol. Biol.* **226**(2): 491-505.
- Makhatadze, G. I. and P. L. Privalov (1993). "Contribution of hydration to protein folding thermodynamics. I. The enthalpy of hydration." *J. Mol. Biol.* **232**(2): 639-659.
- Malakauskas, S. M. and S. L. Mayo (1998). "Design, structure and stability of a hyperthermophilic protein variant." *Nature Struct. Biol.* **5**(6): 470-475.
- Martin, A., I. Kather, et al. (2002). "Origins of the high stability of an in vitro-selected cold-shock protein." *J. Mol. Biol.* **318**(5): 1341-1349.
- Martin, A. and F. X. Schmid (2003). "Evolutionary stabilization of the gene-3-protein of phage fd reveals the principles that govern the thermodynamic stability of two-domain proteins." *J. Mol. Biol.* **328**(4): 863-875.
- Martin, A., F. X. Schmid, et al. (2003). "Proside: a phage-based method for selecting thermostable proteins." *Methods Mol. Biol.* **230**: 57-70.
- Martin, A., V. Sieber, et al. (2001). "In-vitro selection of highly stabilized protein variants with optimized surface." *J. Mol. Biol.* **309**(3): 717-726.
- McCallister, E. L., E. Alm, et al. (2000). "Critical role of beta-hairpin formation in protein G folding." *Nature Struct. Biol.* **7**(8): 669-673.
- Merkel, J. S., J. M. Sturtevant, et al. (1999). "Sidechain interactions in parallel beta sheets: the energetics of cross-strand pairings." *Structure* **7**(11): 1333-1343.
- Minor, D. L., Jr. and P. S. Kim (1994). "Context is a major determinant of beta-sheet propensity." *Nature* **371**(6494): 264-267.
- Minor, D. L., Jr. and P. S. Kim (1996). "Context-dependent secondary structure formation of a designed protein sequence." *Nature* **380**(6576): 730-734.

- Mozo-Villarias, A., J. Cedano, et al. (2003). "A simple electrostatic criterion for predicting the thermal stability of proteins." *Protein Eng.* **16**(4): 279-286.
- Mueller, U., D. Perl, et al. (2000). "Thermal stability and atomic-resolution crystal structure of the *Bacillus caldolyticus* cold shock protein." *J. Mol. Biol.* **297**(4): 975-988.
- Murphy, K. P. and E. Freire (1992). "Thermodynamics of Structural Stability and Cooperative Folding Behavior in Proteins." *Adv. Protein Chem.* **43**: 313-361.
- Myers, J. K., C. N. Pace, et al. (1995). "Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding." *Protein Sci.* **4**(10): 2138-2148.
- Nauli, S., B. Kuhlman, et al. (2001). "Computer-based redesign of a protein folding pathway." *Nature Struct. Biol.* **8**(7): 602-605.
- Nemoto, N., E. Miyamoto-Sato, et al. (1997). "In vitro virus: bonding of mRNA bearing puromycin at the 3'-terminal end to the C-terminal end of its encoded protein on the ribosome in vitro." *FEBS Lett.* **414**(2): 405-408.
- Orencia, M. C., J. S. Yoon, et al. (2001). "Predicting the emergence of antibiotic resistance by directed evolution and structural analysis." *Nature Struct. Biol.* **8**(3): 238-242.
- Pace, C. N. (1992). "Contribution of the hydrophobic effect to globular protein stability." *J. Mol. Biol.* **226**(1): 29-35.
- Pace, C. N., R. W. Alston, et al. (2000). "Charge-charge interactions influence the denatured state ensemble and contribute to protein stability." *Protein Sci.* **9**(7): 1395-1398.
- Park, C. and S. Marqusee (2004). "Probing the high energy states in proteins by proteolysis." *J. Mol. Biol.* **343**(5): 1467-1476.
- Park, C. and S. Marqusee (2005). "Pulse proteolysis: a simple method for quantitative determination of protein stability and ligand binding." *Nat. Methods* **2**(3): 207-212.
- Park, S. H., M. C. Shastry, et al. (1999). "Folding dynamics of the B1 domain of protein G explored by ultrarapid mixing." *Nature Struct. Biol.* **6**(10): 943-947.
- Parsell, D. A. and R. T. Sauer (1989). "The structural stability of a protein is an important determinant of its proteolytic susceptibility in *Escherichia coli*." *J. Biol. Chem.* **264**(13): 7590-7595.
- Pelletier, J. N., K. M. Arndt, et al. (1999). "An in vivo library-versus-library selection of optimized protein-protein interactions." *Nat. Biotechnol.* **17**(7): 683-690.
- Perl, D., U. Mueller, et al. (2000). "Two exposed amino acid residues confer thermostability on a cold shock protein." *Nature Struct. Biol.* **7**(5): 380-383.
- Perl, D. and F. X. Schmid (2001). "Electrostatic stabilization of a thermophilic cold shock protein." *J. Mol. Biol.* **313**(2): 343-357.
- Perl, D., C. Welker, et al. (1998). "Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins." *Nature Struct. Biol.* **5**(3): 229-235.
- Pfeil, W. and P. L. Privalov (1976). "Thermodynamic investigations of proteins. II. Calorimetric study of lysozyme denaturation by guanidine hydrochloride." *Biophys Chem* **4**(1): 33-40.
- Phadtare, S. (2004). "Recent developments in bacterial cold-shock response." *Curr. Issues Mol. Biol.* **6**(2): 125-136.
- Phadtare, S., M. Inouye, et al. (2002). "The nucleic acid melting activity of *Escherichia coli* CspE is critical for transcription antitermination and cold acclimation of cells." *J. Biol. Chem.* **277**(9): 7239-7245.
- Privalov, P. L. (1979). "Stability of proteins: small globular proteins." *Adv. Protein Chem.* **33**: 167-241.
- Privalov, P. L. (1990). "Cold denaturation of proteins." *Crit. Rev. Biochem. Mol. Biol.* **25**(4): 281-305.

- Privalov, P. L. and G. I. Makhatadze (1993). "Contribution of hydration to protein folding thermodynamics. II. The entropy and Gibbs energy of hydration." *J. Mol. Biol.* **232**(2): 660-679.
- Roberts, R. W. and J. W. Szostak (1997). "RNA-peptide fusions for the in vitro selection of peptides and proteins." *Proc. Natl. Acad. Sci. USA* **94**(23): 12297-12302.
- Sanchez-Ruiz, J. M. and G. I. Makhatadze (2001). "To charge or not to charge?" *Trends Biotechnol.* **19**(4): 132-135.
- Sari, N., P. Alexander, et al. (2000). "Structure and dynamics of an acid-denatured protein G mutant." *Biochemistry* **39**(5): 965-977.
- Schindelin, H., M. A. Marahiel, et al. (1993). "Universal nucleic acid-binding domain revealed by crystal structure of the *B. subtilis* major cold-shock protein." *Nature* **364**(6433): 164-168.
- Schindler, T., M. Herrler, et al. (1995). "Extremely rapid protein folding in the absence of intermediates." *Nature Struct. Biol.* **2**(8): 663-673.
- Schmid, F. X. (2005). *Spectroscopic Techniques to Study Protein Folding and Stability. Protein Folding Handbook.* J. Buchner and T. Kiefhaber. Weinheim, WILEY-VCH Verlag GmbH & Co. KGaA. **1**: 22-43.
- Serrano, L. and A. R. Fersht (1989). "Capping and alpha-helix stability." *Nature* **342**(6247): 296-299.
- Sheinerman, F. B. and C. L. Brooks, 3rd (1998). "Calculations on folding of segment B1 of streptococcal protein G." *J. Mol. Biol.* **278**(2): 439-456.
- Sieber, V., A. Plückthun, et al. (1998). "Selecting proteins with improved stability by a phage-based method." *Nat. Biotechnol.* **16**(10): 955-960.
- Smith, C. K. and L. Regan (1995). "Guidelines for protein design: the energetics of beta sheet side chain interactions." *Science* **270**(5238): 980-982.
- Smith, C. K. and L. Regan (1997). "Construction and design of beta-sheets." *Accounts Chem. Res.* **30**(4): 153-161.
- Smith, C. K., J. M. Withka, et al. (1994). "A thermodynamic scale for the beta-sheet forming tendencies of the amino acids." *Biochemistry* **33**(18): 5510-5517.
- Smith, G. P. (1985). "Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface." *Science* **228**(4705): 1315-1317.
- Sobolev, V., A. Sorokine, et al. (1999). "Automated analysis of interatomic contacts in proteins." *Bioinformatics* **15**(4): 327-332.
- Spencer, D. S., K. Xu, et al. (2005). "Effects of pH, salt, and macromolecular crowding on the stability of FK506-binding protein: an integrated experimental and theoretical study." *J. Mol. Biol.* **351**(1): 219-232.
- Steipe, B., B. Schiller, et al. (1994). "Sequence statistics reliably predict stabilizing mutations in a protein domain." *J. Mol. Biol.* **240**(3): 188-192.
- Stengele, I., P. Bross, et al. (1990). "Dissection of functional domains in phage fd adsorption protein. Discrimination between attachment and penetration sites." *J. Mol. Biol.* **212**(1): 143-149.
- Tanford, C. (1970). "Protein denaturation. C. Theoretical models for the mechanism of denaturation." *Adv. Protein Chem.* **24**: 1-95.
- Tanford, C. (1978). "The hydrophobic effect and the organization of living matter." *Science* **200**(4345): 1012-1018.
- Tashiro, M. and G. T. Montelione (1995). "Structures of bacterial immunoglobulin-binding domains and their complexes with immunoglobulins." *Curr. Opin. Struct. Biol.* **5**(4): 471-481.
- Torrez, M., M. Schultehenrich, et al. (2003). "Conferring thermostability to mesophilic proteins through optimized electrostatic surfaces." *Biophys. J.* **85**(5): 2845-2853.

- Vogt, G., S. Woell, et al. (1997). "Protein thermal stability, hydrogen bonds, and ion pairs." *J. Mol. Biol.* **269**(4): 631-643.
- Voigt, C. A., S. L. Mayo, et al. (2001). "Computational method to reduce the search space for directed protein evolution." *Proc. Natl. Acad. Sci. USA* **98**(7): 3778-3783.
- Wassenberg, D., C. Welker, et al. (1999). "Thermodynamics of the unfolding of the cold-shock protein from *Thermotoga maritima*." *J. Mol. Biol.* **289**(1): 187-193.
- Wintrobe, P. L. and F. H. Arnold (2000). "Temperature adaptation of enzymes: lessons from laboratory evolution." *Adv. Protein Chem.* **55**: 161-225.
- Yang, A. S., B. Hitz, et al. (1996). "Free energy determinants of secondary structure formation: III. beta-turns and their role in protein folding." *J. Mol. Biol.* **259**(4): 873-882.
- Yang, A. S. and B. Honig (1995a). "Free energy determinants of secondary structure formation: I. alpha-Helices." *J. Mol. Biol.* **252**(3): 351-365.
- Yang, A. S. and B. Honig (1995b). "Free energy determinants of secondary structure formation: II. Antiparallel beta-sheets." *J. Mol. Biol.* **252**(3): 366-376.
- Zeeb, M. and J. Balbach (2003). "Single-stranded DNA binding of the cold-shock protein CspB from *Bacillus subtilis*: NMR mapping and mutational characterization." *Protein Sci.* **12**(1): 112-123.
- Zhou, H. X. and F. Dong (2003). "Electrostatic contributions to the stability of a thermophilic cold shock protein." *Biophys. J.* **84**(4): 2216-2222.

5. Publikationsliste

- A.** Michael Wunderlich, Andreas Martin and Franz X. Schmid.
Stabilization of the cold shock protein CspB from *Bacillus subtilis* by evolutionary optimization of Coulombic interactions.
Journal of Molecular Biology. **347**, 1063-1076 (2005)
- B.** Michael Wunderlich and Franz X. Schmid
The correlation between protein stability and dipole moment: a critical test
Protein Engineering, Design & Selection. **19** (8), 355-358 (2006)
- C.** Michael Wunderlich, Andreas Martin, Claudia A. Staab and Franz X. Schmid.
Evolutionary protein stabilization in comparison with computational design.
Journal of Molecular Biology. **351**, 1160-1168 (2005)
- D.** Michael Wunderlich, Klaas E. Max, Udo Heinemann and Franz X. Schmid.
Thermodynamical and structural analysis of G β 1 variants derived by *in vitro* evolution and computational design.
Manuskript in Vorbereitung
- E.** Michael Wunderlich and Franz X. Schmid.
In vitro Evolution of a Hyperstable G β 1 Variant.
Journal of Molecular Biology. **363**, 545-557 (2006)
- F.** Michael Wunderlich and Franz X. Schmid.
Energetic determinants for a hyperstable variant of G β 1 obtained from *in-vitro* evolution.
Manuskript in Vorbereitung

6. Darstellung des Eigenanteils

- A. Die Arbeit beschreibt die *in vitro* Evolution des Kälteschockproteins CspB aus *Bacillus subtilis* mit Hilfe des Selektionssystems *Proside*. In das Selektionssystem wurde ich von Andreas Martin eingeführt und er stellte ein aus seinen Arbeiten stammendes Phagenkonstrukt, das als Ausgangspunkt der Arbeit diente, zur Verfügung. Alle in der Arbeit dargestellten Ergebnisse wurden von mir erarbeitet und die Publikation wurde von Franz X. Schmid und mir formuliert.
- B. Alle dargestellten Ergebnisse wurden von mir erarbeitet. Die Publikation wurde von Franz X. Schmid und mir formuliert.
- C. Die Arbeit beschreibt die *in vitro* Evolution der $\beta 1$ Domäne des Streptokokkenproteins G. Die Selektion der Bibliotheken 1 und 2 wurde von Claudia A. Staab im Rahmen ihrer Diplomarbeit durchgeführt und die der Bibliothek 3 von Andreas Martin. Die restlichen dargestellten Ergebnisse wurden von mir erarbeitet. Die Publikation wurde von Franz X. Schmid und mir formuliert.
- D. Alle dargestellten Ergebnisse zur thermodynamischen Charakterisierung von Varianten der $\beta 1$ Domäne des Streptokokkenproteins G wurden von mir erarbeitet. Die Kristallisation und die Strukturbestimmung zweier dieser Varianten wurde von Klaas Max im Labor von Udo Heinemann mit von mir zur Verfügung gestelltem Protein durchgeführt. Das Manuskript wurde von Franz X. Schmid und mir formuliert.
- E. Alle dargestellten Ergebnisse wurden von mir erarbeitet. Das Manuskript wurde von Franz X. Schmid und mir formuliert.
- F. Alle dargestellten Ergebnisse wurden von mir erarbeitet. Das Manuskript wurde von Franz X. Schmid und mir formuliert.

7. Teilarbeiten

7.1 Teilarbeit A

Michael Wunderlich, Andreas Martin and Franz X. Schmid.

Stabilization of the cold shock protein CspB from *Bacillus subtilis* by evolutionary optimization of Coulombic interactions.

Journal of Molecular Biology. **347**, 1063-1076 (2005)

7.2 Teilarbeit B

Michael Wunderlich and Franz X. Schmid

The correlation between protein stability and dipole moment: a critical test

Protein Engineering, Design & Selection. **19** (8), 355-358 (2006)

7.3 Teilarbeit C

Michael Wunderlich, Andreas Martin, Claudia A. Staab and Franz X. Schmid.

