In vitro-Evolution und Analyse der biophysikalischen Grundlagen der Proteinstabilität

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

vorgelegt von
Diplom-Biochemiker
Andreas Martin

Bayreuth 2003

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 September 1998 bis Februar 2003 am Lehrstuhl für Biochemie der Universität Bayreuth unter der Anleitung von Prof. Dr. Franz X. Schmid angefertigt.

Promotionsgesuch eingereicht am:	21. Februar 2003
Tag des wissenschaftlichen Kolloquiums:	4. Juli 2003

Prüfungsausschuss:

Prof. Dr. Franz Xaver Schmid	(Erster Gutachter)
Prof. Dr. Paul Rösch	(Zweiter Gutachter)
Prof. Dr. Heinz Hoffmann	
Prof. Dr. Wolfgang Schumann	
Prof. Dr. Carlo Unverzagt	(Vorsitzender)

Inhaltsverzeichnis

Kurzfassung der Arbeit

Proteine sind die zentralen Bausteine des Lebens, und ihre pharmazeutische und biotechnologische Anwendung gewinnt zunehmend an Bedeutung. Für ihre Funktion ist der Erhalt der dreidimensionalen Struktur essentiell. Somit stellt das Verständnis der biophysikalischen Grundlagen der konformationellen Stabilität von Proteinen ein wichtiges Ziel der biochemischen Forschung dar. Vergleichende Mutationsstudien an homologen Proteinen aus psychrophilen, mesophilen und thermophilen Organismen stoßen aufgrund der oft großen Zahl von Sequenzunterschieden und der damit verbundenen Komplexität der Analysen bald an ihre Grenzen. Die gerichtete Evolution *in vitro* stellt daher eine sehr vielversprechende Alternative für die Untersuchung der Thermostabilität und die Ableitung genereller Stabilitätsprinzipien dar.

In der vorliegenden Arbeit wurde Proside, ein Selektionssystem für stabilisierte Proteine, weiterentwickelt und dafür genutzt, die molekularen Ursachen erhöhter Proteinstabilität zu analysieren. Für diese *in vitro*-Evolutionsexperimente diente das Kälteschockprotein *Bs*-CspB aus *Bacillus subtilis* als Modellprotein. Zusätzlich wurde das Gen-3-Protein (G3P) des Phagen fd, eine essentielle Komponente von Proside, thermodynamisch charakterisiert und sein Faltungsmechanismus aufgeklärt.

Proside basiert auf *phage display* und verknüpft die thermodynamische Stabilität eines Proteins mit einem sehr gut selektierbaren Parameter, der Infektiosität filamentöser Phagen. Die Methode ist generell anwendbar und unabhängig von spezifischen Eigenschaften des zu stabilisierenden Proteins, wie etwa Ligandenbindung oder enzymatischer Aktivität. Die Größe der selektierbaren Mutantenbibliotheken konnte durch Veränderungen des Phagenkonstrukts und der zu ihrer Erstellung verwendeten molekularbiologischen Methoden auf mehr als 10^8 Varianten gesteigert werden. Gleichzeitig wurde mit diesen Modifikationen der rekombinationsbedingte Verlust von Gastproteinsequenzen aus dem Phagengenom deutlich reduziert. Das Gen-3-Protein, in welches diese Gastproteine inseriert werden, wurde mittels Proside um mehr als 10 kJ/mol stabilisiert. Somit sind jetzt *in vitro*-Selektionen bei Temperaturen von bis zu 60 °C möglich, ohne dass die Stabilität der Phagenproteine selbst limitierend für die Optimierung der Gastproteine wirkt.

Das Potential der so optimierten Proside-Methode konnte anhand der Selektionen stark stabilisierter Varianten von *Bs*-Csp gezeigt werden. Dieses Protein unterscheidet sich von seinem um 15,8 kJ/mol stabileren Homologen *Bc*-Csp aus dem thermophilen *Bacillus caldolyticus* in 12 von 67 Resten, die alle an der Oberfläche des Proteins liegen. Die *in vitro*-Selektionen nach Sättigungsmutagenese an sechs dieser zwölf Positionen in *Bs*-CspB (2, 3, 46, 64, 66 und 67) lieferten eine Vielzahl von stabilisierten Mutanten, die, abhängig von den Selektionsbedingungen, unterschiedliche Stabilisierungsprinzipien aufzeigten. Während bei der Selektion in Gegenwart des ionischen Denaturierungsmittels GdmCl vor allem die hydrophoben Wechselwirkungen verbessert wurden, führte die Selektion bei erhöhter Temperatur zusätzlich zur Optimierung der Coulomb´schen Wechselwirkungen auf der Proteinoberfläche.

Für die stabilste der selektierten Varianten von *Bs*-CspB ist der Mittelpunkt des thermischen Entfaltungsübergangs um mehr als 28 Grad relativ zum Wildtypprotein angehoben und liegt damit auch deutlich höher als der Wert des thermostabilen Referenzproteins *Bc*-Csp. Die Variante unterscheidet sich von *Bc*-Csp an allen sechs randomisierten Positionen. Dennoch nutzen die beiden Proteine ähnliche Strategien für eine

hohe Thermostabilität, insbesondere zeichnen sie sich durch eine im Vergleich zu *Bs*-CspB optimierte Verteilung der Oberflächenladungen aus.

Die detaillierte Analyse der Beiträge der einzelnen Aminosäureaustausche in der stabilsten selektierten Variante verdeutlichte, dass Ladungsnetzwerke auf der Proteinoberfläche für die Thermostabilität der Kälteschockproteine sehr wichtig sind. So ist die paarweise Abstoßung zwischen Glu3 und Glu66 in *Bs*-CspB gering im Vergleich zu den globalen, nicht-lokalisierbaren Coulomb´schen Beiträgen der beiden Reste, und auch der stabilisierende Effekt der Einführung einer positiven Ladung in diese Region der Proteinoberfläche erwies sich als unabhängig von ihrer genauen Position und war keiner direkten Wechselwirkung zuzuordnen.

Basierend auf diesen Erkenntnissen war es möglich, ein hyperstabiles Kälteschockprotein mit lediglich vier Austauschen relativ zu *Bs*-CspB zu konstruieren. Der Schmelzpunkt dieser Variante liegt bei 85,6 °C, d.h. 31,6 Grad über dem des Wildtypproteins. Sie ist damit sogar stabiler als das homologe Csp aus dem hyperthermophilen Organismus *Thermotoga maritima*. Durch rationales Design anhand von Sequenzvergleichen wäre eine entsprechende Stabilisierung kaum zu erzielen gewesen.

Die Stabilitätsuntersuchungen des für die Phageninfektion essentiellen N-terminalen Fragments des Gen-3-Proteins reflektieren dessen Aufbau aus den beiden Domänen N1 und N2. Während die Stabilität der Domäne N1 unabhängig von Domäne N2 ist, wird Domäne N2 wesentlich durch die Interdomänenwechselwirkungen mit N1 stabilisiert, und ihre Entfaltung ist mit der Domänendissoziation gekoppelt.

Die vier mittels Proside gefundenen Mutationen in G3P stabilisieren beide Domänen unabhängig voneinander und verdeutlichen die generellen Prinzipien, die der Stabilität von Zweidomänenproteinen zugrunde liegen. Neben der Erhöhung der intrinsischen Stabilitäten der individuellen Domänen spielt die Verbesserung der Domäneninteraktionen eine entscheidende Rolle für die Stabilisierung des gesamten Proteins.

Die Rückfaltung von G3P ist ein sequentieller Prozess. Domäne N1 faltet innerhalb weniger Millisekunden, gefolgt von der Faltung der N2-Domäne, die nach etwa 3 min abgeschlossen ist. Im letzten Schritt assoziieren die beiden Domänen in einer extrem langsamen Reaktion. Sie zeigt eine Zeitkonstante von 6200 s (bei 25 °C) und ist in ihrer Rate limitiert durch die *trans*-*cis*-Isomerisierung an Pro213 in der Gelenksubdomäne von N2. Wie kinetische NMR-Experimente mit einem von G3P abgeleiteten Pentapeptid ergaben, ist die ungewöhnlich niedrige Rate dieser Prolylisomerisierung auf die lokale Sequenzumgebung des *cis*-Pro213 zurückzuführen. Durch die Assoziation von N1 und N2 werden beide Domänen in ihrer Entfaltung gekoppelt, so dass Domäne N1 bis zu 150.000-fach langsamer entfaltet als in isolierter Form.

Die Prolin-limitierte sehr langsame Domänenassoziation ist möglicherweise für die Funktion des G3P bei der Infektion von *E. coli* von Bedeutung. Sie erlaubt es, die Domänen nach Bindung an den F-Pilus so lange dissoziiert zu halten, bis der Pilus zurückgezogen ist und Domäne N1 mit dem Corezeptor TolA an der Zelloberfläche wechselwirken kann.

Abstract

Proteins are the vital molecules in living cells, and they strongly gained in importance for pharmaceutical and biotechnological applications. For their function an intact three-dimensional structure is essential. Thus, it is a central aim in biochemical research to elucidate the biophysical determinants of conformational protein stability. To this end homologous proteins from psychrophile, mesophile and thermophile organisms have been compared in numerous mutational studies, but the usually large number of sequence differences impaired the analysis of increased protein stability at a molecular level. Evolutionary methods provide a promising alternative to understand the basis of protein structure and stability.

In this work Proside, a selection method for stabilizing proteins, was further developed and applied to analyze the molecular origins of protein stability. As a model protein for these *in vitro*-evolution experiments the cold shock protein *Bs*-CspB from *Bacillus subtilis* was used. In addition, the gene-3-protein (G3P) of filamentous phage fd, an essential component of Proside, was thermodynamically characterized and its folding mechanism elucidated.

Proside is based on phage display, and it links the thermodynamic stability of a protein with a well selectable property, the phage infectivity. This method is generally applicable and independent of specific characteristics of the protein to be stabilized, such as ligand binding or enzymatic activity. The size of the phage libraries for selections could be increased to about 10^8 variants by modifications of the phage construct and of the methods that are used to create these libraries. At the same time, the loss of guest protein sequences due to recombination was strongly reduced. By means of Proside, the stability of the phage gene-3-protein itself, in which the guest proteins are inserted, was increased by more than 10 kJ/mol. With this improved phage *in vitro*-selections are now possible at up to 60 °C, and the stability of the phage proteins is not limiting for the evolution of the inserted guest protein.

The efficiency of the Proside method could be demonstrated by the selections of strongly stabilized variants of *Bs*-CspB. *Bs*-CspB is 15.8 kJ/mol less stable than its homolog *Bc*-Csp from the thermophile *Bacillus caldolyticus*. The two proteins differ in 12 out of 67 residues, which are all located at the protein surface. Saturation mutagenesis at six of these 12 positions (2, 3, 46, 64, 66, and 67) and *in vitro*-selections with Proside yielded many stabilized mutants, that revealed different principles for stabilization depending on the selection conditions. In the presence of the ionic denaturant GdmCl variants with improved hydrophobic interactions were selected, whereas the selection at elevated temperature led to an optimization of coulombic interactions at the protein surface. For the most stable selected variant of *Bs*-CspB the midpoint of thermal unfolding transition increased by more than 28 degrees relative to the wild type protein and is thus much higher than the midpoint found for the thermostable reference *Bc*-Csp. The selected variant differs from *Bc*-Csp at all six randomized positions, but the two proteins use similar strategies to reach a high thermostability. In particular, they show an optimized surface charge distribution compared to wild-type *Bs*-CspB.

The detailed analysis of the contributions of single amino acid substitutions in the most stable selected variant suggested that ion networks at the protein surface are of particular importance for the thermostability of the cold shock proteins. Thus, the pairwise repulsion

between the neighboring Glu3 and Glu66 in *Bs*-CspB is small compared to the non-localized destabilizing effects of these two residues, and a positive charge in this region of the protein surface stabilizes *Bs*-CspB irrespective of its exact position.

Based on the selection results a hyperstable cold shock protein with only four mutations relative to *Bs*-CspB was constructed. Its melting point of 85.6 °C is 31.6 degrees higher than the value for the wild-type protein, and even higher than the melting point of the homologous Csp from the hyperthermophile *Thermotoga maritima*. By means of sequence comparisons and rational design such a strong stabilization would have been hardly possible.

The N-terminal fragment of the gene-3-protein of phage fd contains two domains, N1 and N2, and this composition of two structural entities is reflected in the observed thermodynamic parameters. The stability of domain N1 is independent of the presence of domain N2, whereas domain N2 strongly gains in stability by the interdomain contacts with N1 and its unfolding is coupled with domain dissociation.

The four mutations that were selected by Proside within G3P stabilize the two domains in an independent fashion and they point to general principles that govern the stability of two-domain proteins. In addition to the stabilization of individual domains, the improvement of domain interactions contributes considerably to the increased stability of the entire protein.

The kinetics of refolding of G3P turned out to be a sequential process. First domain N1 folds within a few milliseconds, followed by the folding of domain N2, which is completed after about 3 min. In the last step of refolding the two prefolded domains assemble in an extremely slow reaction. This domain docking shows a time constant of 6200 s (at 25 °C) and it is limited in rate by the *trans-cis*-isomerization of Pro213 in the hinge subdomain of N2. The unusually slow prolyl isomerization is caused by the peculiar local sequence context around Pro213, as determined by kinetic NMR-experiments with a pentapeptide derived from G3P. The docking of N1 and N2 links the two domains kinetically, and the unfolding of domain N1 is decelerated up to 150,000-fold by the interaction with N2.

Possibly, this proline-limited slow domain docking is relevant for the function of G3P in the infection process. The kinetic block in the domain re-assembly reaction would ensure that, once N1 and N2 became separated upon binding to the F pilus, domain closure is suppressed long enough so that N1 can interact with its receptor TolA after the pilus has been retracted.

1 Einleitung

1.1 Stabilität von Proteinen

Der Begriff „Proteinstabilität“ hat in der molekularen Biologie und in der Biochemie vielfältige Bedeutung. Im intakten Organismus ist damit oft die Lebensdauer eines Proteins gemeint, während *in vitro* die Stabilität gegenüber Proteasen, chemischen Agenzien oder verschiedenen physikalischen Einflüssen im Mittelpunkt steht. Meist korrelieren jedoch all diese Aspekte mit der thermodynamischen Definition der Stabilität eines Proteins.