Evolutionary protein stabilization in comparison with computational design.

Journal of Molecular Biology. **351**, 1160–1168 (2005)

7.4 Teilarbeit D

Thermodynamical and structural analysis of G β 1 variants derived by *in vitro* evolution and computational design

Michael Wunderlich, Klaas E. Max, Udo Heinemann, and Franz X. Schmid

Manuskript in Vorbereitung

Running title: G β 1 stabilization by *in vitro* evolution and computational design

Thermodynamical and structural analysis of G β 1 variants derived by *in vitro* evolution and computational design

Michael Wunderlich¹, Klaas E. Max², Udo Heinemann², and Franz X. Schmid^{1*}

¹*Laboratorium für Biochemie und Bayreuther Zentrum für Molekulare Biowissenschaften, Universität Bayreuth, D-95440 Bayreuth, Germany*

²*Forschungsgruppe Kristallographie, Max-Delbrück-Zentrum für Molekulare Medizin, Robert-Ressle-Str. 10, D-13125 Berlin, Germany*

**Corresponding author*

Franz X. Schmid

Tel: ++49 921 553660

Biochemie

Fax: ++49 921 553661

Universität Bayreuth

E-mail: fx.schmid@uni-bayreuth.de

D-95440 Bayreuth, Germany

Abbreviations used: G β 1, β 1 domain of the streptococcal protein G; G β 1c3, β 1 domain of the streptococcal protein G including the three core mutations Y3F/L71/V39I; G β 1c3-IIEF, the most stable selected variant of G β 1 including the described core mutations and the four optimized boundary positions T16I/T18I/T25E/V29F; G β 1c3-ILEK, the calculated variant of G β 1 including the described core mutations and the four calculated replacements at the four boundary positions T16I/T18L/T25E/V29K; G β 1-IIEF and G β 1-ILEK, same as above without the three core mutations; T_M , midpoint of a thermal unfolding transition; GdmCl, guanidinium chloride.

(Abstract)

The positions 16, 18, 25 and 29 of the β 1 domain of the streptococcal protein G (G β 1) carry a high potential for stabilization. Computational design and experimental *in vitro* selection both yielded collections of stabilized variants with mutations at these four positions. Good agreement was found for positions 16 and 25 with preferences for Ile at 16 and Glu at 25. For position 18 the calculations yielded Leu for six out of the ten variants, and for position 29 a Lys was found three times. In the 21 variants from the *in-vitro* selection Leu16 or Lys29 were not found. To elucidate the molecular origin of this discrepancy, we measured the contributions to stability of the individual residues, determined the structures of the best variant from the *in-vitro* selections (T16I/T18I/T25E/V29F) and of the variant with Leu18 and Lys29 (T16I/T18L/T25E/V29K) by X-ray crystallography. The mutations T16I and T25E stabilize G β 1 by 5.6 and 2.0 kJ mol⁻¹. In the selected variant, T18I and V29F stabilize by 3.8 and 1.7 kJ mol⁻¹. The corresponding mutations that were found in several computed variants, are not optimal. T18L stabilizes G β 1 by only 1.4 kJ mol⁻¹, and V29K destabilizes by 2.2 kJ mol⁻¹. Double-mutant analyses indicate that the contributions to stability at the four positions are largely additive. The crystal structures of G β 1 variants show that the backbone of the protein is not significantly changed by the mutations. The structures also show that it is very difficult to correlate structural changes with changes in stability. The aliphatic moiety of the Lys29 side chain in the computed variant apparently forms hydrophobic contacts with Leu18, and the amino group seems to form a salt bridge with Glu25. The thermodynamic analysis, however, fails to detect such interactions. This probably explains why the destabilizing effect of the V29K mutation could not be detected in the computational design approach, which ultimately is based on information derived from protein crystal structures.

Keywords: protein stability, *in-vitro* selection, computational design, mutant cycle analysis streptococcal protein G

Introduction

Natural proteins are only marginally stable. Most spontaneous mutations are unfavorable for protein stability, and therefore only the minimal stability that is required to maintain the cellular function of a protein is maintained during evolution.¹ This is encouraging for protein engineering, because it should be possible to optimize protein sequences for enhanced stability, by computational approaches or by directed evolution¹⁻¹⁰.

In pioneering work, Mayo and coworkers used an objective computer algorithm to design stabilized variants of the β 1 domain of protein G ($G\beta$ 1).^{11; 12} $G\beta$ 1 is a small protein, and its chain of 56 amino acids folds into a four-stranded antiparallel β -sheet and a single α -helix, which stretches across the β sheet (Figure 1).¹³⁻¹⁵ The calculations identified three core¹¹ and four partially exposed (“boundary”) positions as good candidate positions for stabilizing $G\beta$ 1.¹² The sequence optimization at the three core positions led to a modest increase in stability. The optimization at the four boundary positions 16, 18, 25 and 29 led to ten computed sequences with closely spaced Monte-Carlo energies, and the best one showed an experimental gain in stability of 14.6 kJ mol⁻¹.¹²

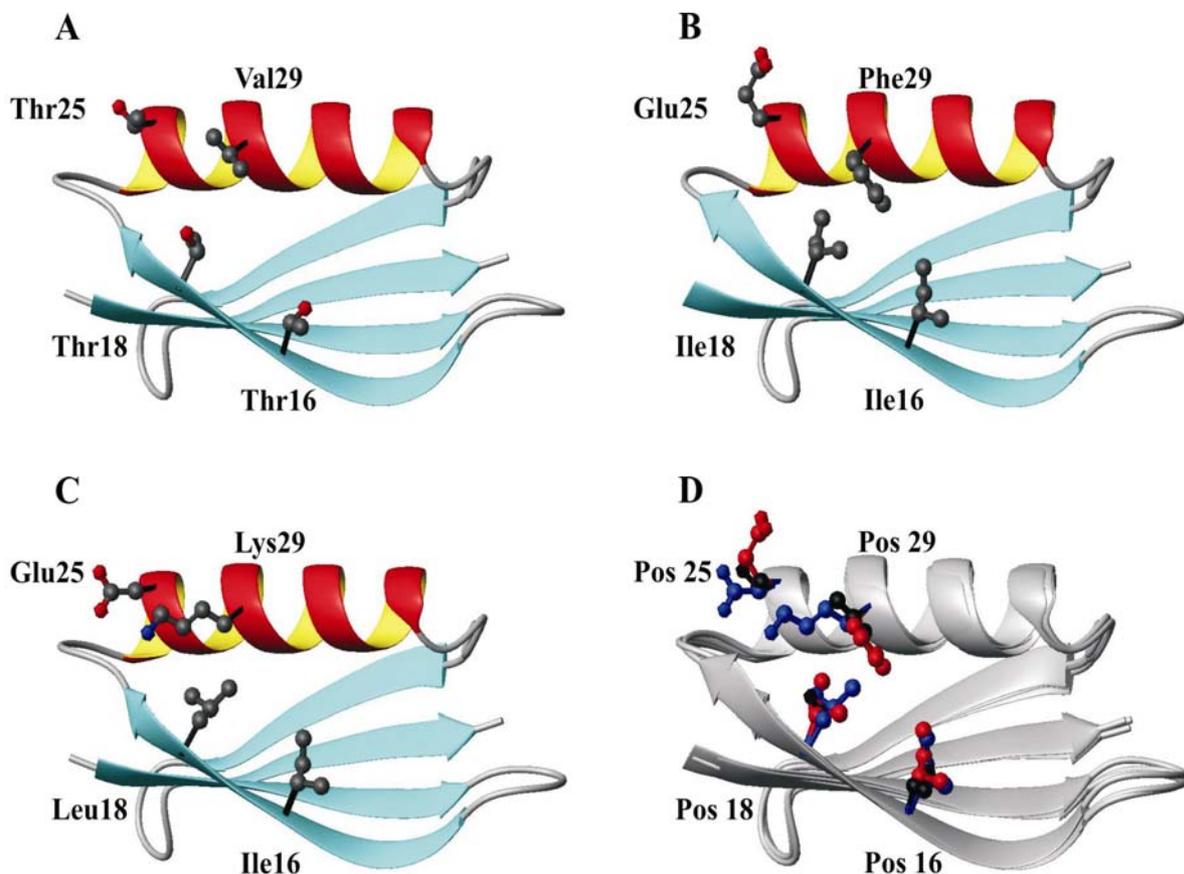


Figure 1. Tertiary structure of A) wild-type $G\beta$ 1, B) $G\beta$ 1c3-IIEF, C) $G\beta$ 1c3-ILEK and D) superposition of wild-type $G\beta$ 1, $G\beta$ 1c3-IIEF and $G\beta$ 1c3-ILEK. The residues at position 16, 18, 25, and 29 are shown in ball-and-stick representation. Sidechain oxygen atoms in Figure 1A, 1B, and 1C are colored red, nitrogen atoms blue and carbon atoms black. In Figure 1D the sidechains of wildtype $G\beta$ 1 are colored black, of $G\beta$ 1c3-IIEF red, and $G\beta$ 1c3-ILEK blue. The Figure was prepared using PDB file 1PGB¹⁵ for wild-type $G\beta$ 1, the determined data for $G\beta$ 1c3-IIEF and $G\beta$ 1c3-ILEK, and MOLMOL.¹⁶

We used the experimental *in vitro* selection method *Proside*¹⁷ to stabilize G β 1. *Proside* links the increased protease resistance of stabilized protein variants with the infectivity of a filamentous phage. To evaluate the performance of *Proside* relative to computational protein stabilization, we employed the same four boundary positions of G β 1 (16, 18, 25 and 29) that had been identified in the computational work, randomized the corresponding codons, and selected stabilized variants of G β 1 from the corresponding phage libraries.¹⁸ More than 20 strongly stabilized variants were obtained by this approach and the two best ones were slightly better than the best variant from the computational design project.

When the ten computed variants were compared with the ten best variants from the *in-vitro* selections, a very good agreement was found for the amino acid replacements at the positions 16 and 25. Both approaches identified Ile16 and Glu25 as the preferred residues for these positions. Significant differences were found, however, for the positions 18 and 29. Six of the ten computed sequences contained a Leu at position 16, which was never found in the *Proside* selections, although Leu is coded by six different base triplets. At position 29 only aromatic residues were found in the selections, but only two of the ten computed sequences contained an aromatic residue at this position. Instead, a positively charged residue, Lys29, occurred in three of the ten computed sequences. Lys29, in turn, was never found in the selections. When the corresponding variants with Leu16 and/or Lys29 were produced by directed mutagenesis, it turned out that these two residues indeed destabilize G β 1.¹⁸ Seven of the ten calculated variants contain Leu18 and / or Lys29.

To understand, why the results of the computational and the experimental stabilization of G β 1 diverged at the positions 18 and 29 we analysed two variants of G β 1. Both contain the two stabilizing mutations T16I and T25E. The first variant contains, in addition, the two suboptimal substitutions T18L and V29K as found in the calculations. We use the shorthand notation G β 1-ILEK for this variant (and G β 1c3-ILEK when it contains also the three stabilizing core mutations Y3F, L7I, and V39I from the initial computational design¹²). The second variant is identical with the best variant that was found in the *Proside* selections. It contains, in addition to T16I and T25E, the mutations T18I and V29F and is denoted as G β 1-IIEF or G β 1c3-IIEF when the three core mutations are also present.

We determined high-resolution crystal structures for these two variants, and measured how all the mutations contribute individually to the Gibbs free energy of stabilization. Potential interaction energies between the mutated residues were determined from several double mutant analyses^{19; 20}.

Results

Structural analysis

The crystal structures of the G β 1c3-IIEF and G β 1c3-ILEK variants were solved by molecular replacement at resolutions of 1.7 Å and 1.35 Å (Table 1), respectively, and refined to *R* values better than 21% and *R*_{free} values better than 23%. The crystals are well packed with solvent contents below 35% and 45%, respectively, and the two structures contain one protein molecule per asymmetric unit. The estimated maximum errors as determined by Luzatti plots are below 0.2 Å. The high resolution of the two crystal structures allows to assess the effects of the mutations with great confidence. The tertiary structures of wild-type G β 1¹⁵, G β 1c3-IIEF, and G β 1c3-ILEK are shown in Figure 1A, 1C, and 1D, respectively. The

superposition of the structures of the two variants (Figure 1B) shows that the protein backbone is virtually unchanged by the mutations. The RMS-value for all backbone atoms is 0.4 Å.

Table 1. Data collection and refinement statistics

Data collection		Gβ1c3-IIEF	Gβ1c3-ILEK
Wavelength (Å)		1.5418	---
Resolution (Å)		19 – 1.60	50 – 1.34
Last shell (Å)		1.70 – 1.60	1.39 – 1.34
Space group		R32	P4
Temperature (K)		110	110
Detector		MAR IP 345	MAR CCD 165
Unit-cell parameters	<i>a</i> (Å)	48.41	52.10
	<i>b</i> (Å)	48.41	52.10
	<i>c</i> (Å)	107.33	22.77
Unique reflections	(last shell)	6358 (918)	13,293 (770)
<i>I</i> / σ (<i>I</i>)	(last shell)	24.65 (5.99)	30.9 (2.03)
Data completeness %	(last shell)	95.3% (84.5%)	94.6% (56.4%)
<i>R</i> _{cryst} %	(last shell)	3.4% (30.1%) ^a	5.3% (36.5%) ^b
Refinement			
Resolution (Å)		19 – 1.7	19 – 1.35
Working set		5296	12044
Free set		256 (4.6%)	635 (5 %)
<i>R</i> _{work} / <i>R</i> _{free} (%)		20.50 / 22.91	15.83 / 19.52
Number of non-hydrogen atoms		481	502
Number of protein molecules		1	1
Number of water molecules		38	56
Mean <i>B</i> factor (Å ²)		30.16	20.64
RMSD:			
bond lengths (Å)		0.021	0.026
bond angles (°)		1.81	2.43
torsion angles (°)		8.18	5.64
planarity (Å)		0.008	0.006
Ramachandran statistics			
Residues in allowed regions		96.0%	96.0%
Residues in add. allowed regions		4.0%	4.0%

^a XDS reports *R*_{meas}, a redundancy independent *R* factor, which correlates intensities from symmetry related reflections (36).

^b HKL2000 reports *R*_{sym},

^c $R_{work,free} = \sum \frac{|F_{obs}| - |F_{calc}|}{|F_{obs}|}$, where the working and free *R* factors are calculated using the working and free reflection sets, respectively. The free reflections were held aside throughout refinement.

Contributions of the individual mutations to stability

In the previous optimizations of Gβ1 by computation¹² or by *in-vitro* selection¹⁸ the variant with the stabilizing core mutations Y3F, L7I, and V39I (Gβ1c3) served as the starting molecule. At the optimal pH (pH 5.5) Gβ1c3-IIEF shows a *T*_M value far above 100 °C. In our previous work¹⁸, we therefore compared the thermodynamic parameters of all selected

variants in the presence of 2.5 M GdmCl and at pH 7.0. Here we analyzed the contributions to stability of the individual mutations by using wild-type G β 1 as the reference molecule. Wild-type G β 1 (without the three core mutations) is 6.0 kJ mol⁻¹ less stable than G β 1c3 and shows a T_M value of 77.0 °C (at pH 7.0). Thus, the use of the ionic denaturant GdmCl could be avoided in the thermal unfolding experiments of all single mutants, and, in addition, the thermal stabilities could be measured as a function of the NaCl concentration, to uncover potential contributions of Coulombic interactions to the stabilities of the variants.

The T_M values of G β 1-IIEF and G β 1-ILEK are increased by 17.3 and 10.2 °C, relative to wild-type G β 1 (Figure 2), which is equivalent to stabilizations by 14.2 and 8.1 kJ mol⁻¹, respectively (Table 2). These values are very similar to those obtained previously in the presence of the core mutations¹⁸.

To determine the individual contributions to stability of the mutations at the four positions 16, 18, 25, and 29, all single mutants were produced and their thermal stabilities determined in the absence and in the presence of 2 M NaCl (Table 2). For the mutations that involve charge changes (T25E and V29K) the stabilities were also measured at intermediate NaCl concentrations (Figure 3). The screenable parts of the $\Delta\Delta G_D$ values were generally smaller than 0.8 kJ/mol (Table 2), suggesting that coulombic effects contribute little to the stability changes caused by these mutations.

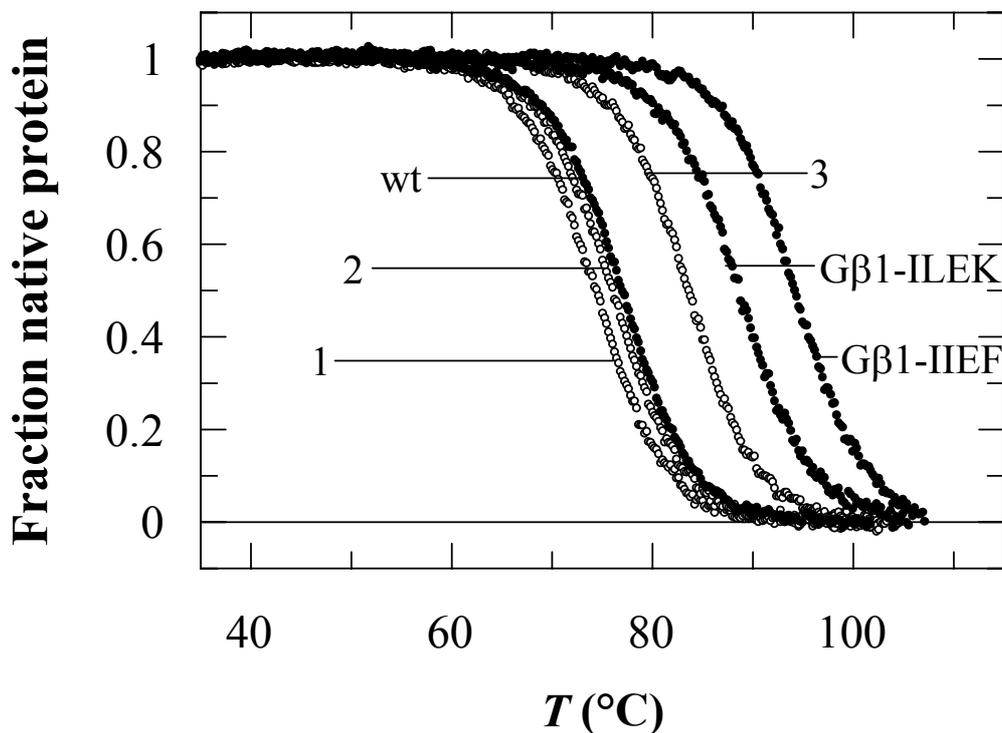


Figure 2. Thermal unfolding transitions of wild-type G β 1, the most stable selected variant G β 1-IIEF, the calculated variant G β 1-ILEK (all filled circles ●) and the variants V29K (1 ○), T18I/V29F (2○) and T18L/V29K (3 ○). The transitions of 4 μ M protein were measured in 0.1 M sodium cacodylate-HCl (pH 7.0) by circular dichroism at 222.6 nm. The fractions of native protein as obtained after two-state analysis are shown as a function of temperature. The corresponding thermodynamic parameters are given in Table 2.

Table 2. Stability data for the variants of G β 1

Variant	T_M^a (°C)	ΔH_D^b (T_M) (kJ mol ⁻¹)	ΔG_D^c (80 °C) (kJ mol ⁻¹)	$\Delta\Delta G_D^d$ (kJ mol ⁻¹)		
				0 M NaCl (total)	0-2 M NaCl (coulombic)	2 M NaCl (non-polar)
wt	77.0	285	-2.5	---	---	---
T16I	83.7	305	3.1	5.6	0.3	5.3
T18I	81.6	296	1.3	3.8	0.3	3.5
T18L	78.7	293	-1.1	1.4	0.3	1.1
T25E	79.4	302	-0.5	2.0	0.8	1.2
V29F	79.1	302	-0.8	1.7	0.3	1.4
V29K	74.4	282	-4.7	-2.2	-0.7	-1.5
T16I/T18I	90.0	303	7.8	10.3	0.1	10.2
T16I/T18L	87.0	297	5.6	8.1	0.1	8.0
T18I/V29F	83.6	305	3.0	5.5	-0.3	5.8
T18L/V29K	76.1	286	-3.3	-0.6	-0.4	-0.2
G β 1-IIIEF (T16I/T18I/ T25E/V29F)	94.3	330	11.7	14.2	1.1	13.1
G β 1-ILEK (T16I/T18L/ T25E/V29K)	87.2	294	5.6	8.1	-0.3	8.4

The thermodynamic parameters were derived from thermal unfolding transitions as shown in Figure 2. The accuracies of T_M and ΔH_D are 0.2 degrees C and 5 kJ mol⁻¹, respectively. The accuracy of ΔG_D (80 °C) is correlated with T_M . For variants with T_M close to 80 °C the accuracy is ≈ 0.3 kJ mol⁻¹, for variants with $75 \text{ °C} \geq T_M \geq 85 \text{ °C}$ the accuracy is ≈ 0.5 kJ mol⁻¹.

^a T_M is the midpoint of the thermal unfolding transitions.

^b ΔH_D is the enthalpy of unfolding at T_M .

^c ΔG_D is the change in Gibbs free energy of unfolding at 80 °C.

^d $\Delta\Delta G_D$ is the change in Gibbs free energy relative to the wild-type protein. $\Delta\Delta G_D$ at 0 M NaCl is the total change, $\Delta\Delta G_D$ at 2 M NaCl represents the non-polar contribution, and the difference between 0 M and 2 M NaCl represents the Coulombic contribution to $\Delta\Delta G_D$.

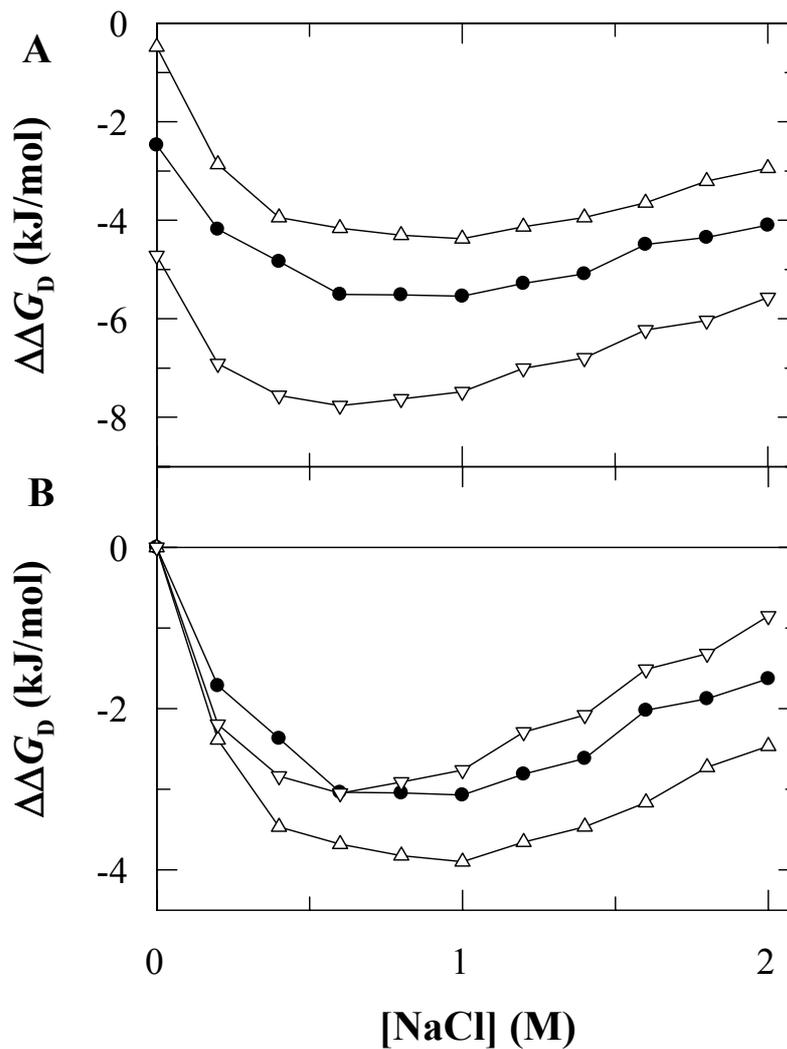


Figure 3. Dependence on NaCl concentration of the change in free energy of unfolding at 80 °C. (A) Absolute values of $\Delta\Delta G_D$ at 80 °C and (B) $\Delta\Delta G_D$ at 80 °C relative to the value measured in the absence of NaCl for wild-type G β 1 (●), G β 1 T25E (Δ) and G β 1 V29K (∇) are shown.

The positions 16 and 25

Ile16 and Glu25 were found as the preferred residues by both computation and *in-vitro* selection. As a single mutation, T25E stabilizes the wild-type protein by 2.4 °C or 2.0 kJ mol⁻¹ (Table 2). About 0.8 kJ/mol of this small stabilization originates from screenable coulombic interactions (Table 2).

The T16I mutation is strongly stabilizing. It increases T_M by 6.7 degrees and ΔG_D by 5.6 kJ mol⁻¹ (Table 2). The extent of stabilization is independent of the NaCl concentration, suggesting that it originates predominantly from improved hydrophobic interactions of Ile16. In the selections a Leu residue was also found repeatedly for position 16. It is slightly less stabilizing than Ile.¹⁸

The positions 18 and 29

The strongest differences between the ten computed and the ten best selected G β 1 variants were observed at the positions 18 and 29.¹⁸ Position 18 is occupied by a Leu residue in six out of ten computed sequences, but a Leu18 was never found in the *in vitro* selections. In fact, Leu should be over-represented in the *Proside* libraries, because it is coded by six different base triplets. Instead, variants with an Ile18 were selected most frequently from these libraries. As a single mutation, T18I stabilizes G β 1 by 3.8 kJ mol⁻¹, T18L, however, stabilizes G β 1 by only by 1.4 kJ mol⁻¹ (Table 2). In both cases, the extent of stabilization is independent of the salt concentration, as expected. When, in the most stable variant G β 1c3-IIEF, Ile18 was changed to Leu18 the stability decreased by 4.7 kJ mol⁻¹.¹⁸ This is almost twofold higher than the value observed for the corresponding single mutations in wild-type G β 1 (Table 2), suggesting that the stability contributions of Ile and Leu at position 18 are slightly context dependent.

For position 29 only aromatic residues were found by *in-vitro* selection. By computation mostly Ile was found, which is slightly less stabilizing than aromatic residues at this position.¹⁸ Lys29 occurred in three out of the ten computed sequences, but was never found by *in-vitro* selection.¹⁸ As single mutations, V29F stabilizes G β 1 by 1.7 kJ mol⁻¹, V29K destabilizes by 2.2 kJ mol⁻¹. The stabilization by Phe29 is salt-independent, the destabilization by Lys29 is reduced to 1.5 kJ/mol in 2 M NaCl, suggesting that at least one third of the destabilizing effect Lys29 originates from unfavorable coulombic interactions of this residue.

Interactions between the residues at the positions 16, 18 and 29

The positions 16, 18, 25, and 29 are close to each other on the same side of G β 1 (Figure 1). To identify potential pairwise interactions between the residues at these positions, we performed double mutant analyses^{19; 20} for both the computed and the selected residues. Positions 18 and 29 are close to each other, and in the wild-type protein the distance between the C $^{\beta}$ atoms of Thr18 and Val29 is only 5.6 Å. In the selected variant (IIEF) these two positions are occupied by Ile18 and Phe29. The double mutation T18I/V29F increased T_M by 6.6 °C and ΔG_D by 5.5 kJ mol⁻¹, equivalent to the sum of the contributions of the single mutations (Figure 4). This indicates that there is no energetic interaction between the two residues.

In the ILEK variant Leu18 and Lys29 are present at these two positions. The double mutation T18L/V29K decreased the T_M value by 0.9 °C (Figure 2) and destabilized G β 1 by 0.6 kJ mol⁻¹ (Table 2). The double mutant analysis gave an insignificant interaction energy of 0.2 kJ mol⁻¹ (Figure 4), suggesting that the stabilizing effect of Leu18 and the destabilizing effect of Lys 29 in the computed variant are also independent of each other.

We also searched for interactions between the residues at the positions 16 and 18. Here the distance between the C $^{\beta}$ atoms is 6.8 Å. In the selected variant (IIEF) these two positions are occupied by Ile residues, and the double mutant T16I/T18I is stabilized by 13.0 degrees and 10.3 kJ mol⁻¹, respectively (Table 2). The sum of the stability contributions of the single mutations T16I and T18I is 9.4 kJ mol⁻¹, suggesting that, relative to the wild-type protein, Ile16 and Ile18 engage in a weakly stabilizing interaction that contributes an additional 0.9 kJ mol⁻¹ to stability (Figure 4). The double mutant analysis for the T16I and T18L

mutations (as present in the ILEK variant) gave an almost identical favorable interaction energy of 1.1 kJ mol^{-1} (Figure 4).

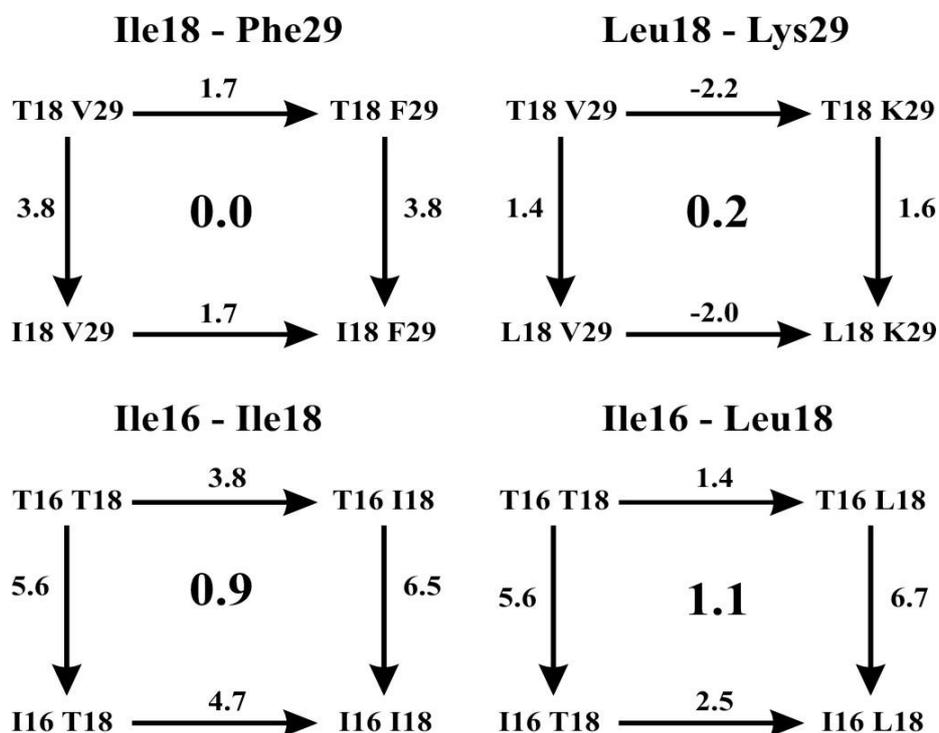


Figure 4. Double mutant cycles with amino acids residues at position 18 and 29 (upper row) and position 16 and 18 (lower row) of the most stable selected variant G β 1-IIEF (left side) and the calculated variant G β 1-ILEK (right side). For every single mutation, the $\Delta\Delta G_D$ values at 80°C (kJ mol^{-1}) are shown and in the center of every cycle, the coupling energy (kJ mol^{-1}) between the respective residues is given.