Für viele kleine Proteine lässt sich die Faltung mit einem einfachen Zweizustandsmodell beschreiben, bei dem der native Zustand N und der entfaltete Zustand U in einem reversiblen Gleichgewicht N⇌U stehen (Privalov, 1979). Strukturell betrachtet stellen beide Zustände Ensembles von Konformationen dar, thermodynamisch jedoch sind sie wohl definiert, und die Proteinstabilität kann durch die freie Enthalpie der Entfaltung, ΔG_D, ausgedrückt werden (Gleichung 1).

$$\Delta G_D = -RT \cdot \ln \frac{[U]}{[N]} \qquad (1)$$

Das Gleichgewicht zwischen N und U lässt sich durch Erhöhung der Temperatur oder Zugabe von Denaturierungsmitteln zum entfalteten Zustand hin verschieben, die konformationelle Stabilität ΔG_D ergibt sich aus der Lage des Gleichgewichts in Abhängigkeit vom Denaturierungsparameter. Für die chemische Entfaltung wird die Abhängigkeit der Stabilität ΔG_D von der Denaturierungsmittelkonzentration oft durch ein lineares Modell sehr gut beschrieben (Tanford, 1970, Gleichung 2), was die Bestimmung von ΔG_D in Abwesenheit des Denaturierungsmittels ($\Delta G_D^{H_2O}$) ermöglicht.

$$\Delta G_D = \Delta G_D^{H_2O} - m \cdot [D] \qquad (2)$$

Bei der thermischen Entfaltung ist die freie Enthalpie ΔG_D mit der Enthalpie ΔH_D und der Entropie ΔS_D der Entfaltung über die Gibbs-Helmholtz-Gleichung (Gleichung 3) verknüpft.

$$\Delta G_D = \Delta H_D - T \cdot \Delta S_D \qquad (3)$$

Durch die Temperaturabhängigkeiten von ΔH_D und ΔS_D nimmt die Wärmekapazität ΔC_p eines Proteins während der Entfaltung zu (Gleichung 4), die thermische Stabilitätskurve (ΔG_D vs T) zeigt einen parabolischen Verlauf.

$$\Delta C_p = \frac{d\ \Delta H_D}{d\ T} = T \cdot \frac{d\ \Delta S_D}{d\ T} \qquad (4)$$

Bei hoher Temperatur sind Enthalpie und Entropie der Entfaltung positiv, das Protein ist somit enthalpisch stabilisiert aber entropisch destabilisiert. Mit abnehmender Temperatur nehmen beide, die Entropie und die Enthalpie der Entfaltung stetig ab und werden bei niedrigen Temperaturen negativ. Dies bedeutet, dass das Protein dann entropisch stabilisiert, enthalpisch jedoch destabilisiert ist (Privalov, 1979).

1.2 Physikalische Grundlagen der Proteinstabilität

Die Stabilität von Proteinen resultiert aus einer Vielzahl schwacher nicht-kovalenter Wechselwirkungen, deren relative Beiträge immer noch kontrovers diskutiert werden.

Den Hauptbeitrag liefert zweifellos der hydrophobe Effekt, das Verbergen unpolarer Seitenketten im Inneren des gefalteten Proteins (Kauzmann, 1959, Dill *et al.*, 1995). Hydrophobe Gruppen können keine Wasserstoffbrückenbindungen mit dem Lösungsmittel ausbilden. Dies verstärkt die Wechselwirkungen zwischen den Lösungsmittelmolekülen selbst und führt so zu einer entropisch sehr ungünstigen Solvatation. Außerdem bilden hydrophobe Gruppen im Inneren des Proteins enthalpisch günstige van-der-Waals-Kontakte aus (Makhatadze & Privalov, 1995).

Einen weiteren wichtigen Beitrag zur Stabilität des nativen Zustands liefern die Wasserstoffbrückenbindungen zwischen verborgenen polaren Gruppen (Makhatadze & Privalov, 1995, Pace *et al.*, 1996). Sie bestimmen mit ihrer hohen Abstands- und Richtungsspezifität die Ausbildung der nativen Struktur ganz entscheidend (Honig & Yang, 1995).

Die langreichweitigen ionischen Wechselwirkungen haben aufgrund der hohen Dielektrizitätskonstante von Wasser ($\varepsilon_r = 80$) auf der Proteinoberfläche nur einen relativ kleinen stabilisierenden Effekt (Honig & Yang, 1995). Im apolaren Proteininneren ($\varepsilon_r = 4$) sind diese Interaktionen wesentlich stärker (Gilson & Honig, 1986), jedoch wird dort ihr Nettobeitrag zur Stabilität durch die bei der Faltung erforderliche Desolvatation der interagierenden geladenen Gruppen extrem reduziert (Hendsch *et al.*, 1996, Xiao & Honig, 1999). Möglicherweise ist der Stabilisierungseffekt von Ionenpaaren bei hoher Temperatur und damit für thermostabile Proteine größer, da die Solvatationsenthalpie polarer Gruppen mit steigender Temperatur sinkt und der Entropiegewinn durch das Freisetzen von Wassermolekülen zunimmt (Elcock, 1998, Makhatadze & Privalov, 1995). Insgesamt kompensieren sich die vielen günstigen und ungünstigen Beiträge zur Proteinstabilität weitgehend, wobei die enorme Abnahme der Kettenentropie während der Faltung den vermutlich größten destabilisierenden Faktor darstellt (Makhatadze & Privalov, 1995).

1.3 Proteine aus thermophilen Organismen

Für Proteine aus mesophilen Organismen liegt die freie Enthalpie der Stabilisierung oft nur zwischen 20 und 80 kJ/mol (Dill, 1990). Dies entspricht der Energie von lediglich zwei bis drei Wasserstoffbrückenbindungen. Ein möglicher Grund für die relativ geringen Stabilitäten ist wohl, dass die meisten im Verlauf der Evolution entstandenen zufälligen Mutationen destabilisierend wirken. Durch die Akkumulation von Mutationen über mehrere Generationen nimmt die Proteinstabilität ab, solange kein entgegenwirkender Selektionsdruck vorhanden ist. Die sukzessive Destabilisierung bis hin zur Entfaltung des nativen Zustands wird *in vivo* jedoch durch das Erfordernis der Funktionalität verhindert, so dass natürlich evolvierte Proteine mit ihrer Stabilität immer nur knapp über einem funktionsbedingten Grenzwert liegen (Wintrode & Arnold, 2000). Eine weitere Erklärung ist, dass sich zu hohe Stabilitäten nachteilig auf die lebende Zelle auswirken können (Benner & Ellington, 1988), da zur Regulation des Zellzyklus und zur Reaktion auf veränderte Umweltbedingungen die intrazellulären Proteinkonzentrationen kurzfristig adaptiert werden müssen.

Die starke Kompensation von stabilisierenden und destabilisierenden Wechselwirkungen sorgt dafür, dass bereits geringe Änderungen in Art oder Zahl der Interaktionen die Proteinstabilität stark beeinflussen können. Dies erschwert die Ableitung allgemeiner Stabilitätsprinzipien anhand von Vergleichen homologer Proteine aus psychrophilen, mesophilen und thermophilen Organismen und das darauf basierende rationale Design ganz erheblich (Wintrode & Arnold, 2000). Meist unterscheiden sich diese Proteine an vielen Positionen in der Sequenz, und es ist schwierig, die Beiträge einzelner Austausche in Mutationsstudien aufzuklären. Zudem macht die irreversible Entfaltung der Proteine eine thermodynamische Charakterisierung oft unmöglich. Untersuchungen zur Thermostabilität beschränken sich daher häufig auf Sequenz- und Strukturvergleiche (Zhang *et al.*, 2001, Dams *et al.*, 2000). Wechselwirkungen lassen sich damit allerdings nicht quantifizieren, und das Risiko von Fehlinterpretationen ist sehr hoch (Strop & Mayo, 2000). Theoretische Ansätze zur Ermittlung der Beiträge einzelner Aminosäuren stehen derzeit noch am Anfang. Die Berechnungen von Mutationseffekten scheiterten bisher in vielen Fällen, weil entropische Effekte nicht berücksichtigt wurden, weil Informationen zur Struktur und Energie des denaturierten Zustands fehlten, und weil die geringe Nettostabilität von Proteinen aus einer Summe großer, sich nahezu kompensierender Beiträge resultiert (Gao *et al.*, 1989). Erleichtert werden die Berechnungen jedoch durch statistische Daten, die sich aus der mittlerweile sehr großen Zahl hochaufgelöster Proteinstrukturen gewinnen lassen. Mit Computeralgorithmen, die diese Daten nutzen, konnten Mutationseffekte auf die Stabilität, Struktur und Funktion von Proteinen bereits erfolgreich vorhergesagt werden (Malakauskas & Mayo, 1998, Street & Mayo, 1999, Looger & Hellinga, 2001, Benson *et al.*, 2000).

Wie aus Sequenz- und Strukturvergleichen mesostabiler und thermostabiler Proteine hervorgeht, existiert keine universelle Strategie zur Stabilisierung (Jaenicke & Böhm, 1998). Thermostabile Proteine zeichnen sich unter anderem durch verstärkte interne Packung unpolarer Gruppen, durch verkürzte Schleifen an der Oberfläche, durch eine größere Zahl von Wasserstoffbrückenbindungen und durch größere Netzwerke von Ionenpaaren bzw. mehr geladene Reste an der Proteinoberfläche aus (Yip *et al.*, 1995, Hatanaka *et al.*, 1997, Russell *et al.*, 1997, Cambillau & Claverie, 2000). Die Analyse natürlich vorkommender Enzyme führte zu der Annahme, dass sich während der Evolution Stärke und Zahl stabilisierender Wechselwirkungen derart entwickelt haben, dass konformationelle Rigidität (für Stabilität) und Flexibilität (für Aktivität) bei der physiologisch relevanten Temperatur optimal aufeinander abgestimmt sind. Meist bleibt aber unklar, welche Sequenzunterschiede tatsächlich für die thermische Adaptation eines Proteins entscheidend sind und bei welchen es sich lediglich um Resultate der neutralen genetischen Drift (Kimura, 1968, Kimura, 1983) bzw. anderer natürlicher Selektionsdrücke im Verlauf der Evolution handelt. Um diese Zusammenhänge zu klären bietet sich das evolutive Proteindesign an, denn: „Nothing in biology can be understood except in the light of evolution“ (Dobzhansky, 1973).

1.4 Gerichtete Evolution *in vitro*

Mit Hilfe der gerichteten Evolution *in vitro* ist es möglich, die physikalischen bzw. thermodynamischen Prinzipien der Proteinstabilität getrennt von anderen biologischen Eigenschaften zu analysieren. Verschiedenste natürliche oder auch nicht-natürliche Evolutionsdrücke lassen sich im Labor definiert nachstellen, so dass die gezielte

Entwicklung bestimmter Eigenschaften von Proteinen durch minimale Änderungen in Sequenz oder Struktur möglich ist. Gerichtete *in vitro*-Evolution ist allerdings nur erfolgreich, wenn wesentliche Grundprinzipien beachtet werden. Insbesondere bestimmt die gewählte Mutationsrate und die Zahl selektierbarer Mutanten die Wahrscheinlichkeit, optimierte Varianten eines Proteins zu finden.

Die Basis der evolutiven Triebkraft zur Verbesserung bestimmter Proteineigenschaften kann analog zum Proteinfaltungsproblem betrachtet werden. Der korrekt gefaltete native Zustand eines Proteins wird trotz der astronomisch hohen Zahl theoretisch möglicher Konformationen sehr schnell gefunden (Levinthal, 1968). Das ist damit erklärbar, dass die Determinanten der Faltung im Konformationsraum eine „Energielandschaft“ formen, in der das Protein bis zum Erreichen des Energieminimums auf bestimmten Faltungspfaden geführt wird (Frauenfelder *et al.*, 1991, Dill *et al.*, 1995). In Analogie zu dem Konformationsraum der Proteinfaltung lässt sich für die Evolution ein Sequenzraum als verbundenes Netzwerk aller Proteinsequenzen definieren, in dem sich das Protein durch Mutationen bewegt (Maynard Smith, 1970). Auch dieser Raum ist riesig, die Zahl der Sequenzpunkte beträgt 20^N, wobei N für die Zahl der Aminosäuren des Proteins steht. Mit den zur Verfügung stehenden Durchmusterungs- und Selektionsmethoden der *in vitro*-Evolution lassen sich zwar nur bis zu 10^{13} Mutanten pro Generation analysieren (Dower & Mattheakis, 2002), jedoch besitzt der Sequenzraum den Vorteil extrem hoher Verknüpfung. Trotz der geringen Größe einer einzelnen Generation kann er in mehreren Zyklen relativ effizient durchsucht werden kann.

Da die Berechnung selbst einfacher Eigenschaften eines Proteins aus dessen Primärsequenz bisher unmöglich ist, werden für die Betrachtung theoretischer Aspekte der Evolution die Punkte im Sequenzraum über sogenannte „Fitnessfunktionen“ mit bestimmten Proteineigenschaften (z.B. der Stabilität) verknüpft. Komplexere Funktionen beziehen in diese Verknüpfung auch die Struktur des Proteins mit ein (Shakhnovich, 1994, Dill *et al.*, 1995). Das heißt, äquivalent zur „Energielandschaft“ im Konformationsraum wird im Sequenzraum eine „Fitnesslandschaft“ definiert, deren statistische Eigenschaften den Verlauf der Evolution bestimmen (Amitrano *et al.*, 1989). Die einfachste Form einer solchen Hyperfläche, beschrieben durch eine ungekoppelte „Fitnessfunktion“, besitzt ein einzelnes, globales Optimum. Jeder Rest im Protein kann dabei unabhängig mutiert werden, die Effekte sind additiv, und die beste Sequenz ist einfach die Kombination der besten Aminosäure an jeder Position. In der Regel kommt es jedoch zu Kopplungen zwischen Positionen, die im Fall der Proteinstabilität durch paarweise Wechselwirkungen wie Wasserstoffbrückenbindungen, Salzbrücken oder van-der-Waals-Kontakte zwischen den Resten bedingt sind (Wells, 1990). Diese Kopplungen bestimmen die Toleranz der Struktur bzw. Stabilität eines Proteins gegenüber Mutationen (Brown & Sauer, 1999) und die Additivität von Mutationsbeiträgen, d.h. die Wahrscheinlichkeit, bereits mit Einzelaustauschen günstige Effekte zu erzielen.

Die am häufigsten verwendete Strategie zur evolutiven Stabilisierung von Proteinen ist die adaptive Evolution („*adaptive walk*“), bei der über mehrere Generationen hinweg einzelne günstige Mutationen zufällig erzeugt, angehäuft und durch einen entsprechenden Selektionsdruck fixiert werden (Wintrode & Arnold, 2000). Die Zufallsmutagenese erlaubt die Optimierung eines Proteins ohne detaillierte Kenntnis der komplexen Wechselwirkungen, die der Stabilität oder Funktion zu Grunde liegen. Da die meisten Mutationen ungünstige Beiträge liefern und miteinander koppelnde Positionen nur schwer zu finden

sind, erhöht die Einführung von nur einzelnen Mutationen bei der Zufallsmutagenese die Wahrscheinlichkeit der Verbesserung von Proteineigenschaften. Gleichzeitig erleichtern Einzelaustausche die Interpretation von Effekten auf molekularer Ebene. Allerdings ist durch die Mutation einzelner Nukleotide, meist durch fehlerbehaftete PCR, lediglich eine beschränkte Zahl von Aminosäuren und damit auch nur ein kleiner Teil des Sequenzraums zugänglich. Insbesondere bei Proteinen, die sich durch komplexere, gekoppelte „Fitnesslandschaften“ mit mehreren lokalen Optima und limitierten Pfaden im Sequenzraum auszeichnen, ist deshalb die Einführung von Mehrfachmutationen für eine erfolgreiche Evolution erforderlich. Genutzt werden dafür beispielsweise die Rekombination (*gene shuffling*, *molecular breeding*, Stemmer, 1994a, Stemmer, 1994b), höhere Mutageneseraten innerhalb beschränkter Proteinregionen und die Sättigungs-mutagenese an bestimmten Positionen, welche vorher anhand von Sequenz- und Strukturanalysen oder durch Zufallsmutagenese identifiziert wurden (Miyazaki & Arnold, 1999, Miyazaki *et al.*, 2000). Die erzeugten Mehrfachmutationen machen unter Umständen die Unterscheidung von funktionalen und neutralen Austauschen erforderlich, zum Beispiel durch Rückkreuzung (Stemmer, 1994b) oder Doppel- und Dreifachmutationszyklen (Horovitz, 1996, Horovitz & Fersht, 1992).