Addition of the individual contributions of the single mutations and the interaction energies from the double mutant cycles in Figure 4 gives values of 14.0 kJ mol^{-1} for G β 1-IIEF and of 8.1 kJ mol^{-1} for G β 1-ILEK. These summed values agree well with the values measured for the corresponding fourfold mutants (Table 2). This suggests that, apart from the weak interaction between the residues at the positions 16 and 18 (Figure 4), there are no additional interactions between the residues at the positions 16, 18, 25, and 29.

Discussion

Mayo and coworkers¹² identified the four boundary positions 16, 18, 25, and 29 as promising targets for stabilizing G β 1. They are close to each other in the crystal structures of G β 1, but the changes in stability caused by mutations at these sites are additive. The backbone conformations of wild-type G β 1 and the mutated variants are identical, which facilitates the structural interpretation of the observed changes in stability and the comparison with theory. The individual sites differ in their potential for stabilization. The mutations at the positions 16 and 18 lead to strong increases in stability, the stabilizations attained at the positions 25 and 29 by computational design or by *in-vitro* selection were much smaller.

A mutation can decrease the conformational stability of a protein indirectly by stabilizing the unfolded state^{21; 22}, typically by reducing the local charge density in the unfolded protein. The V29K mutation increases the positive charge density in the Lys28-Val29-Phe30-Lys31 chain segment. This should destabilize unfolded G β 1 and, thus increase the stability of the folded state. However, the V29K mutation destabilizes folded G β 1, and therefore an electrostatic destabilization of the unfolded protein by the V29K appears unlikely.

A protein can also be stabilized by optimizing its secondary structure. The positions 16 and 18 are located in strand β 2, and, in wild-type G β 1, both are occupied by Thr, which shows a high β -sheet propensity.^{23; 24} In the selected variant (IIEF) these two Thr residues are replaced by Ile, with a similarly high β -sheet propensity. Differing secondary structure preferences thus cannot explain the strong stabilizations caused by the Thr \rightarrow Ile replacements at positions 16 and 18. They can explain, however, the lower stability of the ILEK variant with a Leu18, because Leu shows a significantly lower β -sheet propensity than Ile^{23; 24}.

Thr25 is in the first turn of the α -helix of G β 1. The stabilization by the T25E mutation is probably caused by a favorable interaction of Glu25 with the dipole of this helix²⁵. This interaction is partially screened by salt (Table 2). The strong preference of Phe29 over Lys29 cannot be explained by differences in helix propensity.

Structural interpretation of the stability data

In the following we interpret the thermodynamic consequences of the mutations on the basis of the three-dimensional structures of wild-type G β 1¹⁵ and the IIEF and ILEK variants. Since this discussion is based only on the structure of the folded proteins it might provide an incomplete picture.

Wild-type G β 1 contains a cleft at its surface (see Figure 5A), which is lined by the side chains of Thr16, Thr18, Val29, Tyr33, and Phe30 (see Figure 5B). In the IIEF mutant the side chains of Ile16, Ile18 and Phe29 are larger and more hydrophobic than the corresponding side chain in the wild-type protein, and, as a consequence, the invagination is filled (Figure 5C). In the ILEK mutant the side chain of Lys29 adopts a different orientation than Phe29 in the IIEF mutant (see Figure 1D). As a consequence, in the variant ILEK the cleft is only partially filled (Figure 5D).

Hole filling and improved packing become evident when the extent of intramolecular contact areas in the wild-type protein and the ILEK and IIEF variants are compared. Ile16, which is strongly stabilizing and present in both variants, shows the same orientation in the two structures (Figure 1B). In the wild-type protein, Thr16 forms contact areas (calculated by ref.²⁷) of 6 \AA^2 with Val30 and of 35 \AA^2 with Tyr33. For Ile16 in the IIEF mutant these values increase 23 \AA^2 for the contact with Phe30 and to 54 \AA^2 for the contact with Tyr33. Filling of the invagination is also reflected in the solvent accessibility of Tyr33. It decreases from 79 \AA^2 in the wild-type protein to 65 \AA^2 in the ILEK variant and 51 \AA^2 in the IIEF variant.

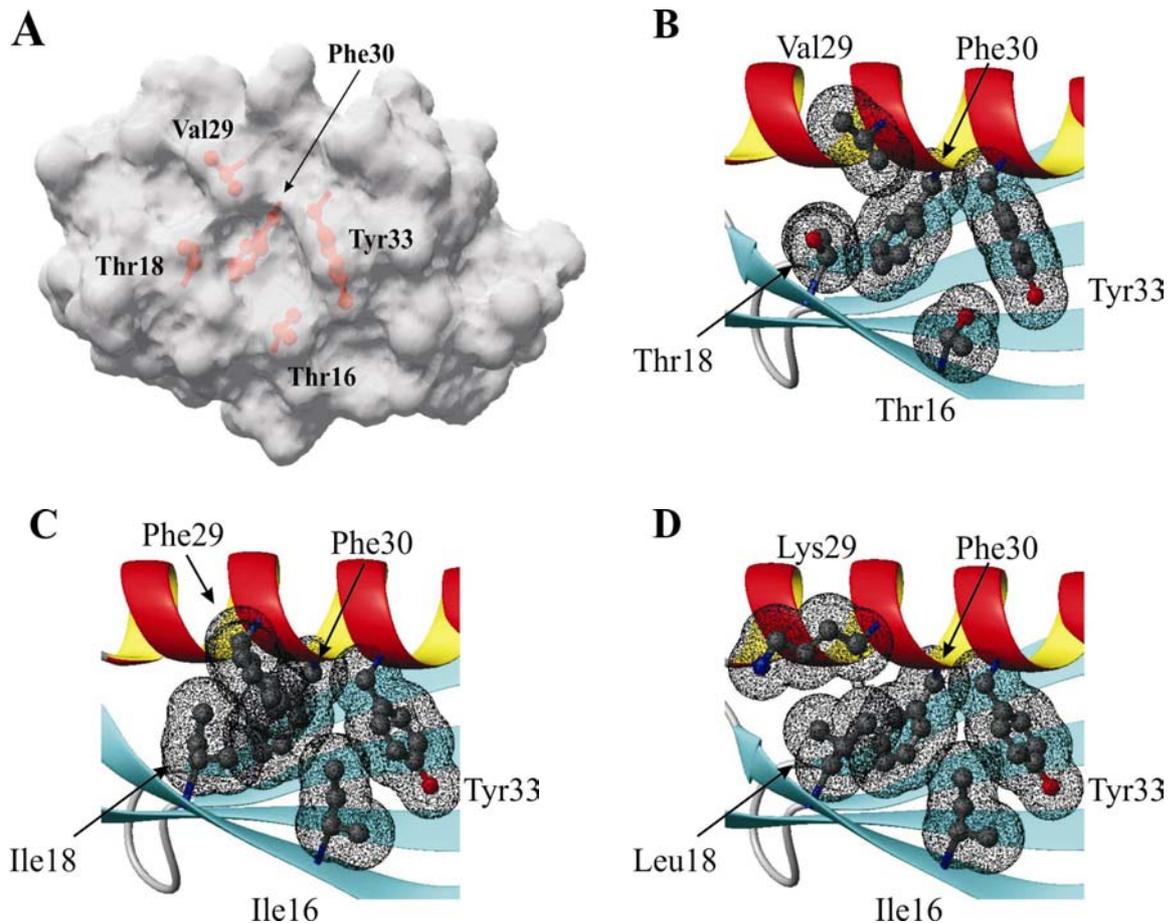
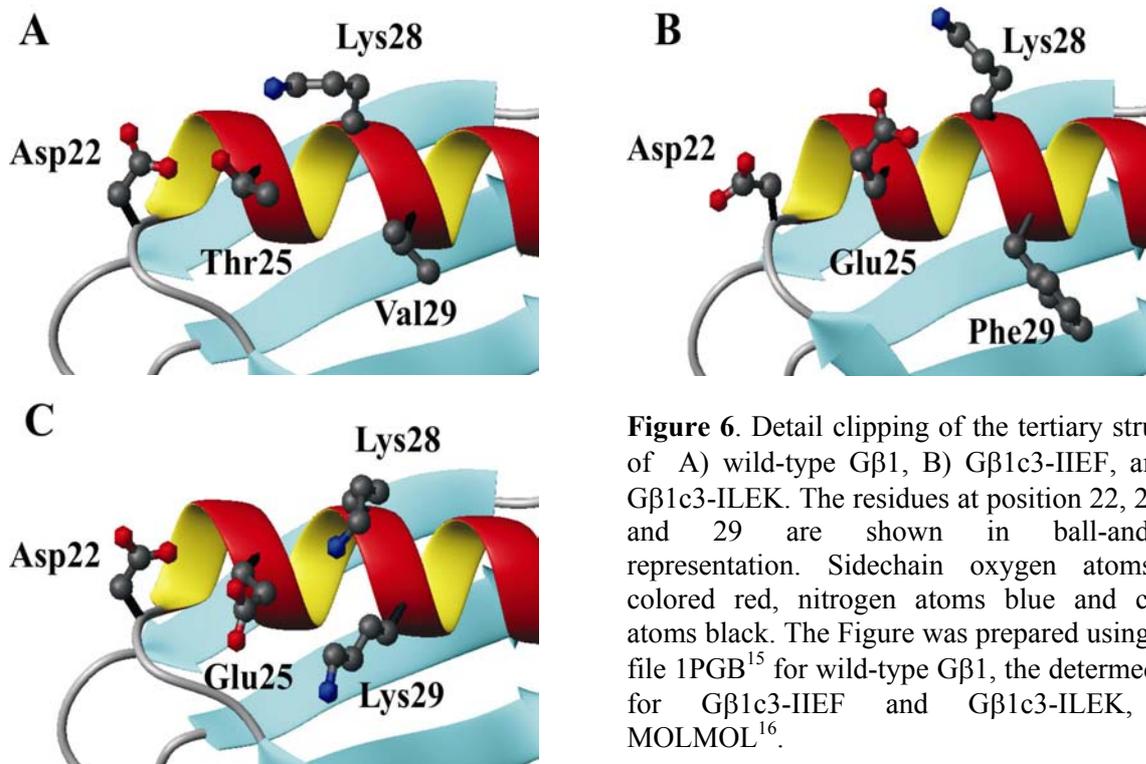


Figure 5. A) Surface illustration of wild-type G β 1. Detail clipping of the tertiary structure of B) wild-type G β 1, C) G β 1c3-IIEF, and D) G β 1c3-ILEK. The residues at position 16, 18, 29, 30 and 33 are shown in ball-and-stick representation and in figure 5B-D the van-der-Waals-radius of the atoms of these sidechains is shown. The Figure was prepared using PDB file 1PGB¹⁵ for wild-type G β 1, the determined data for G β 1c3-IIEF and G β 1c3-ILEK, and the programmes Swiss-PdbViewer²⁶ and MOLMOL¹⁶.

At position 18, the wild-type residue Thr buries 215 \AA^2 , whereas Leu18 in G β 1c3-ILEK and Ile18 in G β 1c3-IIEF bury 239 and 252 \AA^2 , respectively. Wild-type G β 1 and the ILEK and IIEF variants differ also strongly in the contact area between the residues at the positions 18 and 29. In the wild-type protein Thr18 and Val29 form a contact area of only 17 \AA^2 . In the ILEK variant (with Leu18 and Lys29) this value is increased to 40 \AA^2 , and in the IIEF variant (with the residues Ile18 and Phe29) to 61 \AA^2 . We suggest that this increase in buried hydrophobic area between these two residues explains, at least in part, the additional stabilization of the IIEF variant. Experimentally, by double mutant cycle analysis, we found only a very small interaction energy between Leu18 and Lys29 of 0.2 kJ mol^{-1} and we could not detect an interaction between Ile18 and Phe29 (Figure 4). This highlights the difficulty of translating distances and contact areas, as derived from crystal structures, into thermodynamic energies, and vice versa.

The residues 25 and 29 are located in the helix of G β 1 and solvent-exposed. They, as well as at the conserved residues Asp22 and Lys28, show different orientations in the G β 1 variants (Figure 6). Asp22 is the N-cap residue for the helix, and, in the wild-type protein, it is in an optimal position to interact with two unsatisfied NH in the first helix turn and to decrease the

dipole moment of the helix. The hydroxyl group of Thr25 is also within hydrogen bonding distance to the carboxyl group of Asp22 (Figure 6A).



The two mutants show a Glu residue at position 25. The mutation T25E stabilizes Gβ1 by 2.0 kJ mol⁻¹ (Table 2). The orientation of the Glu25 side chain depends on whether position 29 is occupied by a Phe or by a Lys (Figure 6B,C). Phe29 and Lys29 differ strongly in the orientation of their side chains. Lys29 in the ILEK variant is oriented towards Glu25 and the two residues are separated by only 4.3 Å, suggestive of ion pair formation. Phe29 in the IIEF variant is oriented away from Glu25 towards Tyr33 and Ile18 (cf. Figure 5B). As a consequence, the side chain of Glu25 rotates and orients its carboxyl group to form an ion pair with the amino group of Lys28 (Figure 6B). Interestingly, in this variant the carboxylate of Asp22 is switched from its position at the helix N-cap out into the solvent (Figure 6B).

Movements of the protein backbone

Mutational effects are not necessarily confined to changes in the side chains and their interactions. They can also affect the backbone conformation, and thus affect a multitude of interactions in the folded protein. Changes in stability mediated by backbone movements are extremely difficult to predict computationally. The tertiary structure of Gβ1 is determined by the packing of the single helix across the four stranded β sheet. As a measure for changes in the backbone conformation we used the distance between the C^α positions of Phe30 in the α helix and Thr/Ile18 in the β sheet. This distance increases from 8.0 and 8.4 Å in the two structures for the wild-type protein to 8.7 Å in the ILEK and 9.1 Å in the IIEF variant. This increase correlates with the size of the side chain at position 29, which is Val in the wild-type

protein, Lys in the ILEK, and Phe in the IIEF variant. Whether this widening of the molecule affects its stability is not clear.

Conclusions

It remains a major challenge to understand the correlation between the sequence, the structure and the thermodynamic stability of a protein, even for a small protein, such as G β 1. We determined high-resolution crystal structures measured individual contributions to stability by an exhaustive mutational analysis, but still could not elucidate unambiguously, why the two variants of G β 1, one from computational design and the other from experimental *in-vitro* selection, differed strongly in their stabilities.

The conformational stability of a protein is determined not only by contacts in the folded conformations, although they provide the major contribution to the energy, i. e. the enthalpy of unfolding. Changes in steric strain and overpacking can have dramatic influences on stability but are very difficult to evaluate. Changes in the dynamics of the native and the unfolded state and in solvation can occur as well, they change the folding entropy and are very difficult to analyze.

Regarding this complexity, it is encouraging that computational design has led to impressive successes both in the stabilization of existing proteins and in the design of novel stable proteins. However, many false positives are still picked in the computational design, and it is important to validate the designs experimentally. Moreover, positions identified as promising in the calculations should be used to construct libraries of variants and use them for experimental *in-vitro* evolution to select the optimal residues for these positions.

Material and Methods

Protein purification

The gene of *g β 1* was amplified by PCR with primers including restriction sites for *NdeI* at the N terminus and *BamHI* at the C terminus. The T2Q mutation was introduced to avoid the loss of the N-terminal methionine²⁴ and the gene was cloned into the plasmid pET11a (Novagen, Madison, Wisconsin, USA). The mutations were introduced by amplifying the whole plasmid by PCR with primers coding for the mutations, blunt end ligation of the PCR product and transformation into *E. coli* XL1Blue. The mutant form of the plasmid was isolated and transformed into *E. coli* BL21(DE3)pLysS (Stratagene, La Jolla, USA) for protein expression.

The proteins were purified using DNA precipitation by poly(ethyleneimine) and the supernatant was incubated at either 60 °C or 70 °C for 20 min dependent on the expected stability to precipitate heat-labile proteins. In the following anion exchange chromatography and gel filtration was used as described²⁴ with minor modifications.

The two variants G β 1c3-IIEF and G β 1c3-ILEK used for crystallisation are including the core mutations (Y3F, L7I, V39I) and were obtained as described.¹⁸

Thermal and chemical induced unfolding transitions

Thermal unfolding transitions of 4 μM protein in 0.1 M Na cacodylate/HCl (pH 7.0) and different concentrations of NaCl were followed by circular dichroism at 222.6 nm with a 1-nm bandwidth in 10-mm cells using a Jasco J-600 spectropolarimeter equipped with a PTC-348 WI Peltier device. The data were analyzed according to a two-state model using non-linear regression and the program Graft (Erithacus Software, Staines, UK), the heat capacity change ΔC_p was held constant at 4,000 J mol⁻¹ K⁻¹. The reversibility of thermal unfolding was examined by heating protein samples for 5 min at a temperature that was 10 °C above its T_M value and cooled to 20 °C. The subsequent thermal unfolding transitions were identical with those obtained without pre-heating.

Crystallization of G β 1 constructs

Lyophilized protein was dissolved and dialyzed over night in a buffer containing 50 mM Na acetate pH 5.0. Crystallization was performed using the hanging drop method. 0.4 μl protein stock solution was mixed with 0.4 μl crystallization buffer and incubated at 20 °C. The reservoir was filled with 100 μl crystallization buffer. Crystals of sufficient quality and size for data collection formed within 4 to 10 days. The crystallization buffer for G β 1c3-IIEF contained 33% PEG monomethylether, 0.1 M CaCl₂. Crystals of G β 1c3-IIEF were directly frozen in liquid nitrogen. The crystallization buffer for G β 1c3-ILEK contained 2.3 M ammonium sulphate, 0.1 M Na acetate pH 4.0. Crystals of G β 1c3-ILEK were transferred into crystallization buffer with an addition of 25% glycerol for cryoprotection and frozen in liquid nitrogen.

Data collection and processing

Data collection of a G β 1c3-IIEF crystal was performed using our inhouse X-ray generator equipped with a MAR345 image-plate. XDS package²⁸ was used to index and scale diffraction data. X-ray diffraction data of a G β 1c3-ILEK crystal were collected at a wavelength of 0.9184 Å at the Protein Structure Factory beamline BL 14.1 of the Free University of Berlin at BESSY (Berlin) on a MAR165 CCD camera. Diffraction data was processed and integrated using the HKL2000 software. The analysis of the processed data revealed a hemihedral twinning of the G β 1c3-ILEK crystal. The Detwin program²⁹ was used to determine the twin fraction and to detwin the data. A twin fraction of 26 percent was used for detwinning. The quality of the collected diffraction data set is summarized in Table 1.

Model building and refinement

The phase problem was solved by Molecular replacement using the AMoRe software³⁰ and the structural model of wild-type G β 1 (1PGB)¹⁵. The ARP/wARP software³¹ was used to generate initial atomic models of the G β 1 structures. The structure was refined by Refmac5³² version 5.1.24. 5% of the reflections were set aside for cross-validation, and R_{free} was used to adjust the refinement strategy and monitor the refinement progress. Automated building cycles were followed by manual inspection, repositioning of incorrectly oriented groups and building of missing groups in O³³. Search for water molecules was performed both manually and using the ARP/wARP³¹. In the late cycles of automated refinement, B-factors of the G β 1c3-IIEF

structure were refined using TLS refinement with the molecule being defined as one TLS group. The G β 1c3-ILEK dataset provided sufficient data to allow anisotropic B-factors refinement. The quality of the refined model is summarized in Table 1. The resulting structural model and the experimental data of the G β 1c3-IEEF and G β 1c3-ILEK structure were deposited in the Protein Data Bank.

Acknowledgements

We thank the members of our groups for suggestions and comments on the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie.

References

1. Wintrode, P. L. & Arnold, F. H. (2000). Temperature adaptation of enzymes: lessons from laboratory evolution. *Adv. Protein Chem.* **55**, 161-225.
2. Butterfoss, G. L. & Kuhlman, B. (2005). Computer-Based Design of Novel Protein Structures. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 49-65.
3. Korkegian, A., Black, M. E., Baker, D. & Stoddard, B. L. (2005). Computational thermostabilization of an enzyme. *Science* **308**, 857-860.
4. Bolon, D. N., Grant, R. A., Baker, T. A. & Sauer, R. T. (2005). Specificity versus stability in computational protein design. *Proc. Natl. Acad. Sci. USA* **102**, 12724-12729.
5. Kuhlman, B. & Baker, D. (2004). Exploring folding free energy landscapes using computational protein design. *Curr. Opin. Struct. Biol.* **14**, 89-95.
6. Xia, Y. & Levitt, M. (2004). Simulating protein evolution in sequence and structure space. *Curr. Opin. Struct. Biol.* **14**, 202-207.
7. Leemhuis, H., Stein, V., Griffiths, A. D. & Hollfelder, F. (2005). New genotype-phenotype linkages for directed evolution of functional proteins. *Curr. Opin. Struct. Biol.* **15**, 472-478.
8. Roodveldt, C., Aharoni, A. & Tawfik, D. S. (2005). Directed evolution of proteins for heterologous expression and stability. *Curr. Opin. Struct. Biol.* **15**, 50-56.
9. Eijsink, V. G., Gaseidnes, S., Borchert, T. V. & van den Burg, B. (2005). Directed evolution of enzyme stability. *Biomol. Eng.* **22**, 21-30.
10. Amstutz, P., Forrer, P., Zahnd, C. & Plückthun, A. (2001). In vitro display technologies: novel developments and applications. *Curr. Opin. Biotechnol.* **12**, 400-405.
11. Dahiyat, B. I. & Mayo, S. L. (1997). Probing the role of packing specificity in protein design. *Proc. Natl. Acad. Sci. USA* **94**, 10172-10177.
12. Malakauskas, S. M. & Mayo, S. L. (1998). Design, structure and stability of a hyperthermophilic protein variant. *Nature Struct. Biol.* **5**, 470-475.
13. Gronenborn, A. M., Filpula, D. R., Essig, N. Z., Achari, A., Whitlow, M., Wingfield, P. T. & Clore, G. M. (1991). A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. *Science* **253**, 657-661.
14. Gronenborn, A. M. & Clore, G. M. (1993). Identification of the contact surface of a streptococcal protein G domain complexed with a human Fc fragment. *J. Mol. Biol.* **233**, 331-335.

15. Gallagher, T., Alexander, P., Bryan, P. & Gilliland, G. L. (1994). Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR. *Biochemistry* **33**, 4721-4729.
16. Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51-55.
17. Sieber, V., Plückthun, A. & Schmid, F. X. (1998). Selecting proteins with improved stability by a phage-based method. *Nat. Biotechnol.* **16**, 955-960.
18. Wunderlich, M., Martin, A., Staab, C. A. & Schmid, F. X. (2005). Evolutionary protein stabilization in comparison with computational design. *J. Mol. Biol.* **351**, 1160-1168.
19. Horovitz, A. (1996). Double-mutant cycles: a powerful tool for analyzing protein structure and function. *Fold. Des.* **1**, R121-126.
20. Horovitz, A. & Fersht, A. R. (1992). Co-operative interactions during protein folding. *J. Mol. Biol.* **224**, 733-740.
21. Cho, J. H. & Raleigh, D. P. (2005). Mutational analysis demonstrates that specific electrostatic interactions can play a key role in the denatured state ensemble of proteins. *J. Mol. Biol.* **353**, 174-185.
22. Pace, C. N., Alston, R. W. & Shaw, K. L. (2000). Charge-charge interactions influence the denatured state ensemble and contribute to protein stability. *Protein Sci.* **9**, 1395-1398.
23. Smith, C. K. & Regan, L. (1997). Construction and design of beta-sheets. *Accounts Chem. Res.* **30**, 153-161.
24. Smith, C. K., Withka, J. M. & Regan, L. (1994). A thermodynamic scale for the beta-sheet forming tendencies of the amino acids. *Biochemistry* **33**, 5510-5517.
25. Serrano, L. & Fersht, A. R. (1989). Capping and alpha-helix stability. *Nature* **342**, 296-299.
26. Guex, N. & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714-2723.
27. Sobolev, V., Sorokine, A., Prilusky, J., Abola, E. E. & Edelman, M. (1999). Automated analysis of interatomic contacts in proteins. *Bioinformatics* **15**, 327-332.
28. Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795-800.
29. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760-763.
30. Navaza, J. (2001). Implementation of molecular replacement in AMoRe. *Acta Crystallogr D Biol Crystallogr* **57**, 1367-1372.
31. Lamzin, V. S., Perrakis, A. & Wilson, K. S. (2001). The ARP/wARP suite for automated construction and refinement of protein models. *International Tables for Crystallography Vol F*, 720 - 722.
32. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D-Biological Crystallography* **53**, 240-255.
33. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* **47 (Pt 2)**, 110-119.

7.5 Teilarbeit E

Michael Wunderlich and Franz X. Schmid.

In vitro Evolution of a Hyperstable G β 1 Variant.

Journal of Molecular Biology. **363**, 545-557 (2006)

7.6 Teilarbeit F

Energetic determinants for a hyperstable variant of G β 1 obtained from *in-vitro* evolution.

Michael Wunderlich and Franz X. Schmid.

Manuskript in Vorbereitung

Running title: Energetic analysis of a hyperstable G β 1 variant

Energetic determinants for a hyperstable variant of G β 1 obtained from *in-vitro* evolution

Michael Wunderlich and Franz X. Schmid*

*Laboratorium für Biochemie und Bayreuther Zentrum für Molekulare Biowissenschaften,
Universität Bayreuth, D-95440 Bayreuth, Germany*

**Corresponding author*

Franz X. Schmid

Tel: ++49 921 553660

Biochemie

Fax: ++49 921 553661

Universität Bayreuth

E-mail: fx.schmid@uni-bayreuth.de

D-95440 Bayreuth, Germany

Abbreviations used: G β 1, β 1 domain of the streptococcal protein G; G β 1-M2, most stable selected variant of the β 1 domain of the streptococcal protein G including the mutations E15V, T16L, T18I, and N37L; T_M , midpoint of a thermal unfolding transition; GdmCl, guanidinium chloride.

(Abstract)

In previous work, *in-vitro* evolution was used to create a hyperstabilized variant of the $\beta 1$ domain of protein G (G $\beta 1$). This protein, termed G $\beta 1$ -M2, contains the four mutations E15V, T16L, T18I, and N37L. Together, they increase the midpoint of thermal unfolding by 35.1 °C and the Gibbs free energy of stabilization by 28.5 kJ mol⁻¹. Here we analyzed the molecular basis of this strong stabilization. As single mutations, the E15V, T16L, T18I, and N37L substitutions stabilize wild-type G $\beta 1$ by contributions that range between 1.6 and 6.0 kJ mol⁻¹. Their sum (15.3 kJ·mol⁻¹) is much smaller than the stabilization observed when all four mutations are present simultaneously (28.5 kJ·mol⁻¹). Conversely, when the stabilizing mutations were removed individually from the stabilized variant, the sum of the individual losses in stability amounted to 48.5 kJ·mol⁻¹. A series of double mutant analyses with the mutations at the positions 16, 18, and 37 in all possible combinations revealed strong energetic interactions between Leu16, Ile18, and Leu37 in the stabilized variant. Leu37 is more than 10 Å apart from Leu16 and Ile18. The strong energetic coupling the three residues is mediated by the side chain of Tyr33, which is sandwiched between Leu37 and Leu16. In the stabilized variant Leu16, Ile18, Tyr33 and Leu37 form a closely packed cooperative unit near the surface of G $\beta 1$, which fills a cavity that is present in the wild-type protein. This improved packing leads to a strong increase in the unfolding enthalpy by about 70 kJ·mol⁻¹ and in the heat capacity of unfolding by 1.5 kJ·mol⁻¹ K⁻¹.

Keywords: protein stability, *in-vitro* selection, hydrophobic interactions, mutant cycle analysis streptococcal protein G

Introduction

To increase the stability of natural proteins is challenging, and many approaches reaching from computational chemistry to protein *in-vitro* evolution are currently used to tackle this problem¹⁻¹⁰. Protein stabilization projects are important to improve the performance of proteins as tools for biotechnology, and they provide us with novel insights into the principles that govern protein stability. This requires a thorough thermodynamic and structural analysis of the successfully stabilized protein variants

In the past, the $\beta 1$ domain of protein G (G $\beta 1$)¹¹⁻¹³ has been used extensively as a model for investigations of the relation between sequence and structure¹⁴⁻¹⁸ and, in particular, for computational and experimental protein stabilization projects.¹⁹⁻²¹ G $\beta 1$ is a small protein and its chain of 56 amino acids folds into a four-stranded antiparallel β -sheet and a single α -helix, which stretches across the β sheet (Figure 1).