Allgemein wird die für eine *in vitro*-Evolution optimale Mutageneserate bestimmt durch die Größe der selektierbaren Bibliothek und die Struktur der „Fitnesslandschaft“ des zu optimierenden Proteins, d.h. durch das Ausmaß der koppelnden Wechselwirkungen zwischen einzelnen Resten. Beispielsweise ist die gezielte Mutagenese an Positionen der Proteinoberfläche sehr gut geeignet, um thermodynamisch stabilisierte Varianten zu selektieren, da solvensexponierte Reste wesentlich weniger spezifische Interaktionen eingehen und daher geringer gekoppelt sind als Reste im Proteininneren. Geeignete Positionen für Sättigungsmutagenesen lassen sich durch Sequenz- und Strukturvergleiche (Teilarbeit B), elektrostatische Berechnungen (Grimsley *et al.*, 1999) oder Anwendung von Computeralgorithmen (Dahiyat & Mayo, 1997, Dahiyat *et al.*, 1997) identifizieren.

1.5 *In vitro*-Selektionsmethoden

Voraussetzung für die Suche nach optimierten, z.B. thermodynamisch stabilisierten Proteinmutanten in einem großen Repertoire von Sequenzen ist eine effiziente Durchmusterungs- (*screening*-) oder Selektionsmethode. Während beim *screening* jede einzelne Variante untersucht und diejenige mit den gewünschten Eigenschaften gezielt ausgewählt werden muss, reichern sich bei einer Selektion die verbesserten Varianten entsprechend dem angelegten Selektionsdruck von selbst an (Kuchner & Arnold, 1997). Bei dieser Evolution ist die Reproduktion der selektierten Spezies nötig. Für Proteine bedeutet dies, dass sie mit der Information für ihre Vervielfältigung, d.h. der kodierenden Nukleinsäure verbunden sein müssen. Wenn die Funktion eines Proteins mit dem Wachstum oder dem Überleben eines Mikroorganismus gekoppelt werden kann, ist eine Selektion *in vivo* möglich (Kast & Hilvert, 1997, Oshima, 1994, Matsumura & Aiba, 1985). Für Selektionen *in vitro* werden verschiedene sogenannte Displaytechniken genutzt (Dower & Mattheakis, 2002). Das am häufigsten verwendete Verfahren ist das Phagendisplay (Dunn, 1996, Smith & Scott, 1993), bei dem das zu optimierende Protein mit dem Gen-3- oder Gen-8-Protein filamentöser Bakteriophagen fusioniert und so auf deren Oberfläche präsentiert wird. Damit vergleichbar, wenn auch deutlich komplexer, ist das Zelloberflächendisplay bei Bakterien (Georgiou *et al.*, 1997) oder Hefezellen

(Schreuder *et al.*, 1996). Das Gastprotein wird dabei mit zellulären Proteinen der äußeren Membran, Pilinen oder Flagellinen, fusioniert. Displaytechniken, die völlig auf *in vivo*-Schritte verzichten, sind das Ribosomendisplay (Hanes & Plückthun, 1997) und die mRNA-Protein-Fusion (Roberts & Szostak, 1997). Die Verknüpfung von Protein und kodierender Nukleinsäure wird dabei durch das gestoppte, im Komplex stabilisierte Ribosom vermittelt bzw. über eine kovalente Bindung via Puromycin erreicht.

All diese Displaytechniken haben gemeinsam, dass die Selektion über spezifische Bindungseigenschaften erfolgt. Da die Ligandenbindung und die Proteinstabilität thermodynamisch gekoppelt sind, können diese Techniken im Prinzip auch zur Selektion von Proteinvarianten mit erhöhter Stabilität genutzt werden (Jung & Plückthun, 1997, Proba *et al.*, 1998, Jackson *et al.*, 1995, Spada *et al.*, 1998).

1.6 Das Proside-Verfahren zur Selektion thermodynamisch stabilisierter Proteine

Proside (protein stability increased by directed evolution, Sieber *et al.*, 1998) ist eine generell anwendbare Selektionsmethode zur Stabilisierung von Proteinen. Sie ist unabhängig von spezifischen Eigenschaften des Proteins wie Ligandenbindung oder enzymatischer Aktivität. Proside verknüpft die proteolytische Stabilität eines Proteins mit einer sehr gut selektierbaren Eigenschaft, nämlich der Infektiosität filamentöser Phagen.

Die proteolytische und die konformationelle Stabilität eines Proteins sind korreliert, da der denaturierte Zustand in der Regel wesentlich leichter durch Proteasen gespalten wird als der gefaltete native Zustand (Parsell & Sauer, 1989). Dies gilt insbesondere für Chymotrypsin. Diese Protease spaltet spezifisch an aromatischen Aminosäuren, die im gefalteten Protein meist verborgen und damit unzugänglich sind.

Verantwortlich für die Infektiosität der Phagen gegenüber *E. coli* ist das Gen-3-Protein (G3P), das in drei bis fünf Kopien an einem Ende des Phagenpartikels vorkommt. G3P besteht aus drei Domänen, N1 (68 Aminosäuren), N2 (131 Aminosäuren) und CT (150 Aminosäuren), verknüpft durch glycinreiche Linker von 18 bzw. 39 Aminosäuren (Marvin, 1998, Lubkowski *et al.*, 1999, Riechmann & Holliger, 1997). Die CT-Domäne verankert das Protein in der Phagenhülle (Boeke & Model, 1982, Stengele *et al.*, 1990). Domäne N2 initiiert die Infektion von *E. coli* durch Bindung an den primären Rezeptor, die Spitze des F-Pilus (Deng & Perham, 2002), und die N1-Domäne ermöglicht durch ihre Wechselwirkung mit dem Corezeptor TolA nach Zurückziehen des F-Pilus den Eintritt des Phagen in die Zelle (Click & Webster, 1997, Glaser-Wuttke *et al.*, 1989). Damit der Phage infektiös ist, müssen demnach die Domänen von G3P miteinander verknüpft sein (Stengele *et al.*, 1990). Die Proside-Selektionsmethode macht sich zunutze, dass diese Verknüpfung auch durch Gastproteine erfolgen kann, die zwischen die Domänen inseriert sind (Krebber *et al.*, 1997). Ein Repertoire von Sequenzen, die das zu stabilisierende Protein kodieren, wird zwischen die CT- und N2-Domäne eingebaut (Abb. 1) und die resultierende Bibliothek von Phagen einer *in vitro*-Proteolyse mit Trypsin oder Chymotrypsin ausgesetzt. Die Proteolysebedingungen werden dafür so gewählt, dass die Phagenproteine gefaltet bleiben, das eingebaute Gastprotein in seiner Wildtypform jedoch bereits partiell entfaltet ist. Die für die Infektiosität essentielle Verknüpfung der G3P-Domänen bleibt also nur für Varianten des Gastproteins erhalten, die im Vergleich zum Wildtypprotein thermodynamisch stabilisiert, d.h. gefaltet und damit proteaseresistent sind. In mehreren

Zyklen aus Selektion durch *in vitro*-Proteolyse, Infektion von *E. coli* und Phagenpropagation lassen sich diese Varianten in der Bibliothek stark anreichern und identifizieren. Volker Sieber konnte bereits das Potential der Proside-Selektionsmethode zeigen, indem er stabilisierte Varianten von RNase T1 aus einer Bibliothek mit drei randomisierten Positionen selektierte (Sieber *et al.*, 1998).

1.7 Das Kälteschockprotein CspB aus *Bacillus subtilis*, ein Modellprotein zur Untersuchung der Proteinstabilität

Die Kälteschockproteine (Csp) stellen eine Familie sequenzhomologer Proteine dar, die im Rahmen der Kältschockantwort verschiedener Organismen, u. a. *E. coli* und *Bacillus subtilis*, verstärkt exprimiert werden (Graumann & Marahiel, 1999, Graumann & Marahiel, 1996). Sie binden hochaffin an RNA und einzelsträngige DNA (Graumann & Marahiel, 1998, Lopez *et al.*, 1999), so dass ihnen *in vivo* eine Funktion als RNA-Chaperone zur Verhinderung von Sekundärstrukturen in mRNAs bei niedrigen Temperaturen zugeschrieben wird (Jiang *et al.*, 1997). Die zugrunde liegende Fähigkeit zum Aufschmelzen von Nukleinsäuren ist auch verantwortlich für ihre potentielle Wirkung als Transkriptionsantiterminatoren (Bae *et al.*, 2000, Phadtare *et al.*, 2002a; Phadtare *et al.*, 2002b). Die Kälteschockproteine sind klein (63 bis 68 Aminosäuren) und monomer. Sie gehören zur *Ob-fold* (*oligosaccharide-oligonucleotide binding fold*) Faltungsfamilie, und ihre Struktur ist durch ein β-Fass aus 5 antiparallelen β-Strängen gekennzeichnet, α-Helices fehlen (Schindelin *et al.*, 1994; Schindelin *et al.*, 1993, Schnuchel *et al.*, 1993, Newkirk *et al.*, 1994, Müller *et al.*, 2000). Da sie reversibel nach dem Zweizustandsmodell ohne Intermediate entfalten, wurden sie als Modellproteine genutzt, um die Prinzipien der Faltung globulärer Eindomänenproteine aufzuklären (Schindler *et al.*, 1995, Reid *et al.*, 1998, Perl *et al.*, 1998, Perl *et al.*, 2001).

Das Protein aus dem mesophilen Bakterium *Bacillus subtilis*, *Bs*-CspB, besitzt nur geringe thermodynamische Stabilität, sein Homologes *Bc*-Csp aus dem thermophilen Organismus *Bacillus caldolyticus* ist nahezu doppelt so stabil (Perl *et al.*, 1998). Beide unterscheiden sich in lediglich 12 von 67 Resten und stellen somit ein sehr gut geeignetes System zur Untersuchung der molekularen Ursachen der Thermostabilität dar. Dieter Perl analysierte mittels ortsgerichteter Mutagenese die Stabilitätseinflüsse aller 12 Aminosäureaustausche und konnte zwei exponierte Reste in *Bc*-Csp identifizieren, Arg3 und Leu66, die allein für den großen Stabilitätsunterschied zu *Bs*-CspB verantwortlich sind (Perl *et al.*, 2000, Perl & Schmid, 2001).

1.8 Problemstellung

Ziel der hier vorliegenden Arbeit war zunächst die Weiterentwicklung und Optimierung des Proside-Verfahrens zur *in vitro*-Selektion thermodynamisch stabilisierter Proteine. Einen Schwerpunkt bildete dabei die Vergrößerung der selektierbaren Bibliotheken unterschiedlicher Varianten eines Proteins durch Optimierung des Phagenkonstrukts und durch Veränderungen der molekularbiologischen Methoden, die zum Erstellen dieser Bibliotheken genutzt werden. In diesem Zusammenhang galt es auch, die rekombinationsbedingte unerwünschte Eliminierung von Gastproteinsequenzen aus dem Phagengenom zu verhindern bzw. zu reduzieren.

Die erfolgreiche Stabilisierung von RNase T1 hatte bereits das dem Proside-Verfahren zugrunde liegende Selektionsprinzip bestätigt (Sieber *et al.*, 1998, Martin, 1998). Es wurde allerdings noch nicht gezeigt, inwieweit sich die Methode generell zur Evolution globulärer Proteine nutzen lässt. RNase T1 war als disulfidfreie, stark labilisierte Variante in den Phagen eingebaut worden, so dass die proteolytische *in vitro*-Selektion stabilisierter Mutanten unter Nativbedingungen durchgeführt werden konnte. Im Gegensatz dazu erfordert die Proside-Selektion anderer, bereits in ihrer Wildtypform stabilerer Proteine eine Proteolyse unter denaturierenden Bedingungen. Dies hat zur Folge, dass auch der Phage, insbesondere das Gen-3-Protein, diesem erhöhten Selektionsdruck ausgesetzt ist und in seiner Stabilität optimiert werden muss, um nicht selbst limitierend für die Selektion zu wirken.

Für die thermodynamische Stabilisierung mittels Proside wurde das Kältschockprotein *Bs*-CspB aus *Bacillus subtilis* ausgewählt. Es ist klein, globulär und monomer, und seine Entfaltung verläuft sowohl thermisch, als auch denaturierungsmittelinduziert reversibel nach dem Zweizustandsmodell (Schindler *et al.*, 1995, Perl, 1997). Wie schon erwähnt unterscheidet sich *Bs*-CspB nur in 12 exponierten Resten von seinem thermostabilen Homologen Bc-Csp. Durch ortsgerichtete Mutagenese (Perl & Schmid, 2001) können jedoch nur relativ wenig Aminosäureaustausche untersucht werden, und es ist schwer bzw. gar nicht möglich, das Stabilisierungspotential einzelner Positionen im Detail zu analysieren und generelle Stabilitätsprinzipien abzuleiten. Die gerichtete *in vitro*-Evolution bietet dagegen die Möglichkeit, einen sehr großen Sequenzraum nach der stabilsten Variante zu durchsuchen und die optimalen Aminosäurekombinationen an bestimmten Positionen zu identifizieren. Die Stabilisierung von *Bs*-CspB mittels Proside, in Kombination mit der thermodynamischen Charakterisierung selektierter Varianten und der Interpretation von Austauschen anhand hochaufgelöster Proteinstrukturen, sollte Einblicke in die molekularen Ursachen der Thermostabilität liefern. Basierend auf den bereits vorhandenen Mutagenesedaten galt es insbesondere, die Rolle solvensexponierter Reste zu untersuchen.

Bei einer Vielzahl von Phagendisplaymethoden, so auch bei Proside, werden die zu selektierenden Gastproteine mit dem Gen-3-Protein der Phagen fusioniert. Zu den Stabilitäts- und Faltungseigenschaften dieses Proteins ist bisher allerdings fast nichts bekannt. Auch die für die Infektion essentielle Trennung der Domänen wurde bislang nur anhand der Erkenntnisse aus den Strukturen der Komplexe mit den zellulären Rezeptoren beschrieben. Ein weiteres Ziel der Arbeit war deshalb, G3P in isolierter Form thermodynamisch zu charakterisieren. Besonderes Interesse galt der Aufklärung des Faltungsmechanismus des N-terminalen Fragments bestehend aus den Domänen N1 und N2. Beide Domänen stellen strukturell separate Einheiten dar, zeigen zugleich aber extensive Interdomänenwechselwirkungen. Am Beispiel dieses Zweidomänenproteins sollte untersucht werden, ob und wenn ja wie die Faltung der individuellen Domänen mit deren Assoziation zusammenhängt und welche Stabilitäts- und Faltungsprinzipien bei Mehrdomänenproteinen allgemein von Bedeutung sind.