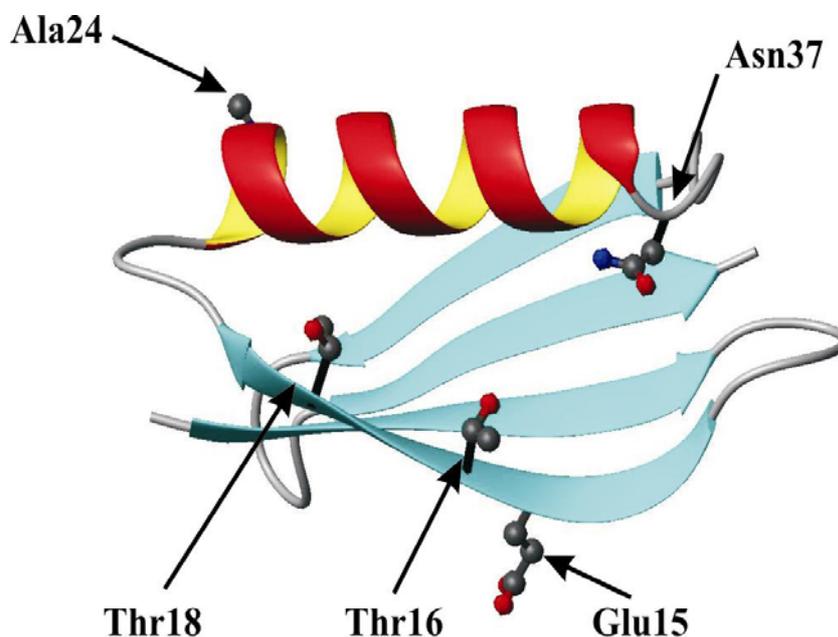


Figure 1. Tertiary structure of wild-type G $\beta 1$. Identified positions by error-prone PCR and in the following randomized by saturation mutagenesis are shown in ball-and-stick representation. Sidechain oxygen atoms are colored red, nitrogen atoms blue and carbon atoms black. The Figure was prepared using PDB file 1PGB¹³ and MOLMOL.²²

By computational design, Mayo and coworkers optimized the sequence of G $\beta 1$ at three core¹⁹ and four partially exposed (“boundary”) positions²⁰, and the resulting sevenfold mutant showed a strongly increased thermodynamic stability. Our group uses the *in vitro* selection system *Proside*²³ for protein stabilization. *Proside* links the increased protease resistance of stabilized protein variants to the infectivity of a filamentous phage. We employed the four boundary positions of G $\beta 1$ that were identified in the computational work, randomized the corresponding codons, and selected strongly stabilized G $\beta 1$ variants from the corresponding phage libraries²¹. Two of them were more stable than the best variant from the computational design project.

In subsequent *in-vitro* evolution work, an entirely experimental approach was used to stabilize G $\beta 1$. Sequence positions with a potential for stabilization were identified in libraries of G $\beta 1$ variants that were created by error-prone PCR, and then the most promising positions

were randomized simultaneously by saturation mutagenesis. In both cases the stabilized variants were selected by *Proside* from the corresponding libraries. These *in-vitro* selections finally led to a stabilized protein with the four mutations E15V, T16L, T18I, and N37L. For this variant, termed G β 1-M2, the midpoint of the thermal unfolding transition, T_M , increased by 35.1 °C, and the Gibbs free energy of denaturation, ΔG_D , increased by 28.5 kJ mol⁻¹ (at 70 °C) (Wunderlich et al., submitted). G β 1-M2 was thus much more stable than the seven-fold variants from the hybrid approach that combined computational and experimental approaches ($\Delta\Delta G_D = 22.7$ kJ mol⁻¹ at 70 °C).²¹

Here we analyzed the energetic basis of the strong stabilization of the G β 1-M2 variant by the four selected mutations. First, the mutations at the positions 15, 16, 18, and 37 were introduced individually into the wild-type protein, and then, in a reciprocal approach, they were removed individually from the stabilized variant. Finally, the three mutations at the positions 16, 18, and 37 were introduced into G β 1 in all possible combinations for a series of double mutant analyses^{24; 25}. Strong cooperative interactions between the mutated residues were uncovered in this way. They are communicated over a wide distance along the protein surface.

Results and Discussion

Individual contributions to stability of the mutations at the positions 15, 16, 18, and 37

Eight variants were constructed to measure the individual contributions of the stabilizing mutations. Four of them are based on wild-type G β 1 and contained the individual mutations E15V, T16L, T18I, or N37L. The other four variants are based on the stabilized variant G β 1-M2 and contained the respective back mutations V15E, L16T, I18T, or L37N. The stabilities of these variants were determined from thermal unfolding transitions measured in the presence of 1.5 M GdmCl by the change in circular dichroism (CD) at 222.6 nm (Figure 2). Adding a denaturant was necessary to shift the thermal unfolding transitions to the temperature range below 100 °C. The stability data for all variants are given in Table 1.

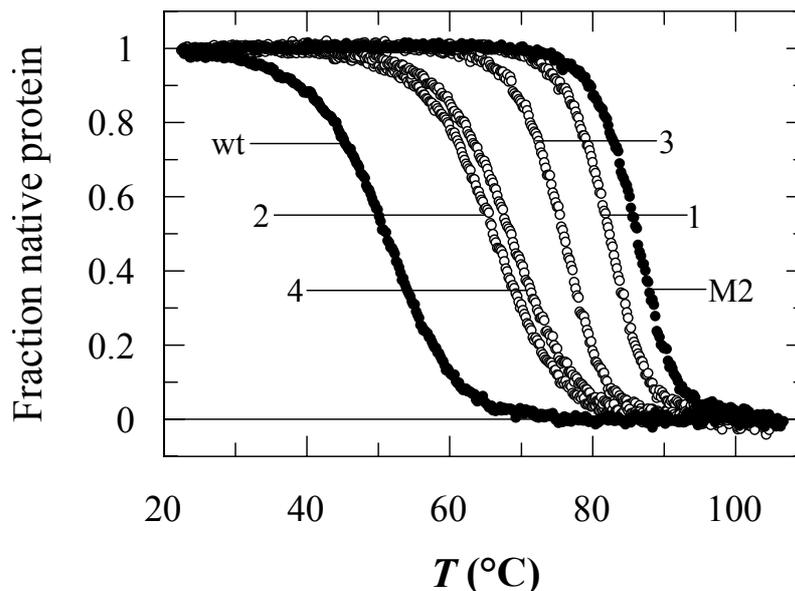


Figure 2. Thermal unfolding transitions of wild-type G β 1 (●), the variants 1-4 (○), and the variant M2 (●) (for the numbering, see Table 1). The transitions of 4 μ M protein were measured in 0.1 M sodium cacodylate-HCl (pH 7.0) by circular dichroism at 222.6 nm in the presence of 1.5 M GdmCl. The fractions of native protein as obtained after two-state analysis are shown as a function of temperature. The corresponding thermodynamic parameters are given in Table 1.

Table 1. Stability data for the characterised variants of G β 1

Variant	Position				T_M^a (°C)	ΔH_D^b (kJ mol $^{-1}$)	$\Delta G_D(70\text{ °C})^c$ (kJ mol $^{-1}$)	$\Delta\Delta G_D^d$ (kJ mol $^{-1}$)
	15	16	18	37				
wt	<i>Glu</i>	<i>Thr</i>	<i>Thr</i>	<i>Asn</i>	51.2	186	-12.9	0
G β 1-M2	Val	Leu	Ile	Leu	86.3	378	15.6	28.5
1	<i>Glu</i>	Leu	Ile	Leu	82.3	366	11.8	24.7
2	Val	<i>Thr</i>	Ile	Leu	66.3	206	-2.3	10.6
3	Val	Leu	<i>Thr</i>	Leu	75.8	330	5.3	18.2
4	Val	Leu	Ile	<i>Asn</i>	68.5	212	-0.9	12.0
5	Val	<i>Thr</i>	<i>Thr</i>	<i>Asn</i>	53.8	196	-11.3	1.6
6	<i>Glu</i>	Leu	<i>Thr</i>	<i>Asn</i>	56.8	184	-8.4	4.5
7	<i>Glu</i>	<i>Thr</i>	Ile	<i>Asn</i>	55.9	200	-9.7	3.2
8	<i>Glu</i>	<i>Thr</i>	<i>Thr</i>	Leu	59.8	203	-6.9	6.0
9	<i>Glu</i>	Leu	Ile	<i>Asn</i>	66.0	221	-2.6	10.3
10	<i>Glu</i>	Leu	<i>Thr</i>	Leu	71.9	315	1.7	14.6
11	<i>Glu</i>	<i>Thr</i>	Ile	Leu	63.9	196	-3.8	9.1

The thermodynamic parameters were derived from thermal unfolding transitions as shown in Figure 2. The variant M2 is identical to our previous work, the most stable variant derived by the *in vitro* evolution. The variants 1-4 represent the deletion of one mutation back to the wild-type residue in the variant G β 1-M2 and the variants 5-8 the single mutation variants based on the wild-type protein. The variants 9-11 are constructed to complete the double mutant cycles for the analysis of the coupling energies. The residues of each variant identical to the wild-type protein are written in italics. The accuracies of T_M and ΔH_D are ± 0.2 degrees and ± 5 kJ mol $^{-1}$, respectively. The accuracy of $\Delta G_D(70\text{ °C})$ is correlated with T_M . For variants with T_M close to 70 °C the accuracy is ± 0.3 kJ mol $^{-1}$, for variants with $65\text{ °C} \geq T_M \geq 75\text{ °C}$ the accuracy is ± 0.5 kJ mol $^{-1}$.

^a T_M is the midpoint of the thermal unfolding transitions.

^b ΔH_D is the enthalpy of unfolding at T_M .

^c ΔG_D is the change in Gibbs free energy of unfolding at 70 °C.

^d $\Delta\Delta G_D$ is the change in Gibbs free energy relative to the wild-type protein.

Position 15

The residues 15, 16, and 18 are in the edge strand (β_2) of the four-stranded β -sheet of G β 1 (Figure 1). The side chain of Glu15 in the wild-type protein is solvent-exposed, but in the selections for variants with improved stability Val was found most frequently. As a single mutation, E15V stabilizes the wild-type protein by 1.6 kJ mol⁻¹ (compare variant 5 and wt, Table 1). In the best variant, G β 1-M2, the reverse V15E mutation decreased the stability by 3.8 kJ mol⁻¹ (compare variants G β 1-M2 and 1 in Figure 2 and Table 1). The β -sheet-forming propensity for Val is much higher than for Glu^{18, 26}, which provides the simplest explanation for the stabilizing effect of the mutation E15V. In the wild-type protein Glu15 is close to Lys4, but the contribution of a potential Glu15-Lys4 salt bridge to stability is probably screened by the 1.5 M GdmCl, which had to be added to decrease the T_M value of G β 1-M2 to a value below 100 °C.

Position 18

The side-chain hydroxyl of Thr18 in the wild-type protein points to the surface of the protein. The high stabilizing potential at position 18 was revealed also by computational design²⁰ and in our previous *Proside in-vitro* selections²¹, (Wunderlich, unpublished results). In all these approaches, T18I was identified as the optimal substitution. As a single mutation, it stabilizes the wild-type protein moderately by 3.2 kJ mol⁻¹ (compare variant 7 and wt, Table 1). The effect of reverse mutation I18T in the best variant G β 1-M2 is, however, much larger: it decreases ΔG_D by 10.2 kJ mol⁻¹ (compare variants 3 and G β 1-M2 in Table 1).

Position 16

Position 16 is occupied by Thr in the wild-type protein and by Leu in the variant G β 1-M2. The T16L mutation stabilizes wild-type G β 1 by 4.5 kJ mol⁻¹ (compare variant 6 and wt in Table 1). The effect of the reverse mutation L16T in the stabilized variant G β 1-M2 is fourfold larger. It decreases in the T_M value by 20.0 °C (Figure 2) and reduces the stability by 17.9 kJ mol⁻¹ (compare G β 1-M2 and variant 1, Table 1). This enormous difference of 13.4 kJ mol⁻¹ (Figure 3) between the ΔG_D values for introducing Leu16 into wild-type G β 1 and removing it from the stabilized variant G β 1-M2 indicates that there are strong cooperative interactions between the mutated residues in the M2 variant.

In the computational design²⁰ and in the previous *in vitro* evolution²¹ of the four boundary positions 16, 18, 25, and 29 always Ile16 was found as the best replacement for Thr16. In the variant G β 1-M2 (with optimized residues at the positions 15, 16, 18, and 37) Leu16 was found in all the selected variants. Changing Leu16 to Ile in G β 1-M2 in fact destabilized this variant by 2.1 kJ mol⁻¹, which confirms that the mutational effects at position 16 are context-dependent.

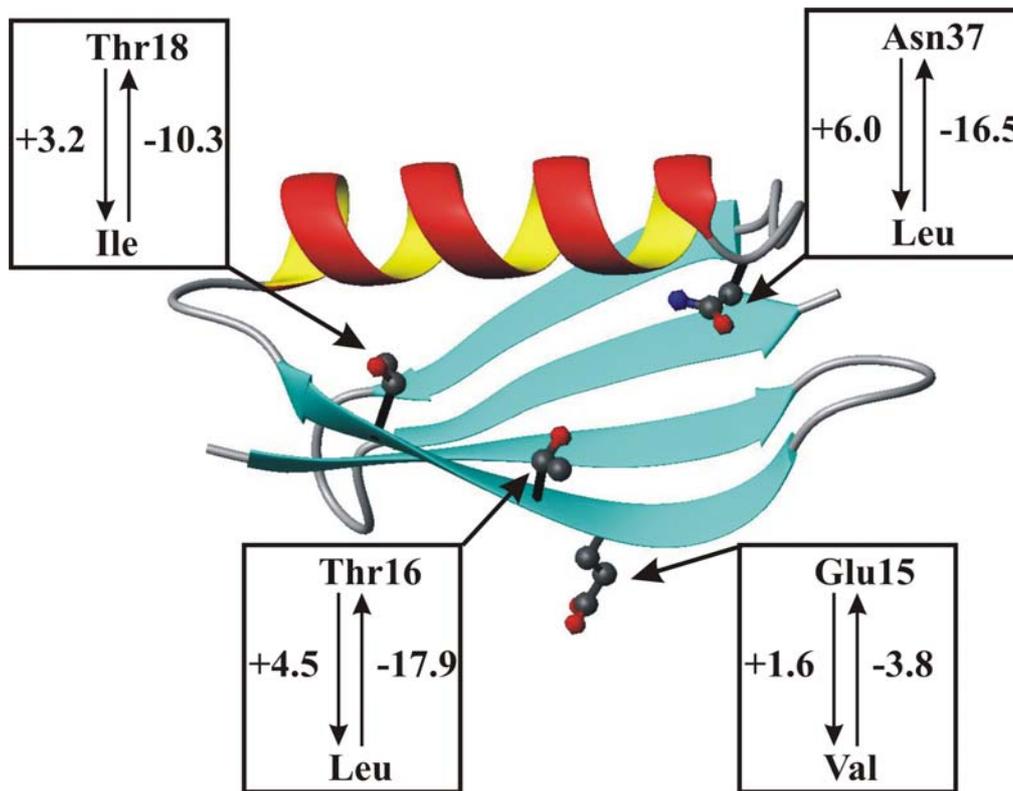


Figure 3. Tertiary structure of wild-type G β 1. The residues at position 15, 16, 18 and 37 are shown in ball-and-stick representation. Sidechain oxygen atoms are colored red, nitrogen atoms blue and carbon atoms black. The stabilizing effect of the introduction of each single mutation in the wild-type protein and the effect of the mutation to the wild-type residue at one position is given (kJ mol^{-1}). The Figure was prepared using PDB file 1PGB¹³ and MOLMOL.²²

Position 37

In wild-type G β 1, Asn37 is located in the loop that connects the α -helix with strand β 3 (Figure 1), and in the *Proside* selections Leu was consistently found at this position. The mutation N37L stabilizes wild-type G β 1 by 6.0 kJ mol^{-1} , which is the highest stabilizing effect observed thus far for a single mutation (Table 1). This strong stabilizing potential at position 37 was not discovered in the computational design of Mayo and coworkers²⁰ The effect of the reverse mutation L37N in the best variant (G β 1-M2) is even stronger. It decreases the stability by 16.5 kJ mol^{-1} (compare G β 1-M2 and variant 4, Table 1). This indicates that Leu37 establishes strongly stabilizing interactions with the other mutated residues in the M2 variant, although they are remote from Leu37 (Figure 1).