2 Zusammenfassung und Diskussion der Ergebnisse

2.1 Weiterentwicklung des Proside-Selektionsverfahrens für stabilisierte Proteine

2.1.1 Optimierung des Phagenkonstrukts und der Methoden zum Erstellen von Mutantenbibliotheken

Das Proside-Selektionssystem konnte bereits erfolgreich zur Stabilisierung von RNaseT1 eingesetzt werden (Sieber *et al.*, 1998, Martin, 1998). Das dafür verwendete Phagenkonstrukt basierte auf dem Phagmid fCKCBS (Krebber *et al.*, 1997) mit Gastprotein, inseriert zwischen der N2- und CT-Domäne von G3P. Sequenzhomologe Bereiche innerhalb der flankierenden Linker in G3P waren bereits ausgetauscht worden, um homologe Rekombination zu vermeiden (Sieber *et al.*, 1998). Trotzdem kam es im Verlauf der Selektionen bei einzelnen Spezies zur Eliminierung des Gastproteins aus dem Phagengenom und damit zur Bildung extrem proteaseresistenter Wildtypphagen, die bereits nach wenigen Selektionszyklen die Bibliotheken dominierten. Um stabilisierte Proteinvarianten mit dem Proside-System selektieren zu können, mussten diese Phagen regelmäßig aus den Repertoires entfernt werden, was durch Amplifikation der Gastproteingene und Neukonstruktion der Bibliotheken erfolgte. Vor allem die illegitime Rekombination (Ehrlich *et al.*, 1993, Kuzminov, 1999) scheint für den Verlust der inserierten Proteine verantwortlich zu sein, der trotz fehlender homologer Bereiche und der Verwendung des RecA-defizienten *E. coli*-Stammes XL1Blue zur Vermehrung der Phagen weiterhin auftrat. Dabei erhielten speziell die Phagen einen hohen selektiven Vorteil, bei denen die Rekombination innerhalb der Linker zwischen Gastprotein und G3P stattfand. Nur in diesen Fällen blieben die Domänen von G3P intakt, und das inserierte Protein wurde nahezu vollständig entfernt, so dass keine unstrukturierten proteasesensitiven Fragmente zurückblieben. Begünstigt wurde eine derartige Rekombination weil die Linker, die das Gastprotein flankierten, mit 21 bzw. 156 Nukleotiden relativ lang waren. Der Linker zur CT-Domäne wurde deshalb auf 24 Nukleotide verkürzt und dabei gleichzeitig mit weiteren Restriktionsschnittstellen zur Klonierung von Gastproteinen versehen (Teilarbeit A). Auf die Funktion des Gen-3-Proteins, d.h. die Infektiosität der Phagen gegenüber *E. coli*, hatte diese Verkürzung des Linkers keinen Einfluss, auch nicht bei Gastproteinen wie dem λ-Repressor, der β1-Domäne des Staphylokokkenproteins G oder humanem FKBP12, deren N- und C-Termini in der nativen Struktur relativ weit voneinander entfernt sind. Mit dem kurzen Linker konnte der Verlust der inserierten Gene durch Rekombination im Verlauf der Selektionen zwar nicht vollständig verhindert, jedoch stark reduziert werden.

Bei der ersten Anwendung von Proside (zur Stabilisierung von RNaseT1) wurden nur relativ kleine Bibliotheken aus maximal 6500 verschiedenen Varianten selektiert (Martin, 1998). Um alle Mutanten mit 99 %iger Wahrscheinlichkeit zu repräsentieren waren daher lediglich $3 \cdot 10^5$ Klone nötig (vgl. Clarke & Carbon, 1976). Konstruiert wurden die Bibliotheken durch Hybridisierung mutierter, für das Gastprotein kodierender DNA-Fragmente an den (+)-DNA-Strang von Phagen mit inseriertem Wildtypgen, gefolgt von der Synthese des kompletten (-)-Strangs durch *Pfu*-DNA-Polymerase und anschließender Transformation von *E. coli* XL1Blue mit den resultierenden doppelsträngigen Hybriden. Die Ausbeuten waren hoch genug, um die erforderlichen Bibliotheksgrößen zu erreichen und bei der Selektion die zahlenmäßige Dominanz von Phagen mit Wildtypinsert (mindestens 50% des initialen Repertoires) zu tolerieren. Für das Erstellen größerer,

möglichst umfassender Bibliotheken war die Effizienz dieses Verfahrens allerdings nicht ausreichend. Außerdem war zu beobachten, dass Sequenzbereiche mit Mutationen nahe des 5´-Endes der hybridisierten DNA-Fragmente (C-terminale Mutationen im Gastprotein) häufig verloren gingen, entweder aufgrund der Exonuclease- und Strangverdrängungs-aktivität der eingesetzten DNA-Polymerase oder durch Reparaturmechanismen in *E. coli*, die vermutlich vom noch vorhandenen Einzelstrangbruch im neusynthetisierten (-)-Strang der Phagen-DNA ausgingen (Lindahl, 1982). Der Versuch, durch dam-Methylierung der DNA vor der Transformation eine eventuelle *mismatch*-Reparatur in *E. coli* zu unterdrücken, zeigte keinen Erfolg beim Erhalt terminaler Mutationen im inserierten Gen. Auch die Verlängerung der zur Hybridisierung eingesetzten mutierten DNA-Fragmente um 50 Basen über die letzte Mutation hinaus sowie die Verwendung von Oligonukleotiden mit nicht-hydrolysierbaren Thiophosphatbindungen zur Herstellung dieser Fragmente (Ott & Eckstein, 1987, Skerra, 1992) führten zu keiner Verbesserung. Erfolgreich war jedoch der Einsatz der thermostabilen *Taq*-Ligase bei der Synthese des (-)-Strangs (Teilarbeit A). Durch das Schließen des Einzelstrangbruchs am 5´-Ende des hybridisierten Fragments nach vollständiger Synthese des (-)-Stranges wurde die Eliminierung von Mutationen deutlich unterdrückt. Gleichzeitig konnten damit die Ausbeuten an doppelsträngigem Hybrid und dessen Transformationseffizienz gesteigert werden, wodurch Bibliotheksgrößen von bis zu 10^8 Mutanten möglich wurden.

Da sich bei Bibliotheken mit theoretisch 10^7 oder 10^8 verschiedenen Varianten die Transformationseffizienz von *E. coli* bereits limitierend auswirkt, können die einzelnen Spezies nur noch mit geringerer Wahrscheinlichkeit repräsentiert werden. Unter Umständen kommen Phagen mit bestimmten stabilen Varianten des Gastproteins lediglich in wenigen Kopien im Repertoire vor. Dass bei solch niedrigen Kopienzahlen die Regeln der statistischen Thermodynamik nur noch eingeschränkt gelten, macht die Selektion basierend auf dem Faltungszustand eines Proteins schwierig. Der selektive Vorteil von Phagen mit Wildtypinsert nimmt, trotz geringerer thermodynamischer Stabilität, aufgrund ihrer zahlenmäßigen Dominanz in den initialen Bibliotheken stark zu.

Um die Zahl dieser Phagen zu reduzieren, wurde in die zum Erstellen der Bibliotheken genutzte Matrizen-DNA direkt benachbart zum Wildtypgen des Gastproteins ein amber-Kodon eingebaut (Abb. 1). In den mutierten DNA-Fragmenten, die an diese Matrize hybridisiert und zum kompletten (-)-Strang ergänzt wurden, war das Stopkodon nicht enthalten. Mit den resultierenden doppelsträngigen Hybriden wurde zunächst der Suppressor-Stamm XL1Blue (*sup*E44) transformiert, der dann sowohl Phagen mit mutiertem Gastprotein (kodiert durch den neusynthetisierten (-)-Strang), als auch Phagen mit Wildtypinsert (kodiert durch den Matrizen-(+)-Strang mit amber-Kodon) produzierte. Die Phagen mit Wildtypinsert ließen sich anschließend selektiv durch Infektion des Nichtsuppressor-Stammes *E. coli* ABLE K eliminieren. Das amber-Kodon der Wildtyp-kopie verhinderte in diesem Nichtsuppressor-Stamm die vollständige Translation von G3P, so dass die entsprechenden Phagen von den Zellen nicht freigesetzt werden konnten und es zur Bildung von Polyphagen auf der Zelloberfläche kam (Rakonjac & Model, 1998). Außer den Phagen mit Wildtypinsert konnten mit diesem Verfahren ebenso diejenigen Varianten aus den initialen Bibliotheken entfernt werden, bei denen der Verlust C-terminaler Mutationen durch Reparatur oder Rekombination zum Auftreten des amber-Kodons auch im mutierten Strang geführt hatte. Von besonderer Bedeutung war dies unter anderem für die Proside-Selektion von *Bs*-CspB (Teilarbeit B), bei der unter anderem auch die Positionen 64, 66 und 67 direkt am C-Terminus des Proteins randomisiert wurden.

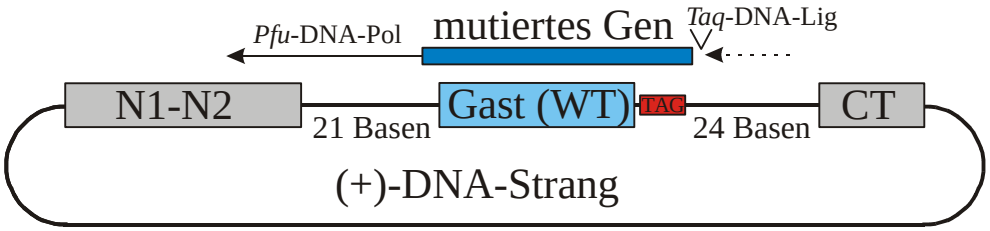


Abbildung 1. Schematische Darstellung des Phagenkonstrukts zur Proside-Selektion stabilisierter Proteinvarianten. Die Domänen N1, N2 und CT des Gen-3-Proteins sowie das inserierte Gastprotein in seiner Wildtypform (WT) sind hervorgehoben, ebenso wie das amber-Kodon direkt benachbart zum Gastprotein. Um die Bibliotheken von Varianten eines Proteins im Phagen zu erstellen, werden die mutierten Fragmente des Gastgens mit dem (+)-DNA-Strang des Phagen hybridisiert und durch *Pfu*-DNA-Polymerase zum kompletten (-)-Strang ergänzt, wobei der verbleibende Einzelstrangbruch durch *Taq*-DNA-Ligase geschlossen wird.

2.1.2 Selektion stabilisierter Varianten des Gen-3-Proteins filamentöser Phagen

Die *in vitro*-Proteolyse der Phagenbibliothek bildet das Herzstück des Proside-Selektionssystems. Sie wird unter Bedingungen durchgeführt, bei denen das Gastprotein in seiner Wildtypform bereits partiell entfaltet ist. Nur so lässt sich die Korrelation zwischen Proteaseresistenz und thermodynamischer Stabilität zur Diskriminierung unterschiedlicher Varianten nutzen. Voraussetzung ist allerdings, dass die Proteine des Phagen selbst, insbesondere das Gen-3-Protein, unter den Selektionsbedingungen stabil gefaltet bleiben. G3P ist schon bei Raumtemperatur sensitiv gegenüber Subtilisin und Proteinase K, gegenüber Chymotrypsin und Trypsin ist es allerdings resistent (Sieber *et al.*, 1998). Diese Proteasen sind aufgrund ihrer Spezifität ideal für Proside geeignet. Oberhalb von 40 °C wird G3P aber auch durch diese Enzyme gespalten.

Für die ersten Proside-Selektionen von RNaseT1 war die Stabilität des Wildtyp-G3P ausreichend, da die RNase in ihrer disulfidfreien, stark destabilisierten Form in den Phagen eingebaut wurde (Sieber *et al.*, 1998). Um das Selektionssystem auch zur Optimierung stabilerer Proteine nutzen zu können, wurde Proside auch eingesetzt, um G3P selbst zu stabilisieren. Da für G3P keine thermodynamischen Stabilitätsdaten verfügbar waren und auch keine thermostabilen Homologen für Sequenzvergleiche existieren, wurde zur Stabilisierung die adaptive Evolution, d.h. die schrittweise Einführung zufälliger Mutationen, gefolgt von proteolytischer Selektion gewählt. Insertfreie Phagen wurden dazu im Mutator-Stamm *E. coli* XL1Red vermehrt und durch *in vitro*-Proteolyse bei zunächst 45 °C und in den folgenden Zyklen bei schrittweise bis auf 55 °C gesteigerten Temperaturen selektiert. Die erste so erhaltene stabilisierende Mutation, T101I, war in Domäne N2 lokalisiert. Nach erneuter Propagation der Phagen in XL1Red und Selektion über mehrere Zyklen wurde der Austausch D209Y, ebenfalls in N2, erhalten. Durch weitere *in vivo*-Mutagenesen und Selektionszyklen bei 55 °C bzw. 57,5 °C wurden die Dreifachmutante T13I/T101I/D209Y und, davon ausgehend, schließlich die Vierfachmutante T13I/T101I/Q129H/D209Y gefunden. Drei der vier stabilisierenden Mutationen, T101I, Q129H und D209Y, befinden sich in der N2-Domäne, nur T13I in Domäne N1 (Abb. 1a in Teilarbeit A).

Bei der späteren thermodynamischen Charakterisierung von G3P stellte sich heraus, dass die N2-Domäne tatsächlich deutlich labiler ist als N1 (Teilarbeit D). Das sukzessive Erscheinen der Mutationen im Verlauf dieses Evolutionsexperiments spiegelte deren Beiträge zur thermodynamischen und damit proteolytischen Stabilität von G3P wider (Abb. 2). Nachdem Domäne N2 durch die Austausche T101I und D209Y stabilisiert war, wurde die Gesamtstabilität von G3P nun durch N1 limitiert. Die dritte Mutation, T13I, trat deshalb in Domäne N1 auf, bevor mit Q129H die Domäne N2 weiter stabilisiert wurde. Alle vier Mutationen zusammen erhöhen die Temperatur, bei der die Phagen sensitiv gegenüber Spaltung durch Chymotrypsin werden, von 40 °C auf fast 60 °C. Das Proside-System lässt sich mit diesem stabilisierten Phagen für thermische Selektionen bei bis zu 60 °C einsetzen, ohne dass die Stabilitäten der Phagenproteine selbst limitierend für die *in vitro*-Evolution werden. Die Auswirkungen der selektierten Mutationen auf die konformationelle Stabilität von G3P werden in Teilarbeit D analysiert.