Analysis of the interaction energies between the residues at the positions 16, 18, and 37

The variant with the three stabilizing mutations at the positions 16, 18, and 37 (variant 1, Table 1) is 24.7 kJ mol^{-1} more stable than wild-type G β 1, but the single mutations account for only 13.7 kJ mol^{-1} of this additional stability. The non-additivity of the contributions to stability becomes most evident when the increases in the stability of the wild-type protein caused by the individual mutations are compared with the losses in stability caused by the corresponding back mutation in the best stabilized variant (Figure 3). At the positions 16 and 37, in particular, these values differ by more than 10 kJ mol^{-1} . This points to strong

cooperative interactions between the mutated residues, especially between the residues at the positions 16, 18, and 37.

Interaction energies between two residues can be obtained from a double mutant cycle, those between three residues from a triple mutant cube^{24; 25; 27}. Such a cube consists of six double mutant cycles, which represent the faces of the cube. The corners of the cube are formed by the eight species which contain the wild-type and the mutated residues (for G β 1 at the positions 16, 18 and 37) in all possible combinations. Table 1 presents the stability data for these eight variants, and Figure 4 the six faces of the corresponding mutant cube.

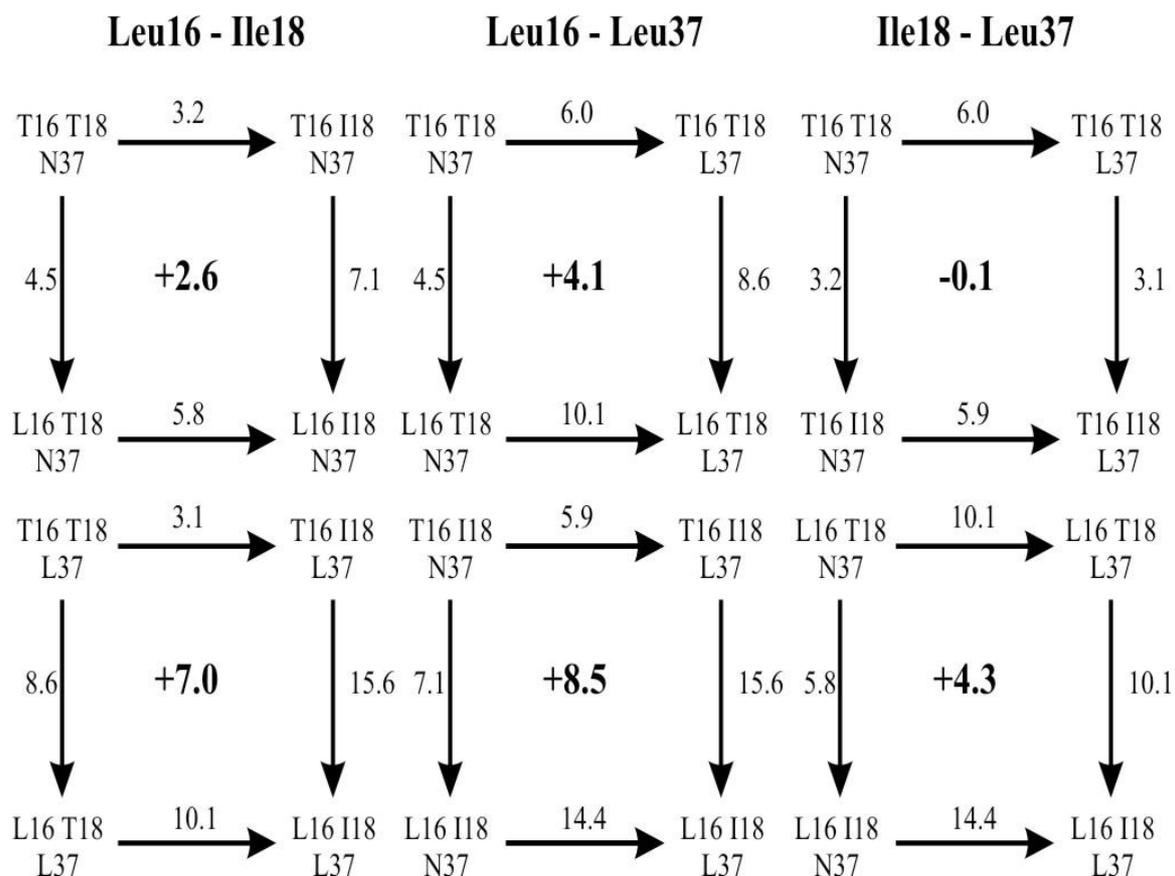


Figure 4. Double mutant cycles with all possible combinations of amino acid residues as they occur in wild-type G β 1 and the most stable variant G β 1-M2. For every single mutation, the $\Delta\Delta G_D$ values at 70 °C (kJ mol^{-1}) are shown. In the centre of every cycle, the coupling energy between the respective two amino acid residues is given (kJ mol^{-1}). The experimental conditions are given in Table 1. The left pair of cycles determines the interaction energies between Leu16 and Ile18, in the absence (upper cycle) and in the presence (lower cycle) of Leu37. The middle pair gives the interaction between Leu16 and Leu37, in the absence and in the presence of Ile18, and the right pair show the interaction energies between Ile18 and Leu37, in the absence and in the presence of Leu16.

The double mutant cycles on opposing faces of the cube reflect the interaction between the same two residues, in the presence or in the absence of the third mutation. The left pair in Figure 4 thus represents the interaction between Leu16 and Ile18 in the presence of Asn37 as in the wild-type protein (top) or in the presence of Leu37, as in the stabilized variant G β 1-M2 (bottom).

Leu16 and Ile18 show a favorable interaction free energy of $2.6 \text{ kJ}\cdot\text{mol}^{-1}$ when position 37 is occupied by Asn (as in the wild-type protein, upper left cycle, Figure 4). In the wild-type protein Thr16 and Thr18 do not contact each other (Figure 1). Leu16 and Ile18 have larger side chains, and possibly they form a direct contact in the M2 variant, which might give rise to the interaction energy of $2.6 \text{ kJ}\cdot\text{mol}^{-1}$. When the third mutation, N37L is introduced, the interaction energy between Leu16 and Ile18 increases almost threefold to $7.0 \text{ kJ}\cdot\text{mol}^{-1}$ (lower left cycle, Figure 4). This is a remarkable finding, because Leu16 is 10 \AA and Ile18 is 14 \AA away from residue 37 (Figure 1).

The middle pair of mutant cycles in Figure 4 represents the interaction between Leu16 and Leu37 in the presence of Thr18 (top) or Ile18 (bottom), respectively. The interaction energy between the Leu16 and Leu37 is $4.1 \text{ kJ}\cdot\text{mol}^{-1}$ in the presence of the wild-type residue Thr18 and rises to $8.5 \text{ kJ}\cdot\text{mol}^{-1}$ in the presence of Ile18, as in the M2 variant. This shows that, despite their large spatial separation, Leu16 and Leu37 establish a strongly stabilizing energetic interaction. This interaction is further enhanced when position 18 is occupied by Ile.

The right pair of cycles (Figure 4) shows that Ile18 and Leu37 do not interact energetically when position 16 is occupied by Thr (as in the wild-type protein). In the presence of Leu16, however, an energetic link is established between Ile18 and Leu37, leading to an additional stabilization by $4.4 \text{ kJ}\cdot\text{mol}^{-1}$.

The triple mutant analysis reveals an intriguing energetic communication between the three mutation sites. (i) Ile18 interacts energetically with its neighbor Leu16, but not with Leu37, which is about 14 \AA away. (ii) Leu16 and Leu37, which are also remote from each other at the surface of $G\beta 1$, engage in a long-distance stabilizing energetic interaction of $4.1 \text{ kJ}\cdot\text{mol}^{-1}$. This interaction is strongly increased to $8.5 \text{ kJ}\cdot\text{mol}^{-1}$ when the third stabilizing residue (Ile18) is also present. (iii) The T16L mutation creates an energetic link between Ile18 and Leu37 and leads to an interaction energy of $4.3 \text{ kJ}\cdot\text{mol}^{-1}$. Correspondingly, the presence of Leu37 at this remote position increases the interaction energy between the neighboring residues Leu16 and Ile18 from 2.6 to $7.0 \text{ kJ}\cdot\text{mol}^{-1}$.

Structure of the stabilized variant

In wild-type $G\beta 1$ the residues at the positions 16 and 37 (Thr16 and Asn37) are separated by the side chain of Tyr33 (Figure 5). Asn37, but not Thr16, is in van der Waals contact with the aromatic ring of Tyr33. The side chains of Thr16 and Thr18 are on the same side of strand $\beta 1$, but they do not contact each other. Thr18 also does not contact Tyr33.

In the M2 variant small polar residues (2×Thr and Asn) are replaced by larger, hydrophobic residues (Ile and 2×Leu) (Figure 5). As a consequence, the hydrophobic and van der Waals interactions between Leu37 and Tyr33 are improved, and this explains the stabilizing effect of the N37L single mutation. Unlike Thr16 and Thr18 in the wild-type protein, Leu16 and Ile18 in the M2 variant contact each other, which explains the interaction energy between these two residues (Figure 4). Wild-type $G\beta 1$ contains a cleft that separates the side chains of Thr16 and Thr18 on one side from the aromatic ring of Tyr33 and the sidechain of Val29 on the other side. At the bottom of this cleft part of the side chain of Phe30 is exposed (Figure 6). In the stabilized variant, this cleft probably is no longer present. It is filled by the larger side chains of Leu16 and Ile18, which now contact each other as well as

the side chain of Tyr33. Thus, a continuous hydrophobic region is formed that encompasses the side chains of all four residues, Leu16, Ile18, Tyr33 and Leu37 (Figure 6).

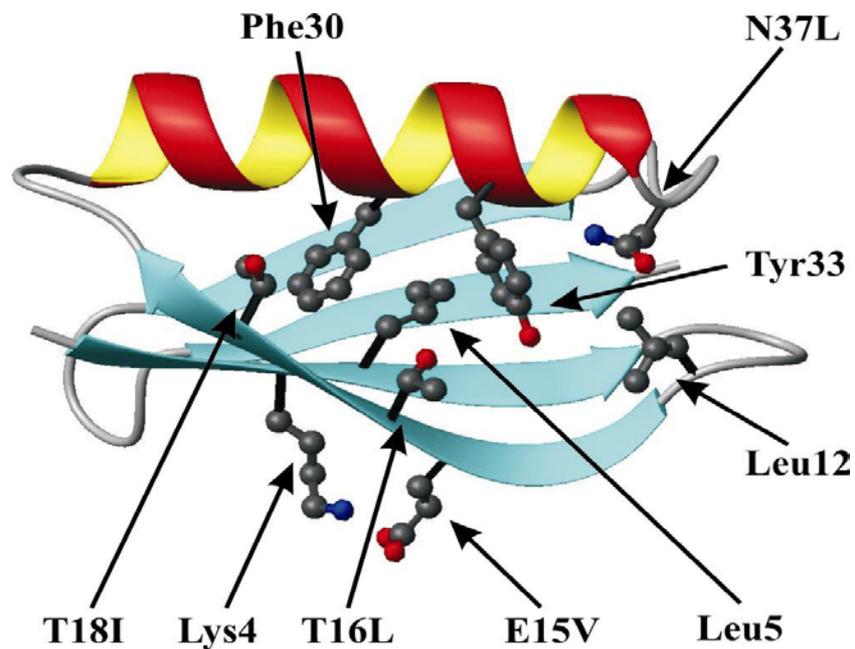


Figure 5. Tertiary structure of wild-type G β 1. The wild-type residues of the stabilizing mutations E15V, T16L, T18I and N37L and the surrounding positions Lys4, Leu5, Leu12, Phe30 and Tyr33 are shown in ball-and-stick representation. Sidechain oxygen atoms are colored red, nitrogen atoms blue and carbon atoms black. The Figure was prepared using PDB file 1PGB¹³ and MOLMOL.²²

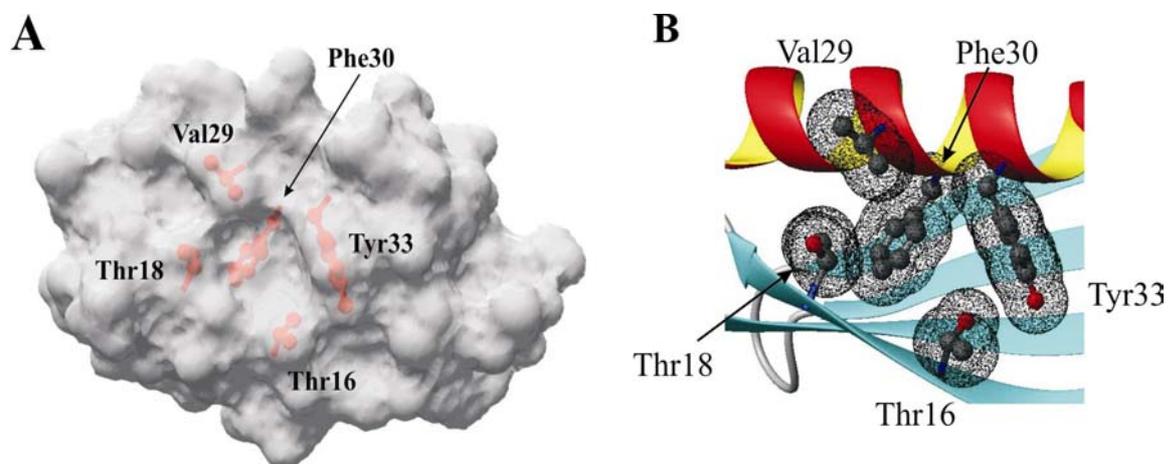


Figure 6. A) Surface illustration of wild-type G β 1. B) Detail clipping of the tertiary structure of wild-type G β 1. The residues at position 16, 18, 29, 30 and 33 are shown in ball-and-stick representation and in figure 5B the van-der-Waals-radius of the atoms of these sidechains is shown. The Figure was prepared using PDB file 1PGB¹³ and the programmes Swiss-PdbViewer²⁸ and MOLMOL²².

These contacts and the filling of the cleft provide the structural explanation for the increased stability of the M2 variant and, in particular, for the observed strong energetic cooperation between Leu16, Ile18 and Leu37. Clearly, the mutual interactions of these residues with Tyr33 mediate a long-range energetic cooperativity along the surface of the G β 1

molecule. This indirect cooperative effect contributes $11 \text{ kJ}\cdot\text{mol}^{-1}$ to the stability of the M2 variant.

Analysis of the enthalpy of unfolding

The variants of G β 1 that contain both Leu16 and Leu37 show a strongly increased van't Hoff enthalpy of unfolding, ΔH_D . This becomes apparent when the variants 2, 3, and 4 in Table 1 are compared. Variant 2 contains only Leu37 and variant 4 contains only Leu16. They show similar T_m values and a similar ΔH_D of about $210 \text{ kJ}\cdot\text{mol}^{-1}$. Variant 3 contains both Leu16 and Leu37, and its ΔH_D is $110 \text{ kJ}\cdot\text{mol}^{-1}$ higher than the ΔH_D values of variants 2 and 4, although the difference between the T_M values is only $10 \text{ }^\circ\text{C}$. The ΔC_p value for the unfolding of G β 1 is $3\text{-}4 \text{ kJ}\cdot\text{mol}^{-1}$, which accounts for maximally $40 \text{ kJ}\cdot\text{mol}^{-1}$ of the difference in ΔH_D . The simultaneous presence of Leu16 and Leu37 thus leads to an increase in ΔH_D by about $70 \text{ kJ}\cdot\text{mol}^{-1}$.

The heat capacity of unfolding, ΔC_p , is increased as well. The ΔC_p values for the wild-type protein and for the stabilized variant G β 1-M2 were determined from plots of the van't Hoff-enthalpies of unfolding ΔH_D as a function of the T_M values, as derived from thermal unfolding transitions in the presence of varying concentrations of GdmCl (Figure 7). These ΔC_p values are higher than the real ΔC_p because GdmCl solvates the unfolded protein^{29; 30}, but this apparent increase in ΔC_p is the same for the two variants. The comparison in Figure 6 thus shows that the stabilizing mutations increase ΔC_p of unfolding by $1.5 \text{ kJ}\cdot(\text{mol}\cdot\text{K})^{-1}$, which is equivalent to an about 40 % increase relative to the ΔC_p of wild-type G β 1.

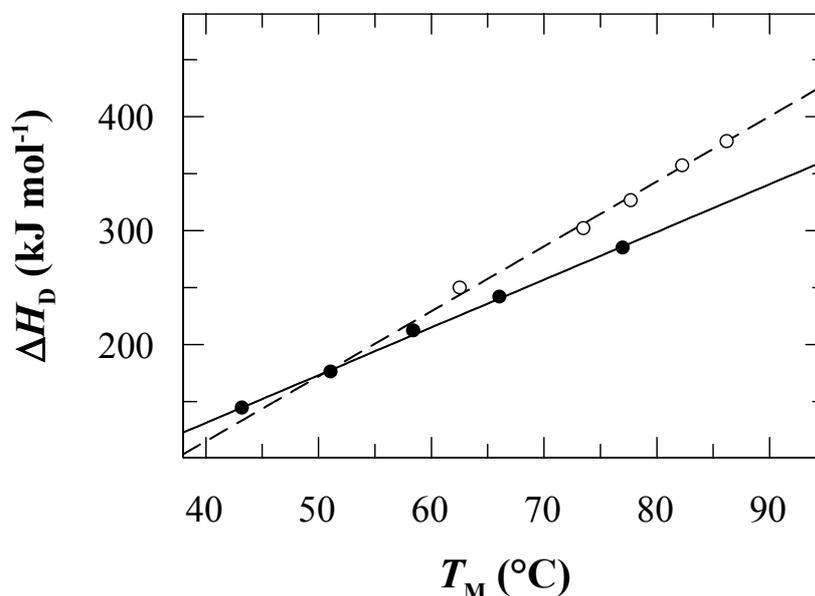


Figure 7. Analysis of Δc_p for wild-type G β 1 (●) and the variant G β 1-M2 (○). For the determination of Δc_p thermal transitions of $4 \mu\text{M}$ protein were measured in 0.1 M sodium cacodylate-HCl (pH 7.0) by circular dichroism at 222.6 nm in the presence of different concentrations of GdmCl. The transitions were analysed according two-state behaviour and the derived enthalpies of unfolding are plotted against the melting points.

The increase in ΔH_D by $70 \text{ kJ}\cdot\text{mol}^{-1}$ suggests that the stabilization of the G β 1-M2 variant results from additional enthalpic interactions. Forming these strong interactions is

accompanied by a large loss in entropy, equivalent to $T\Delta S = -40 \text{ kJ}\cdot\text{mol}^{-1}$, resulting in a ΔG_D of $30 \text{ kJ}\cdot\text{mol}^{-1}$. This strong enthalpy/entropy compensation and the increase in ΔC_p are hallmarks of hydrophobic interactions. These changes in the thermodynamics of unfolding thus support our conclusion that the stabilizing mutations in the G β 1-M2 variant led to an improved packing of the side chains of Leu16, Ile18, Tyr33, and Leu37 into a cooperative unit near the surface of the protein.

Conclusions

The results clearly show the two possibilities for a protein to escape proteolytic cleavage in the selection steps. On the one side of course the increase in the T_M value (in our case $\Delta T_M = 35.1 \text{ }^\circ\text{C}$) is the major factor, on the other side the higher cooperativity by about 50 kJ mol^{-1} leads to a up-shift of the onset of proteolytic accessibility (fraction unfolded protein $\geq 1\%$) of about $3 \text{ }^\circ\text{C}$. Both effects are contributing to the strong stabilization of G β 1-M2 by 28.5 kJ mol^{-1} .

The stabilization of G β 1-M2 is based on about one half on stabilizing interactions of the mutated residues. The residues Leu16, Ile18 and Leu37 are forming a hydrophobic cluster mediated by Tyr33. This result show the importance of simultaneous randomization to optimize not only residue at one position but to find to best combination. In our previous *in vitro* evolution³¹ of the cold shock protein *Bs-CspB* from *Bacillus subtilis* we identified four strongly stabilizing single mutations by selection of libraries created by error-prone PCR. The sum of the stabilizing effects of all four mutations is 34.3 kJ mol^{-1} , in contrary the four-fold mutant is only stabilized by 14.1 kJ mol^{-1} . By manuell combination of the mutations we could identify a three-fold mutant stabilized by 20.9 kJ mol^{-1} . The reason for the non-additivity of the effect of the four mutations are strong unfavorable electrostatic interactions. Although in this case by a simultaneous randomization the optimal combination could be identified not to find additional stabilizing interactions but to avoid unfavorable interactions.

Material and Methods

Protein purification

The gene of *g β 1* was amplified by PCR with primers including restriction sites for *NdeI* at the N terminus and *BamHI* at the C terminus. The T2Q mutation was introduced to avoid the loss of the N-terminal methionine¹⁸ and the gene was cloned into the plasmid pET11a (Novagen, Madison, Wisconsin, USA). The mutations were introduced by amplifying the whole plasmid by PCR with primers coding for the mutations, blunt end ligation of the PCR product and transformation into *Escherichia coli* XL1Blue. The mutant form of the plasmid was isolated and transformed into *E. coli* BL21(DE3)pLysS (Stratagene, La Jolla, USA) for protein expression.

The proteins were purified using DNA precipitation by poly(ethyleneimine) and the supernatant was incubated at either $60 \text{ }^\circ\text{C}$ or $70 \text{ }^\circ\text{C}$ for 20 min depend on the expected stability to precipitate heat-labile proteins. In the following anion exchange chromatography and gel filtration was used as described¹⁸ with minor modifications.

Thermal and chemical induced unfolding transitions

Thermal unfolding transitions of 4 μM protein, in 0.1 M Na cacodylate/HCl (pH 7.0), 1.5 M GdmCl were followed by circular dichroism at 222.6 nm with a 1-nm bandwidth in 10-mm cells using a Jasco J-600 spectropolarimeter equipped with a PTC-348 WI Peltier device. To analyse the change of ΔC_p , for some variants thermal unfolding transitions in the presence of different GdmCl concentrations were measured. The data were analyzed according to a two-state model using non-linear regression and the program Grafit (Erithacus Software, Staines, UK), the heat capacity change ΔC_p was held constant at 4,000 J mol⁻¹ K⁻¹. The reversibility of thermal unfolding was examined by heating protein samples for 5 min at a temperature that was 10 °C above its T_M value and cooled to 20 °C. The subsequent thermal unfolding transitions were identical with those obtained without pre-heating.

Acknowledgements

We thank the members of our group for suggestions and comments on the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie.

References

1. Butterfoss, G. L. & Kuhlman, B. (2005). Computer-Based Design of Novel Protein Structures. *Annu. Rev. Biophys. Biomol. Struct.* 35, 49-65.
2. Korkegian, A., Black, M. E., Baker, D. & Stoddard, B. L. (2005). Computational thermostabilization of an enzyme. *Science* 308, 857-860.
3. Bolon, D. N., Grant, R. A., Baker, T. A. & Sauer, R. T. (2005). Specificity versus stability in computational protein design. *Proc. Natl. Acad. Sci. USA* 102, 12724-12729.
4. Kuhlman, B. & Baker, D. (2004). Exploring folding free energy landscapes using computational protein design. *Curr. Opin. Struct. Biol.* 14, 89-95.
5. Xia, Y. & Levitt, M. (2004). Simulating protein evolution in sequence and structure space. *Curr. Opin. Struct. Biol.* 14, 202-207.
6. Leemhuis, H., Stein, V., Griffiths, A. D. & Hollfelder, F. (2005). New genotype-phenotype linkages for directed evolution of functional proteins. *Curr. Opin. Struct. Biol.* 15, 472-478.
7. Roodveldt, C., Aharoni, A. & Tawfik, D. S. (2005). Directed evolution of proteins for heterologous expression and stability. *Curr. Opin. Struct. Biol.* 15, 50-56.
8. Eijsink, V. G., Gaseidnes, S., Borchert, T. V. & van den Burg, B. (2005). Directed evolution of enzyme stability. *Biomol. Eng.* 22, 21-30.
9. Amstutz, P., Forrer, P., Zahnd, C. & Plückthun, A. (2001). In vitro display technologies: novel developments and applications. *Curr. Opin. Biotechnol.* 12, 400-405.
10. Winthrope, P. L. & Arnold, F. H. (2000). Temperature adaptation of enzymes: lessons from laboratory evolution. *Adv. Protein Chem.* 55, 161-225.
11. Gronenborn, A. M., Filpula, D. R., Essig, N. Z., Achari, A., Whitlow, M., Wingfield, P. T. & Clore, G. M. (1991). A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. *Science* 253, 657-661.

12. Gronenborn, A. M. & Clore, G. M. (1993). Identification of the contact surface of a streptococcal protein G domain complexed with a human Fc fragment. *J. Mol. Biol.* 233, 331-335.
13. Gallagher, T., Alexander, P., Bryan, P. & Gilliland, G. L. (1994). Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR. *Biochemistry* 33, 4721-4729.
14. Merkel, J. S., Sturtevant, J. M. & Regan, L. (1999). Sidechain interactions in parallel beta sheets: the energetics of cross-strand pairings. *Structure* 7, 1333-1343.
15. Minor, D. L., Jr. & Kim, P. S. (1994). Context is a major determinant of beta-sheet propensity. *Nature* 371, 264-267.
16. Minor, D. L., Jr. & Kim, P. S. (1994). Measurement of the beta-sheet-forming propensities of amino acids. *Nature* 367, 660-663.
17. Dalal, S., Balasubramanian, S. & Regan, L. (1997). Protein alchemy: changing beta-sheet into alpha-helix. *Nature Struct. Biol.* 4, 548-552.
18. Smith, C. K., Withka, J. M. & Regan, L. (1994). A thermodynamic scale for the beta-sheet forming tendencies of the amino acids. *Biochemistry* 33, 5510-5517.
19. Dahiyat, B. I. & Mayo, S. L. (1997). Probing the role of packing specificity in protein design. *Proc. Natl. Acad. Sci. USA* 94, 10172-10177.
20. Malakauskas, S. M. & Mayo, S. L. (1998). Design, structure and stability of a hyperthermophilic protein variant. *Nature Struct. Biol.* 5, 470-475.
21. Wunderlich, M., Martin, A., Staab, C. A. & Schmid, F. X. (2005). Evolutionary protein stabilization in comparison with computational design. *J. Mol. Biol.* 351, 1160-1168.
22. Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* 14, 51-55.
23. Sieber, V., Plückthun, A. & Schmid, F. X. (1998). Selecting proteins with improved stability by a phage-based method. *Nat. Biotechnol.* 16, 955-960.
24. Horovitz, A. (1996). Double-mutant cycles: a powerful tool for analyzing protein structure and function. *Fold. Des.* 1, R121-126.
25. Horovitz, A. & Fersht, A. R. (1992). Co-operative interactions during protein folding. *J. Mol. Biol.* 224, 733-740.
26. Smith, C. K. & Regan, L. (1997). Construction and design of beta-sheets. *Accounts Chem. Res.* 30, 153-161.
27. Perl, D. & Schmid, F. X. (2001). Electrostatic stabilization of a thermophilic cold shock protein. *J. Mol. Biol.* 313, 343-357.
28. Guex, N. & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18, 2714-2723.
29. Pfeil, W. & Privalov, P. L. (1976). Thermodynamic investigations of proteins. II. Calorimetric study of lysozyme denaturation by guanidine hydrochloride. *Biophys Chem* 4, 33-40.
30. Makhatadze, G. I. & Privalov, P. L. (1992). Protein interactions with urea and guanidinium chloride. A calorimetric study. *J. Mol. Biol.* 226, 491-505.
31. Wunderlich, M., Martin, A. & Schmid, F. X. (2005). Stabilization of the cold shock protein CspB from *Bacillus subtilis* by evolutionary optimization of Coulombic interactions. *J. Mol. Biol.* 347, 1063-1076.

Danksagung

Die vorliegende Arbeit von April 2002 bis April 2006 am Lehrstuhl Biochemie der Universität Bayreuth unter der Leitung von Herrn Professor Dr. Franz X. Schmid angefertigt. Bei ihm möchte ich mich sehr herzlich für sein großes Engagement und für die hervorragende wissenschaftliche Betreuung bedanken. Durch seine zahlreichen Ratschläge und Ideen hatte er maßgeblichen Anteil an dem Entstehen dieser Arbeit.

Meinen Laborkolleginnen und -kollegen Jochen Balbach, Barbara Eckert, Roman Jakob, Insa Kather, Christian Löw, Stefan Lorenz, Christine Magg, Andreas Martin, Claudia Staab und Markus Zeeb gilt mein Dank für die angenehme Laboratmosphäre und die Hilfe bei den kleinen und großen Problemen des Laboralltags.

Andreas Martin sei gedankt dafür, dass er mich in die Geheimnisse des *phage display* eingeführt hat, und Claudia Staab für den Einsatz im Rahmen Ihrer Diplomarbeit.

Für die sehr angenehme und erfolgreiche Zusammenarbeit zur Bestimmung der Strukturen von verschiedenen Proteinvarianten möchte ich Klaas Max und Prof. Dr. Udo Heinemann danken.

Meiner Schwester Martina und Bettina Theissen möchte ich für das Korrekturlesen dieser Arbeit danken.

Und zum Abschluss möchte ich allen danken, die wissentlich oder unwissentlich zum erfolgreichen Entstehen dieser Arbeit beigetragen haben.

Erklärung

Hiermit erkläre ich, dass ich die Arbeit selbständig verfasst und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet habe.

Ferner erkläre ich, dass ich nicht anderweitig mit oder ohne Erfolg versucht habe, diese Dissertation einzureichen. Ich habe keine gleichartige Doktorprüfung an einer anderen Hochschule endgültig nicht bestanden.

Bayreuth, den 28. Juni 2006