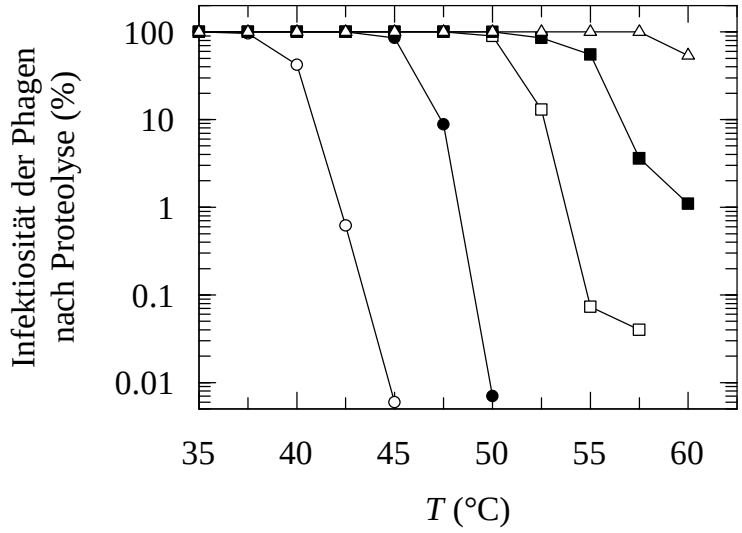


Abbildung 2. Erhaltene Infektiositäten des Wildtypphagen fd (○) und der selektierten Phagenvarianten mit den Gen-3-Protein-Mutationen T101I (●), T101I/D209Y (□), T13I/T101I/D209Y (■) und T13I/T101I/Q129H/D209Y (△) nach *in vitro*-Proteolyse für 15 min mit 0,25 µM Chymotrypsin bei unterschiedlichen Temperaturen. Die Proteolyse wurde durchgeführt wie in den Teilarbeiten A und D beschrieben.

2.2 Untersuchung der Prinzipien der Proteinstabilität am Beispiel von *Bs*-CspB

2.2.1 *In vitro*-Evolution von *Bs*-CspB mittels Proside

Proteine aus thermophilen Organismen zeichnen sich häufig dadurch aus, dass an ihren Oberflächen vermehrt geladene Aminosäuren auftreten (Jaenicke & Böhm, 1998), die möglicherweise über zusätzliche ionischen Wechselwirkungen stabilisierend wirken. Die Verteilung der Ladungen auf der Oberfläche ist dabei von besonderer Bedeutung, wie durch Experimente mit Ubiquitin gezeigt werden (Loladze *et al.*, 1999, Grimsley *et al.*, 1999, Spector *et al.*, 2000). Oberflächenladungen sind auch für die Stabilität der Kälteschock-proteine wichtig. *Bs*-CspB und sein thermostabiles Homologes *Bc*-Csp unterscheiden sich an 12 Positionen, die alle an der Oberfläche der Proteine liegen, und fünf dieser Aminosäureaustausche betreffen geladene Reste (Abb. 3).

	1 5 10 15 20 25 30
***Bs*-CspB**	M**L****E**GKVKWFN**S**EKG**F**GFIEVEG**Q****D**DVFVHF**S**AIQ
***Bc*-Csp**	M**Q****R**GKVKWFN**N**EKG**Y**GFIEVEG**G****S**DVFVHF**T**AIQ
	35 40 45 50 55 60 65
***Bs*-CspB**	GEGFKTLEEGQ**A**VSFEIV**E**GNRGPQAANV**T**K**E****A**
***Bc*-Csp**	GEGFKTLEEGQ**E**VSFEIV**Q**GNRGPQAANV**V**K**L**-

Abbildung 3. Sequenzvergleich der Kälteschockproteine *Bs*-CspB und *Bc*-Csp. Die zwölf Sequenz-unterschiede sind fett hervorgehoben. Positiv geladene Aminosäuren an diesen zwölf Positionen sind blau, negative geladene rot eingefärbt.

Wie Abbildung 6 zeigt, weist die Oberfläche von *Bs*-CspB eine Region sehr hoher negativer Ladungsdichte auf. Verantwortlich dafür sind vor allem die Glutamatreste 3 und 66, deren Seitenketten in der Kristallstruktur des Proteins allerdings nicht aufgelöst sind (Schindelin *et al.*, 1993). Die Modellierung der beiden Reste zeigt aber, dass ihre Carboxylgruppen vermutlich weniger als 5 Å voneinander entfernt sind, was zu elektro-statischer Abstoßung führen könnte (Abb. 4). Im thermostabilen *Bc*-Csp sind diese Glutamate durch Arg3 und Leu66 ersetzt, eine Abstoßung ist hier nicht möglich, und die Ladungen auf der Oberfläche sind deutlich gleichmäßiger verteilt (Abb. 6). Dieter Perl konnte mittels systematischer ortsgerichteter Mutagenese zeigen, dass tatsächlich nur die beiden Austausche E3R und E66L für die Stabilitätsdifferenz von 15,8 kJ/mol zwischen *Bs*-CspB und *Bc*-Csp verantwortlich sind (Perl *et al.*, 2000).

Mit Hilfe der *in vitro*-Evolution durch Proside sollte unter anderem das Stabilisierungs-potential an diesen Schlüsselpositionen 3 und 66 in *Bs*-CspB näher untersucht werden. In den vorangegangenen Mutageneseexperimenten waren lediglich die Austausche gegen die in *Bc*-Csp vorhandenen Aminosäuren charakterisiert worden. Inwieweit andere Mutationen vergleichbar oder eventuell sogar noch stärker stabilisierend wirken, war nicht bekannt.

Die gerichtete Evolution bietet die Möglichkeit, sehr komplexe Kombinatorik zu nutzen und einen riesigen Sequenzraum nach stabilisierten Varianten abzusuchen. Für Proside wurden daher neben den Resten 3 und 66 zusätzlich die vier räumlich benachbarten Positionen 2, 46, 64 und 67, an denen sich *Bs*-CspB und *Bc*-Csp ebenfalls unterscheiden, ausgewählt (Abb. 4). Mit der gleichzeitigen Randomisierung und Selektion dieser sechs

oberflächenexponierten Reste sollte nach eventuellen alternativen Aminosäurekombinationen gesucht werden, deren Vergleich mit Sequenzen anderer thermostabiler Homologe Hinweise zu generellen Stabilisierungsprinzipien der Kälteschockproteine geben könnten.

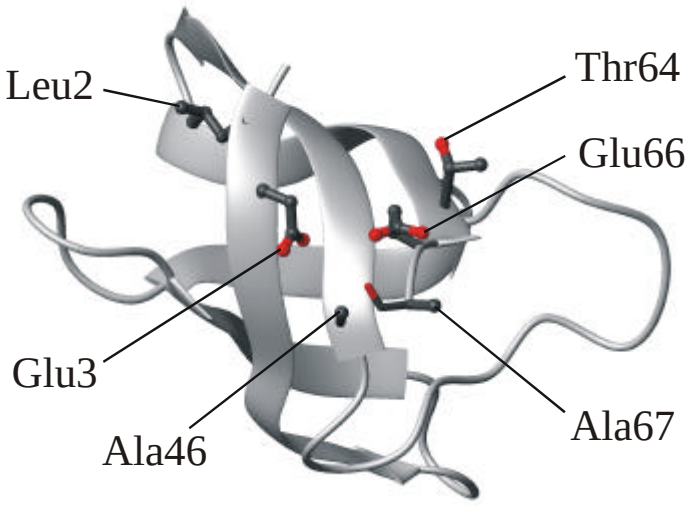


Abbildung 4. Tertiärstrukturdarstellung von *Bs*-CspB, basierend auf den Strukturdaten von Schindelin *et al.* (1993). Die Aminosäureseitenketten an den sechs für die *in vitro*-Evolution ausgewählten Positionen sind in der Kugel-Stab-Darstellung abgebildet. Die in der Kristallstruktur nicht aufgelösten Glutamatreste 3 und 66 wurden mit *Swiss-pdb viewer* (Guex & Peitsch, 1997) modelliert.

Zur Erstellung der Bibliothek von *Bs*-CspB-Varianten wurden in einer Sättigungsmutagenese alle 20 Aminosäuren an den Positionen 2, 3, 46, 64, 66 und 67 zugelassen. Zunächst erfolgte die Selektion dieser Bibliothek auf Stabilität gegenüber GdmCl-induzierter Entfaltung. Die Phagen wurden dazu in sechs Selektionszyklen der *in vitro*-Proteolyse mit 0,25 µM Chymotrypsin in Gegenwart von 1,5 M und später 1,75 M GdmCl bei 25 °C ausgesetzt. Nach dem letzten Zyklus konnten fünf verschiedene Varianten identifiziert werden. Die überexprimierten und gereinigten Proteine waren alle deutlich stabiler als das Wildtypprotein, und ihre Mittelpunkte des thermischen Entfaltungsübergangs (t_M) lagen zwischen 13 und 22 Grad über dem des *Bs*-CspB von 54 °C. Die stabilste selektierte Variante G-5 (Tab. 1 und Abb. 5A) erreichte mit t_M = 76,0 °C nahezu die Stabilität des thermostabilen Referenzproteins *Bc*-Csp (t_M = 76,9 °C), obwohl sie nur an einer der sechs randomisierten Positionen (Leu66) mit dem thermostabilen Protein übereinstimmte. Der große Beitrag der E66L-Mutation zur Thermostabilisierung von *Bs*-CspB kam auch dadurch zum Ausdruck, dass vier der fünf selektierten Varianten diesen Austausch zeigten.

Insgesamt konnten bei dieser Selektion kaum geladene Reste gefunden werden, vermutlich weil das ionische Denaturierungsmittel GdmCl verwendet wurde. Es schirmt Ladungen an der Proteinoberfläche ab, schwächt stabilisierende elektrostatische Wechselwirkungen und vermindert so die Vorteile der entsprechenden Varianten in der Selektion. Beispielsweise trat nur in einer der Varianten ein Arginin an Position 3 auf, obwohl dieser Rest eine Schlüsselrolle für die hohe Stabilität des thermostabilen *Bc*-Csp spielt (Perl *et al.*, 2000).

In einer zweiten Selektion wurde deshalb die Phagenbibliothek bei erhöhter Temperatur und niedriger Ionenstärke auf stabilisierte Varianten von CspB hin selektiert. Über insgesamt 10 Runden bei 57,5 °C wurde der Selektionsdruck der Proteolyse durch schritt-

weise Steigerung der Chymotrypsinkonzentration erhöht. Die 22 im Verlauf der Selektion identifizierten Mutanten, von denen 15 gereinigt und thermodynamisch charakterisiert wurden, sind in Tabelle 2 in Teilarbeit B aufgelistet. Nach der letzten Selektionsrunde zeigten sich an drei der sechs randomisierten Positionen deutliche Präferenzen für bestimmte Aminosäuren. Alle Varianten enthielten einen positiv geladenen Rest, Arginin oder Lysin, an Position 64, bei acht von zehn Varianten war auch an Position 2 Arginin oder Lysin zu finden, und Position 66 wurde in ebenfalls acht von zehn Varianten von einer Aminosäure mit aliphatischer Seitenkette besetzt. An Position 3 stimmten alle selektierten Varianten darin überein, dass das Glutamat des Wildtypproteins eliminiert worden war. Eine positiv geladene Aminosäure, so wie Arginin in *Bc*-Csp, kam allerdings nur einmal vor. Bemerkenswerterweise zeigte die stabilste Variante der thermischen Selektion T-10/10 an keiner der sechs randomisierten Positionen eine Übereinstimmung mit dem thermostabilen Homologen *Bc*-Csp (Tab. 1). Mit einem Mittelpunkt des thermischen Entfaltungsübergangs von t_M = 82,2 °C war diese Variante aber deutlich stabiler als *Bc*-Csp (Abb. 5A) und erreichte sogar die Stabilität von *Tm*-Csp (t_M = 82,0 °C), dem Kälteschockprotein des hyperthermophilen Organismus *Thermotoga maritima* (Wassenberg *et al.*, 1999).

Protein	Pos. 2	Pos. 3	Pos. 46	Pos. 64	Pos. 66	Pos. 67	t_M (°C)	$\Delta\Delta G_D^{70°C}$ (kJ/mol)	$[GdmCl]_M$ (M)	$\Delta\Delta G_D^{2M}$ (kJ/mol)
Bs-CspB	**Leu**	**Glu**	**Ala**	**Thr**	**Glu**	**Ala**	54,0	0	1,52	0
G-5	**Tyr**	**Ile**	**Gln**	**Leu**	**Leu**	**Pro**	76,0	15,1	2,77	9,6
T-10/10	**Arg**	**Val**	**Leu**	**Arg**	**Val**	**Gly**	82,2	19,0	2,82	9,4
Bc-Csp	**Gln**	**Arg**	**Glu**	**Val**	**Leu**	**-**	76,9	15,6	2,68	8,6

Tabelle 1. Thermodynamische Parameter für das *Bs*-CspB-Wildtypprotein, die stabilste Variante G-5 der Selektion in Gegenwart von GdmCl, die stabilste Variante T-10/10 der thermischen Selektion, und *Bc*-Csp. Angegeben sind die Aminosäurereste an den sechs mutierten Positionen, der Mittelpunkt des thermischen Entfaltungsübergangs (t_M), die Änderung der freien Enthalpie bei 70 °C relativ zu *Bs*-CspB ($\Delta\Delta G_D^{70°C}$), der Mittelpunkt des GdmCl-induzierten Entfaltungsübergangs ($[GdmCl]_M$) sowie die Änderung der freien Enthalpie bei 2 M GdmCl, 25 °C relativ zu *Bs*-CspB ($\Delta\Delta G_D^{2M}$). Die Parameter wurden abgeleitet von den thermischen und GdmCl-induzierten Gleichgewichtübergängen in 100 mM Na-Cacodylat, pH 7,0, die Messungen erfolgten wie in Teilarbeit B beschrieben.

Die Abhängigkeit der Stabilität eines Proteins von der Salzkonzentration gibt Aufschluss über die elektrostatischen Wechselwirkungen an der Proteinoberfläche. Ionen schirmen exponierte Ladungen des Proteins ab und schwächen so deren Interaktionen. Anhand des Verlaufs von ΔG_D, bestimmt aus thermischen Entfaltungsübergängen in Gegenwart von 0 bis 2 M NaCl, lassen sich somit Aussagen über die Coulomb´schen Beiträge zur Gesamtstabilität eines Proteins treffen. Wie aufgrund der hohen negativen Oberflächenladungsdichte von *Bs*-CspB zu erwarten, wirken dessen elektrostatische Interaktionen insgesamt destabilisierend. Ihre Abschirmung bei höherer Ionenstärke stabilisiert das Protein, und ΔG_D nimmt durch Zugabe von 0,5 M NaCl um etwa 4 kJ/mol zu (Abb. 5B). Im Gegensatz dazu sinkt die Stabilität von *Bc*-Csp in Gegenwart von 0,5 M NaCl um ungefähr 1 kJ/mol. Diese Stabilitätsabnahme deutet auf günstige elektrostatische Wechselwirkungen hin, vermutlich bedingt durch die gleichmäßigere Verteilung der Oberflächenladungen. Oberhalb von 1,0 M NaCl nehmen die Stabilitäten beider Proteine im gleichen Ausmaß zu, was sich mit dem Hofmeistereffekt von NaCl erklären lässt (von Hippel & Wong, 1965, Baldwin, 1996).

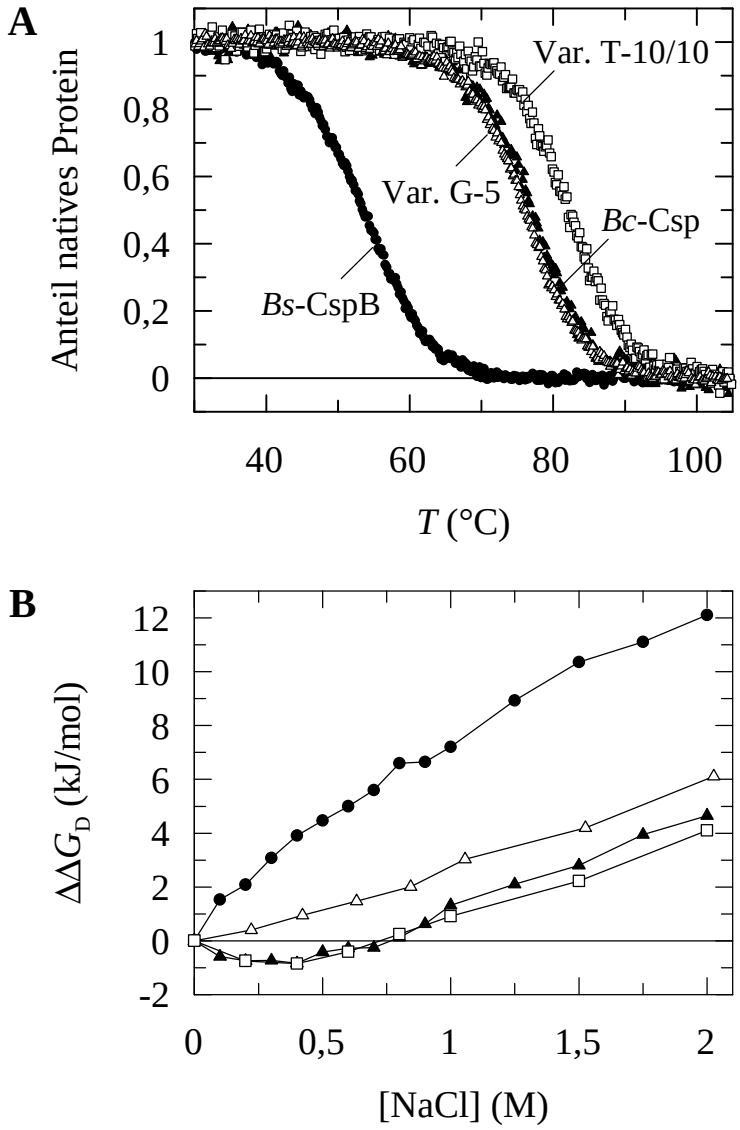
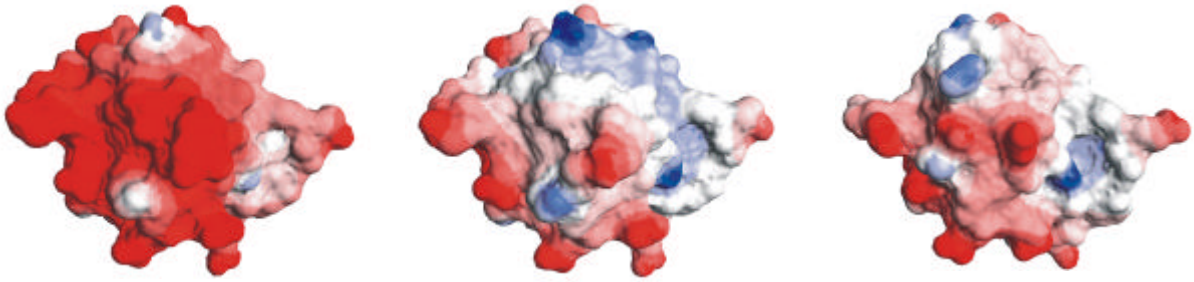


Abbildung 5. (A) Thermische Entfaltungsübergänge von *Bs*-CspB (●), der selektierten Varianten G-5 (△) und T-10/10 (▲) sowie von *Bc*-Csp (□) in 100 mM Na-Cacodylat, pH 7,0. Gezeigt ist der Anteil nativen Proteins als Funktion der Temperatur nach Analyse der Übergänge entsprechend dem Zweizustandsmodell. Die Messungen der Entfaltungsübergänge wurden durchgeführt wie in Teilarbeit B beschrieben.
(B) Änderung der freien Enthalpie der Entfaltung bei 70 °C als Funktion der NaCl-Konzentration relativ zur freien Enthalpie in Abwesenheit von NaCl ($\Delta\Delta G_D$) für *Bs*-CspB (●), G-5 (△), T-10/10 (▲) und *Bc*-Csp (□).

Wie das Salzprofil der stabilsten CspB-Variante G-5 aus der GdmCl-Selektion zeigt, wurden deren ionische Wechselwirkungen im Vergleich zum Wildtypprotein durch die Eliminierung von Glu3 und Glu66 zwar etwas verbessert, ihre Nettobeiträge zur Stabilität sind jedoch immer noch ungünstig. In Gegenwart des ionischen Denaturierungsmittels GdmCl spielten elektrostatische Beiträge offensichtlich nur eine untergeordnete Rolle bei der Selektion optimierter Varianten. Das Protein wurde primär durch verbesserte hydrophobe Wechselwirkungen stabilisiert. Dies macht sich auch bei der Stabilität gegenüber GdmCl-induzierter Entfaltung bemerkbar, die Variante G-5 ist hier sogar stabiler als *Bc*-Csp (Tab. 1).

Bei der thermischen Selektion hingegen wurden sowohl die hydrophoben als auch die elektrostatischen Interaktionen deutlich verbessert. Das Salzprofil der stabilsten Variante T-10/10 ähnelt der des thermostabilen *Bc*-Csp mit einem Stabilitätsminimum bei etwa 0,4 M NaCl (Abb. 5B). Die Ladungen sind auf der Proteinoberfläche dieser Variante wesentlich günstiger verteilt (Abb. 6) und die ionischen Beiträge wirken so insgesamt stabilisierend. Wie bei der stabilsten Variante aus der GdmCl-Selektion sind auch bei T-10/10 die sich potentiell abstoßenden Glutamate an 3 und 66 durch neutrale hydrophobe Reste ersetzt. Die günstigen elektrostatischen Wechselwirkungen in dieser Variante scheinen also auf die selektierten Argininreste an den Positionen 2 und/oder 64 zurückzuführen zu sein. Offenbar kommt den positiven Ladungen auf der Oberfläche der Kälteschockproteine eine besondere Bedeutung zu. Die molekularen Ursachen der erhöhten Thermostabilität werden in Teilarbeit C im Detail analysiert.



***Bs*-CspB** **T-10/10** ***Bc*-Csp**

Abbildung 6. Oberflächenladungen des *Bs*-CspB-Wildtypproteins, der stabilsten selektierten Variante T-10/10 sowie *Bc*-Csp. Die Oberflächen wurden generiert mit *Swiss-pdb viewer* (Guex & Peitsch, 1997) unter Berücksichtigung der geladenen Aminosäuren Asp, Glu, Arg und Lys sowie der N- und C-Termini. Potentiale von -10 kT/e oder weniger sind rot, neutrale Potentiale (0 kT/e) weiß und Potentiale von + 10 kT/e oder mehr blau gefärbt. Die Darstellungen basieren auf den Strukturdaten von Schindelin *et al.* (1993) für *Bs*-CspB und Müller *et al.* (2000) für *Bc*-Csp.

Die starke Stabilisierung von *Bs*-CspB durch Mutationen der solvensexponierten Reste 2, 3, 46, 64, 66 und 67 hatte auf die Nukleinsäurebindungsaktivität des Proteins keinen Einfluss. Die mittels Fluoreszenztitration bestimmten Affinitäten von Wildtyp-CspB und den beiden Varianten G-5 und T-10/10 für das Heptanukleotid $(dT)_7$ waren nahezu identisch (K_D = 7 nM bei 15 °C). Zwar handelt es sich bei der Bindung nicht um eine enzymatische Aktivität, die komplexe Konformationsänderungen des Proteins erfordert, aber trotzdem belegt dies, dass Thermostabilität und Aktivität bei niedriger Temperatur keineswegs invers korreliert sein müssen. Bei natürlichen Enzymen ist oft eine solche inverse Korrelation zu beobachten, die häufig mit den Einflüssen der Proteinflexibilität auf Stabilität und Aktivität begründet wird. *Protein engineering* und die gerichtete *in vitro*-Evolution von Enzymen zeigten aber bereits, dass ein Kompromiss zwischen konformationeller Stabilität und Aktivität bei niedrigen Temperaturen physikalisch nicht notwendig, sondern durch die natürliche Evolution der Proteine bedingt ist (Giver *et al.*, 1998, Van den Burg *et al.*, 1998, Zhao & Arnold, 1999, Miyazaki & Arnold, 1999, Gershenson & Arnold, 2000).

2.2.2 Verbesserte Ladungsnetzwerke an der Proteinoberfläche als Ursache erhöhter Thermostabilität von *Bs*-CspB

Bs-CspB konnte also durch die *in vitro*-Evolution mittels Proside sehr stark stabilisiert werden. Die stabilste selektierte Variante T-10/10 (L2R/E3V/A46L/T64R/E66V/A67G) zeigte ein $\Delta\Delta G_D$ von 19 kJ/mol, und ihr Schmelzpunkt lag mehr als 28 Grad über dem des Wildtypproteins. Um generelle Prinzipien der Proteinstabilität ableiten zu können, war es nötig, Art und Größe der Beiträge der einzelnen Aminosäureaustausche im Detail zu analysieren. Die selektierten Reste an den sechs Positionen der Variante T-10/10 wurden daher zunächst einzeln in das Wildtypprotein eingeführt und die entsprechenden Mutanten thermodynamisch charakterisiert. Die anschließende Kombination der Mutationen, unter anderem in einem Dreifachmutationszyklus für die Positionen 3, 64 und 66, war essentiell, um energetische Wechselwirkungen insbesondere zwischen geladenen Gruppen aufzuklären. Qualitativ wurden die elektrostatischen Beiträge der Austausche über die Salzabhängigkeiten der Mutantenstabilitäten analysiert. Zur Quantifizierung dieser Effekte konnten die Differenzen zwischen den Stabilitäten bei 0 M und 2 M NaCl verwendet werden, da in Gegenwart von mehr als 1,5 M NaCl die Interaktionen zwischen Ladungen weitgehend abgeschirmt sind.

Gly67 ist für die erhöhte Stabilität der selektierten Variante unwichtig. Dies erklärt auch die hohe Variabilität an dieser Position in homologen Kälteschockproteinen.

Die L2R-Mutation wirkt überraschenderweise destabilisierend ($\Delta\Delta G_D$ = -1,5 kJ/mol). Die positive Ladung an dieser Position ist elektrostatisch günstig ($\Delta\Delta G_D$ = 2,0 kJ/mol), allerdings wird dies durch die im Vergleich zum Leucin des Wildtypproteins schlechteren hydrophoben Wechselwirkungen der Arginins überkompensiert. Der Grund für die Selektion der Arg2-Mutation bleibt unklar, insbesondere da Arg2 auch als Einzelmutation im CspB-Phagenkonstrukt zu keiner Stabilisierung gegenüber *in vitro*-Proteolyse führte.

Leu46 dagegen stabilisiert das Protein deutlich, und zwar durch verbesserte hydrophobe Wechselwirkungen ($\Delta\Delta G_D$ = 3,2 kJ/mol). In Gegenwart des selektierten Val66 ist diese Stabilisierung sogar noch größer ($\Delta\Delta G_D$ = 4,9 kJ/mol), möglicherweise weil die beiden Aminosäuren direkt interagieren. Ihre Cβ-Atome sind in der Struktur des Wildtypproteins weniger als 5 Å voneinander entfernt.

Val3 und Val66, welche Glu3 und Glu66 des Wildtypproteins ersetzen, liefern mit $\Delta\Delta G_D$ = 2,8 kJ/mol und $\Delta\Delta G_D$ = 4,7 kJ/mol ebenfalls hydrophobe Beiträge zur hohen Stabilität des selektierten Proteins. Diese Beiträge sind additiv, d.h. eine direkte Interaktion der beiden benachbarten Valin-Reste konnte so ausgeschlossen werden. Dies war wichtig für die Analyse der elektrostatischen Veränderungen, da sich Wechselwirkungsenergien aus Mehrfachmutationszyklen nur dann verlässlich interpretieren lassen, wenn die eingeführten Aminosäuren keine zusätzlichen direkten oder indirekten Interaktionen eingehen (Horovitz & Fersht, 1992). Die elektrostatischen Beiträge der beiden Glu→Val-Mutationen ($\Delta\Delta G_D$ = 4,6 bzw. 2,5 kJ/mol) bestätigten die Relevanz von Glu3 und Glu66 für die niedrige Thermostabilität von *Bs*-CspB. Kombiniert führten die Austausche gegen Valin zu einer Coulomb´schen Stabilisierung um 5,6 kJ/mol. Wie der Doppelmutationszyklus an beiden Positionen ergab, resultierten allerdings nur 1,5 kJ/mol davon aus der Aufhebung einer direkten repulsiven Wechselwirkung zwischen den beiden Glutamaten (Abb. 3 in Teilarbeit C). Für die destabilisierende Wirkung von Glu3 und Glu66 in *Bs*-CspB ist demnach weniger die paarweise Abstoßung als vielmehr ein globaler Effekt der hohen negativen Ladungsdichte in dieser Region der Proteinoberfläche verantwortlich.

Im Gegensatz zu diesen Daten wurde die Energie der direkten abstoßenden Wechselwirkung zwischen Glu3 und Glu66 in der R3E/L66E-Doppelmutante von *Bc*-Csp zu 4,0 kJ/mol bestimmt (Perl & Schmid, 2001). Eine mögliche Ursache für die unterschiedlichen Ergebnisse ist, dass im *Bc*-Csp die elektrostatischen Effekte des Glu3 relativ zum ausgetauschten positiv geladenen Arginin und nicht zu einer neutralen Aminosäure an dieser Position gemessen wurden. Der Unterschied der Ergebnisse könnte allerdings auch durch Ala67 bedingt sein, das im *Bc*-Csp fehlt, im *Bs*-CspB aber möglicherweise über die terminale Carboxylgruppe mit Glu3 interagiert.

Auch die Ladung des selektierten Arg64 hat, ähnlich wie die von Glu3 und Glu66, einen eher globalen Effekt auf die Stabilität von *Bs*-CspB. Die Coulomb´schen Beiträge der T64R-Mutation ($\Delta\Delta G_D$ = 2,5 kJ/mol) sind nur wenig von der Ladungsumgebung, d.h. vom Vorhandensein anderer Mutationen abhängig. Der kombinierte Austausch der Reste an den Positionen 3, 64 und 66 in einem Dreifachmutationszyklus (Abb. 3 in Teilarbeit C) ergab schwache elektrostatische Interaktionen zwischen Arg64 und Glu3 (1,3 kJ/mol) sowie zwischen Arg64 und Glu66 (1,2 kJ/mol). Beide Energien liegen also in der Größenordnung der direkten abstoßenden Wechselwirkung zwischen Glu3 und Glu66 (1,5 kJ/mol). Wenn Arg64 im Protein vorhanden war, konnte zwischen Glu3 und Glu66 keine Abstoßung mehr detektiert werden, und umgekehrt, wenn eines der beiden Glutamate fehlte, zeigte Arg64 keine anziehende Wechselwirkung mehr mit dem verbleibenden Glutamat. Arg64 hebt also mit seiner positiven Ladung die schwache Abstoßung zwischen Glu3 und Glu66 auf, ohne dabei selbst mit einem der beiden Reste direkt zu interagieren. Eventuell kommt es durch die Einführung des Argininrests zu alternativen Orientierungen der räumlich benachbarten Glu3 und Glu66. Dies wäre möglich, da die oberflächenexponierten Reste sowohl bei *Bs*-CspB als auch bei *Bc*-Csp sehr flexibel sind (Müller *et al.*, 2000).

Wie mit ortsgerichteter Mutagenese gezeigt werden konnte, sorgt in *Bs*-CspB bereits der Einzelaustausch E3R für insgesamt günstige elektrostatische Wechselwirkungen an der Proteinoberfläche (Perl & Schmid, 2001). Die Eliminierung der negativen Ladung und gleichzeitige Einführung einer positiven Ladung an Position 3 stabilisiert *Bs*-CspB um 11,1 kJ/mol, und die Salzabhängigkeit der Stabilität der E3R-Mutante ist nahezu identisch mit der von *Bc*-Csp. In der stabilsten selektierten Variante T-10/10 werden die gleichen Ladungsänderungen durch zwei Mutationen erreicht, E3V und T64R. Ihr gemeinsamer Stabilisierungsbeitrag entspricht mit 10,7 kJ/mol ungefähr dem der E3R-Mutation. Allerdings sind diese beiden Mutationen noch nicht ausreichend, um insgesamt günstige elektrostatische Interaktionen zu bewirken, die Salzabhängigkeit ähnelt lediglich der der E3V/E66V-Doppelmutante von *Bs*-CspB (Abb. 7A). Für eine Salzabhängigkeit entsprechend der von *Bc*-Csp sind in der selektierten Variante der Austausch beider negativ geladener Reste an 3 und 66 und die zusätzliche Einführung einer positiven Ladung nötig. Sowohl Arg 64 als auch Arg2 bewirken in Kombination mit Val3 und Val66 diese günstigen, stabilisierenden ionischen Wechselwirkungen, trotz unterschiedlicher Position und Orientierung ihrer Seitenkette (Abb. 7A). Es ist also festzustellen, dass die Mutationen zu Arginin an den Positionen 3, wie in *Bc*-Csp, an 2 oder an 64 zwar unterschiedlich große, qualitativ aber identische Coulomb´sche Effekte haben.

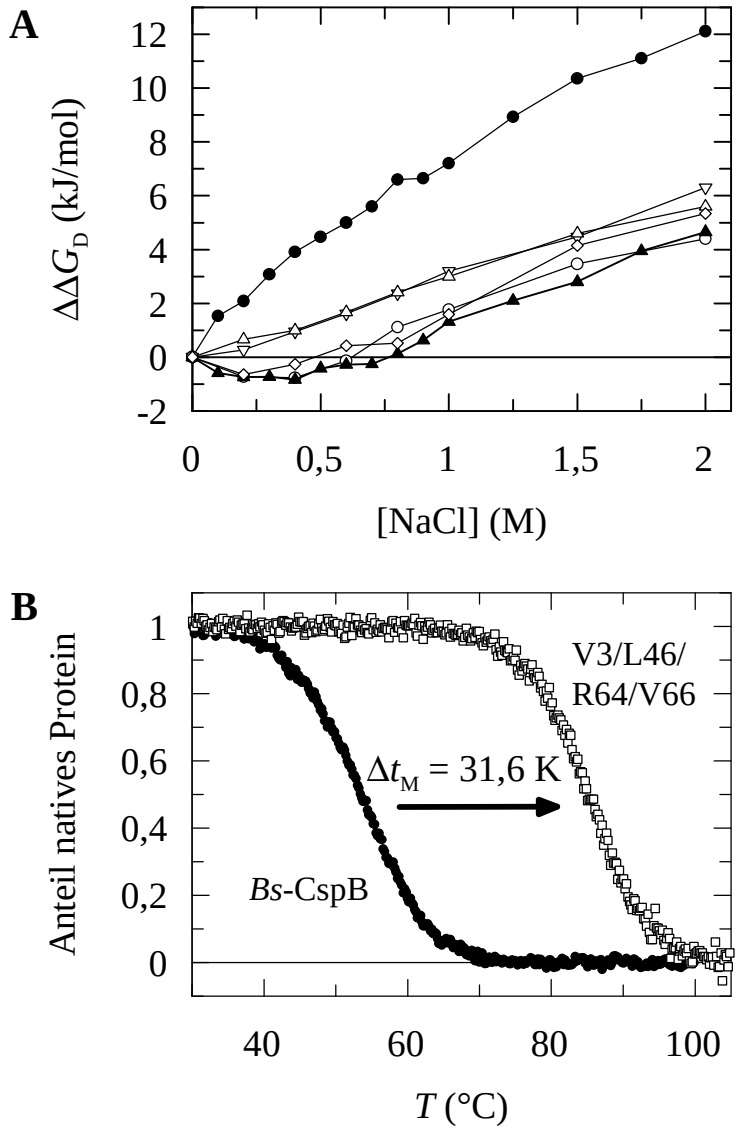


Abbildung 7. (A) Änderung der freien Enthalpie der Entfaltung bei 70 °C als Funktion der NaCl-Konzentration relativ zur freie Enthalpie in Abwesenheit von NaCl ($\Delta\Delta G_D$) für *Bs*-CspB (●), die Mutanten E3V/E66V (▽), E3V/T64R (△), L2R/E3V/E66V (◇) und E3V/T64R/E66V (○) sowie *Bc*-Csp (□). Die Messung der thermischen Übergänge erfolgte wie in Teilarbeit C beschrieben. **(B)** Thermische Entfaltungsübergänge von *Bs*-CspB und der hyperstabilen Vierfachmutante E3V/A46L/T64R/E66V in 100 mM Na-Cacodylat, pH 7,0. Gezeigt ist der Anteil des nativen Proteins als Funktion der Temperatur nach Analyse der Übergänge entsprechend dem Zweizustandsmodell.

Basierend auf dieser detaillierten Analyse war es möglich, ein hyperstabiles Kälteschockprotein mit nur vier Austauschen relativ zu *Bs*-CspB-Wildtyp zu konstruieren. Die Vierfachmutante mit Val3, Leu46, Arg64 und Val66 zeigt eine Stabilisierung um 21,4 kJ, ihr Schmelzpunkt liegt bei t_M = 85,6 °C, d.h. 31,6 Grad höher als der des Wildtypproteins (Abb. 7B). Sie ist damit sogar deutlich stabiler als das Kälteschockprotein *Tm*-Csp aus dem hyperthermophilen Organismus *Thermotoga maritima* (Wassenberg *et al.*, 1999).

Durch die *in vitro*-Evolution mittels Proside konnte ein Stabilisierungspotential auch für Positionen im Protein identifiziert werden, an denen die zwischen *Bs*-CspB und *Bc*-Csp existierenden Sequenzunterschiede für die Thermostabilität unwichtig waren (Perl & Schmid, 2001). Ein Vergleich der in der Selektion gefundenen alternativen Aminosäurekombinationen mit den Sequenzen der thermostabilen Homologen von *Bs*-CspB erlaubt eine Diskriminierung zwischen global wirkenden und lokal beschränkten Effekten, so dass eine Ableitung grundsätzlicher Stabilitätsprinzipien möglich ist. Die stabilste selektierte Variante T-10/10 unterscheidet sich an allen sechs randomisierten Positionen vom thermostabilen *Bc*-Csp, was zunächst auf verschiedene Prinzipien zur Erhöhung der Thermostabilität hinzudeuten scheint. Doch beide Proteine nutzen die gleichen Strategien für eine hohe Stabilität. Sowohl bei der natürlichen Evolution als auch bei der gerichteten Evolution *in vitro* erfolgte eine Verbesserung der elektrostatischen Wechselwirkungen durch Optimierung der Oberflächenladungsverteilung (s. Abb. 6). Die negativen Ladungen an den Positionen 3 und 66 wurden eliminiert und eine positive Ladung in diese Region der Proteinoberfläche eingeführt. Vergleichbare Sequenz- und Oberflächeneigenschaften sind auch bei den homologen Proteinen der hyperthermophilen Organismen *Aquifex aeolicus* und *Thermotoga maritima* zu erkennen (Abb. 1b in Teilarbeit C).

Dass oberflächenexponierte Ladungen für die Stabilität, speziell thermostabiler und hyperthermostabiler Proteine eine wichtige Rolle spielen können, ist mittlerweile unumstritten (Akke & Forsen, 1990, Grimsley *et al.*, 1999, de Bakker *et al.*, 1999, Xiao & Honig, 1999, Sanchez-Ruiz & Makhatadze, 2001, Karshikoff & Ladenstein, 2001). Weitgehend unklar ist allerdings, welche Bedeutung paarweise ionische Wechselwirkungen dabei haben. Für Ionenpaare sind die Desolvatationsenthalpien und der Verlust von Seitenkettenentropie durch Immobilisierung der interagierenden Gruppen besonders groß (Strop & Mayo, 2000, Xiao & Honig, 1999). Wesentlich günstiger scheinen wechselwirkende Ionenpaare bzw. größere Ladungsnetzwerke zu sein, wie sie die Strukturuntersuchungen verschiedener thermostabiler Proteine nahe legen (Hatanaka *et al.*, 1997, Korndörfer *et al.*, 1995). Allerdings hat es sich als sehr schwierig erwiesen, die energetischen Beiträge dieser Netzwerke durch Mutagenese zu analysieren oder sie in mesostabile Homologe zu übertragen (Lebbink *et al.*, 1998, Lebbink *et al.*, 1999). Im Gegensatz zu Ionenpaaren profitieren Ladungsnetzwerke von der hohen Flexibilität geladener Reste, da so alternative Interaktionen und eine schnelle Reorientierung als Antwort auf lokale Fluktuationen möglich sind.

Auch für die Thermostabilität der Kälteschockproteine scheinen hohe Flexibilitäten der solvensexponierten Seitenketten und die langreichweitigen Coulomb´schen Wechselwirkungen entscheidend wichtig zu sein. Ein Indiz dafür ist, dass die Einführung einer positiven Ladung *Bs*-CspB relativ unabhängig davon stabilisiert, wo genau die Ladung auf der Oberfläche lokalisiert ist. Auf diese Flexibilität bei der Positionierung geladener Gruppen deuten auch die hochaufgelösten Kristallstrukturen des Wildtypproteins und fünf weiterer Varianten von *Bc*-Csp hin (Müller *et al.*, 2000, Delbrück *et al.*, 2001). Nicht nur bei den verschiedenen Varianten, sondern auch in den beiden Monomeren der jeweiligen Einheitszellen zeigt die Seitenkette von Arg3 sehr unterschiedliche Orientierungen (Abb. 8).

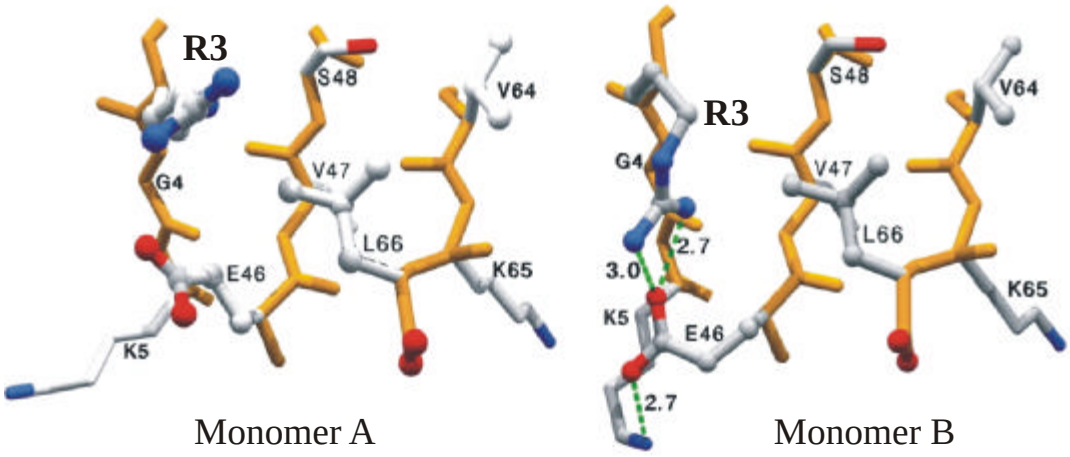


Abbildung 8. Alternative Orientierungen der Seitenketten von Arg3 und der umgebenden Aminosäuren in der Kristallstruktur von *Bc*-Csp. Dargestellt sind die Hauptkonformere der Reste in den Monomeren A und B der asymmetrischen Einheit. Potentielle Salzbrücken sowie Wasserstoffbrückenbindungen sind als grüne Linien wiedergegeben, die entsprechenden Abstände zwischen den Heteroatomen sind in Å angegeben. Die Abbildung wurde von Delbrück *et al.* (2001) übernommen und modifiziert.

Für die Existenz von Ladungsnetzwerken an der Oberfläche der Kälteschockproteine sprechen auch die in den Mehrfachmutationszyklen bestimmten Wechselwirkungen zwischen geladenen Aminosäuren. Die direkte paarweise Abstoßung zwischen Glu3 und Glu66 in *Bs*-CspB ist relativ gering im Vergleich zu den global destabilisierenden ionischen Beiträgen dieser beiden Reste. Außerdem zeigt Arg64 einen stabilisierenden Effekt, der von der Ladungsumgebung, d.h. der Präsenz von Glu3 und Glu66 abhängt, jedoch keiner direkten Interaktion mit einem der beiden Glutamate zugeordnet werden kann. Ein ähnliches Verhalten wurde auch für Arg3 in *Bs*-CspB bzw. *Bc*-Csp beobachtet (Perl & Schmid, 2001).

Sowohl bei den mit Proside selektierten stabilisierten Varianten (Tabelle 2 in Teilarbeit B) als auch bei den thermostabilen und hyperthermostabilen Homologen von *Bs*-CspB (Abb. 1 in Teilarbeit C) zeigt sich an den Positionen 2 bzw. 3 eine deutliche Präferenz für Arg gegenüber Lys. Das allgemein bei thermostabilen Proteinen erkennbare höhere Arg/Lys-Verhältnis (Merkler *et al.*, 1981, Mrabet *et al.*, 1992, Narinx *et al.*, 1997) kann ebenfalls mit der Existenz von Ladungsnetzwerken und der Flexibilität beteiligter geladener Seitenketten erklärt werden. Arginin ist aufgrund der Ladungsresonanz der Guanidinium-gruppe in der Lage, mehr als ein Ionenpaar auszubilden.

Durch die *in vitro*-Evolution mittels Proside war es möglich, die Unterschiede in den Oberflächenladungsverteilungen zwischen *Bs*-CspB und seinen thermostabilen Homologen als Ursache für deren Stabilitätsdifferenzen zu identifizieren. Die starke Stabilisierung von *Bs*-CspB, wie sie durch die Verbesserung der Ladungsnetzwerke erreicht werden konnte, verdeutlicht das große Stabilisierungspotential oberflächenexponierter Reste und die Bedeutung elektrostatischer Wechselwirkungen für die Thermostabilität von Proteinen allgemein. Die gleichzeitige Sättigungsmutagenese an sechs benachbarten Positionen, welche anhand von Mutagenese- und Strukturdaten ausgewählt worden waren, erlaubte die Suche nach optimalen Aminosäurekombinationen in einem riesigen Sequenzraum und offenbarte die gravierenden Vorteile der *in vitro*-Evolution gegenüber der ortsgerichteten Mutagenese einiger weniger Reste.

2.3 Thermodynamische Charakterisierung und Analyse des Faltungsmechanismus des Gen-3-Proteins filamentöser Phagen

2.3.1 Die Domänen N1 und N2 von G3P: separate Faltungseinheiten mit sehr unterschiedlichen Stabilitäten

Wie schon erwähnt, besteht das Gen-3-Protein filamentöser Phagen wie fd oder M13 aus drei Domänen, N1, N2 und CT, die durch glycinreiche Linker miteinander verknüpft sind. CT ist teilweise in die Phagenhülle eingebettet (Boeke & Model, 1982, Stengele *et al.*, 1990) und sehr hydrophob; es existieren daher keine Strukturinformationen zu dieser Domäne. Die beiden aminoterminalen Domänen N1 und N2, die bei der Infektion für die Wechselwirkungen mit den Rezeptoren der *E. coli*-Zelle verantwortlich sind, ragen dagegen von der Phagenoberfläche weg und sind sehr gut löslich. Sie bilden zusammen eine hufeisenförmige Struktur, in der die N1-Domäne teilweise von der größeren N2-Domäne umfasst wird (Lubkowski *et al.*, 1998, Lubkowski *et al.*, 1999, Holliger *et al.*, 1999). Die Interaktionsfläche zwischen beiden beträgt ungefähr 1.000 $Å^2$ (Lubkowski *et al.*, 1999). Domäne N2 besteht aus einem globulären Teil, der strukturell der Domäne N1 ähnelt, sowie einer gestreckten Gelenksubdomäne, die den größten Teil der Wechselwirkungen mit N1 vermittelt. Basierend auf Vergleichen der Kristallstrukturen von G3P aus fd- und M13-Phagen wird dieser Subdomäne eine Scharnierfunktion bei der für die Infektion essentiellen Dissoziation der Domänen zugeschrieben (Holliger *et al.*, 1999).

Um Erkenntnisse zur konformationellen Stabilität des N1N2-Fragments von G3P (Aminosäuren 1-217, bezeichnet als G3P*) sowie zu den Stabilisierungseffekten der mit Proside gefundenen Mutationen zu erhalten, wurden das G3P*-Wildtypprotein und die vier selektierten Varianten T101I, T101I/D209Y, T13I/T101I/D209Y und T13I/T101I/Q129H/D209Y gereinigt und thermodynamisch charakterisiert (Teilarbeit D). Die Gleichgewichtsstabilitäten der Proteine wurden sowohl durch thermische als auch durch GdmCl-induzierte Entfaltungsübergänge analysiert.

Als Sonden für die thermische Denaturierung dienten die CD-Signale bei 210 und 230 nm. Die Übergänge bei 210 nm waren einphasig und reflektierten selektiv die Entfaltung der labileren N2-Domäne. Dagegen waren bei 230 nm zweiphasige Verläufe zu beobachten, deren erster Übergang wiederum der Entfaltung von N2 zugeordnet werden konnte, während der zweite Übergang die Entfaltung der stabileren N1-Domäne widerspiegelte. Dieser zweite Übergang verläuft deutlich weniger kooperativ als der erste, da Domäne N1 nur etwa halb so groß ist wie Domäne N2. Zusätzlich zu den Volllängenproteinen wurde auch die isolierte N1-Domäne (Aminosäuren 1-67 von G3P) mit der Mutation T13I gereinigt und charakterisiert. Ihr thermischer Übergang bei 230 nm war einphasig und identisch mit dem zweiten Übergang der entsprechenden G3P*-Varianten mit T13I-Mutation. Erfolglos blieb leider der Versuch, den globulären Teil von N2 (Aminosäuren 124 bis 202 von G3P) ohne die Gelenksubdomäne zu exprimieren.

Die Zuordnung der CD-Signaländerungen zu den Entfaltungen der Domänen N1 und N2 konnte durch kalorimetrische Messungen bestätigt werden. Sie zeigen, dass die thermisch induzierte Gleichgewichtsentfaltung von G3P* sequentiell verläuft (Abb. 9). Im ersten Schritt entfaltet Domäne N2 (t_M = 48 °C), gleichzeitig mit der Dissoziation beider Domänen, und im zweiten Schritt erfolgt die Entfaltung von Domäne N1 (t_M = 59,8 °C). Die Kopplung der Domänendissoziation mit der Entfaltung der labileren N2-Domäne ist ein Indiz für den hohen Beitrag der Interdomänenwechselwirkungen zur Stabilität dieser Domäne.

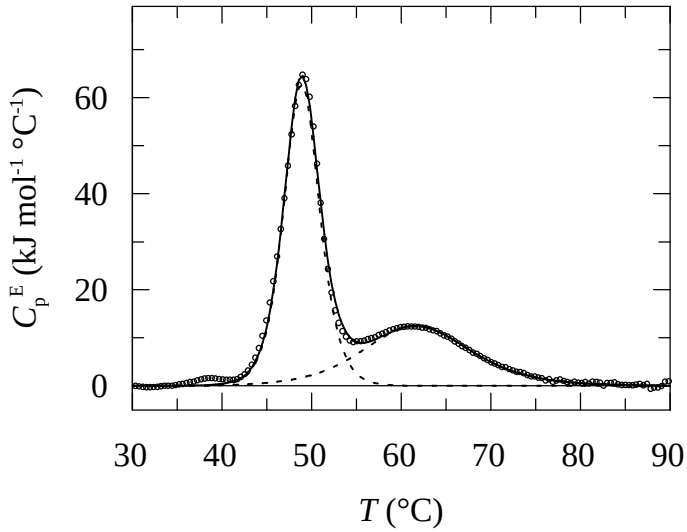


Abbildung 9. Wärmekapazität C_p^E (T) der Entfaltung des G3P*-Wildtypproteins. Die Symbole zeigen die experimentellen Daten, die durchgezogene Linie gibt die Ergebnisse der Levenberg-Marquardt-Analyse nach dem Nicht-Zweizustandsmodell wieder, und die gestrichelten Linien repräsentieren die Dekonvolution in zwei unabhängige Entfaltungsübergänge für die Domänen N2 und N1. Die kalorimetrischen Messungen erfolgten wie in Teilarbeit D beschrieben.

Die GdmCl-induzierten Entfaltungsübergänge von G3P* und seinen Varianten wurden über Änderungen der Proteinfluoreszenz verfolgt. G3P* ist reich an Tyr- und Trp-Resten, die ungleichmäßig über das Protein verteilt sind (s. Abb. 1b in Teilarbeit D) und sich dadurch hervorragend als selektive Sonden zur Detektion der Entfaltungsreaktionen der individuellen Domänen eignen. Domäne N1 enthält drei Tryptophane und drei Tyrosine, wohingegen Domäne N2 elf Tyr-Reste und nur ein einzelnes, stark lösungsmittel-exponiertes Tryptophan aufweist. Von den elf Tyrosinen in N2 befinden sich sieben im globulären Teil und vier in der Gelenksubdomäne. Wie die Analyse der Spektren bei verschiedenen GdmCl-Konzentrationen ergab, kann die Denaturierung von Domäne N1 separat über die Trp-Fluoreszenz bei 360 nm nach Anregung bei 295 nm detektiert werden, die Tyr-Fluoreszenz bei 310 nm nach Anregung bei 280 nm ist dagegen spezifisch für die Entfaltung von Domäne N2.

Die anhand der thermischen und GdmCl-induzierten Entfaltungsübergänge bestimmten thermodynamischen Stabilitäten der vier selektierten G3P*-Mutanten sind in Tabelle 1 und 2 in Teilarbeit D aufgelistet. Die Beiträge der Einzelmutationen korrelierten sehr gut mit den beobachteten proteolytischen Stabilitäten der entsprechenden Phagenvarianten, so dass die erhöhten Proteaseresistenzen in der Tat auf gesteigerte konformationelle Stabilitäten von G3P zurückgeführt werden konnten. Damit wurde das dem Proside-System zugrunde liegende Selektionsprinzip erneut bestätigt.

Die erste stabilisierende Mutation, T101I in Domäne N2, erhöhte den Schmelzpunkt dieser Domäne gegenüber dem Wildtypprotein um mehr als 8 Grad und ΔG_D um 16,5 kJ/mol. In ihrer Stabilität kam Domäne N2 damit der N1-Domäne sehr nahe. Die im Vergleich zum Wildtypprotein höhere Kooperativität des Übergangs von N2 in G3P*(T101I) deutete ebenfalls darauf hin, dass die Entfaltung der N2-Domäne bereits gekoppelt mit der Entfaltung von N1 verläuft, die beiden Domänen also aufgrund der sehr ähnlichen Stabilitäten nicht mehr sequentiell und unabhängig voneinander entfalteten.

Möglicherweise ist diese Kopplung der Grund dafür, dass die zweite selektierte Mutation in Domäne N2, D209Y, keinen messbaren Effekt auf die thermische Stabilität hatte. Generell gilt, dass die Analyse der Entfaltung eines Zweidomänenproteins basierend auf einem sequentiellen Zweischrittmechanismus nur verlässlich ist, solange die beiden Domänen deutlich voneinander separierte Übergänge zeigen (Tsalkova & Privalov, 1985, Brandts *et al.,* 1989). Anders als bei den thermischen Entfaltungen war bei der GdmCl-induzierten Entfaltung ein klarer Stabilisierungsbeitrag der D209Y-Mutation zu beobachten (Tab. 2 in Teilarbeit D). Eventuell wirkt GdmCl durch Salzeffekte unterschiedlich stark denaturierend auf N1 und N2, so dass bei den Varianten G3P*(T101I) bzw. G3P*(T101I/D209Y) die GdmCl-induzierten Übergänge der beiden Domänen immer noch weit genug getrennt sind, um eine Kopplung der Entfaltungsreaktionen zu vermeiden. Ein Vergleich der Stabilitäts-differenzen zwischen N1 und N2 aus den thermischen und den GdmCl-induzierten Entfaltungen würde darüber Aufschluss geben. Allerdings sind die Extrapolationen von ΔG_D auf Standardbedingungen fehlerbehaftet, und daher ist es schwierig, verlässliche Aussagen zu treffen.

Die dritte selektierte Mutation, T13I in Domäne N1, steigerte den Schmelzpunkt von N1 um 6 Grad auf mehr als 66 °C und stellte so den deutlichen Stabilitätsunterschied zwischen den beiden Domänen wieder her. Da damit die N2-Domäne wieder limitierend für die gesamte konformationelle und proteolytische Stabilität von G3P* im Phagen war, wurde entsprechend bei der weiteren *in vitro*-Evolution erneut eine Mutation in N2 selektiert. Q129H erhöhte, gemeinsam mit den Mutationen T101I und D209Y, den Schmelzpunkt von N2 im isolierten Protein auf 62 °C, der damit etwa 14 Grad über dem des G3P*-Wildtyp-proteins lag. Der Mittelpunkt des GdmCl-induzierten Entfaltungsübergangs wurde auf 2,5 M GdmCl angehoben. Die Einführung des Histidins an Position 129 führte zu einer starken Abnahme der Tyr-Fluoreszenz des nativen Proteins. Vermutlich löscht His129 die Fluoreszenz der im nativen Zustand direkt benachbarten Tyr151, Tyr166 und Tyr168. Mit der räumlichen Trennung während der Denaturierung fällt diese Löschung weg und die Fluoreszenz steigt entsprechend an. Dieser spezifische Effekt der Q129H-Mutation bestätigte die Annahme, dass die Tyr-Fluoreszenz selektiv den Faltungszustand der Domäne N2 detektiert.

Die stabilisierenden Effekte der selektierten Mutationen waren immer auf die entsprechenden Domänen beschränkt. So wirkte die T13I-Mutation nur auf Domäne N1 und ließ die Übergänge von Domäne N2 unverändert. Umgekehrt hatten die Mutationen in Domäne N2 keinen Einfluss auf die Stabilität von N1 (s. Abb. 9c in Teilarbeit D). Darüber hinaus zeigt die Domäne N1(T13I) identische Entfaltungsübergänge als Teil von G3P* und als isolierte Domäne. Dies verdeutlicht, dass die Stabilität von N1 unabhängig von der Präsenz der labileren N2-Domäne ist.

Dass Domäne N2 in G3P* in ihrer Stabilität zu einem großen Teil durch die Wechsel-wirkungen mit N1 bestimmt wird, spiegelt sich auch darin wider, dass zwei der drei Austausche, T101I und D209Y, sich in der Gelenksubdomäne befinden, welche die meisten der stabilisierenden Interaktionen mit Domäne N1 vermittelt.

Domäne N2 zeigte bei den thermischen und den GdmCl-induzierten Entfaltungübergängen deutlich höhere Kooperativitäten (ΔH_{vH} bzw. m) als Domäne N1. Dies ist durch die Kopplung von Entfaltung der N2-Domäne und Trennung der Interdomänenkontakte zu N1 bedingt. Dadurch bilden der globuläre Teil und die Gelenksubdomäne von N2 eine kooperative Faltungseinheit, die nahezu doppelt so groß ist wie Domäne N1. Ob diese

kooperative Einheit allein durch die thermodynamische Kopplung zwischen Entfaltung und Domänendissoziation bedingt ist, oder ob die komplette N2-Domäne unter günstigen Bedingungen auch in Abwesenheit der Kontakte zu Domäne N1 falten kann, lässt sich mit Gleichgewichtsexperimenten nicht beantworten. Aufschluss darüber gibt allerdings die Analyse von kinetischen Faltungsintermediaten des G3P* (Teilarbeit E).

Die Stabilitätsbeiträge der selektierten Mutationen können relativ gut anhand der Struktur des nativen G3P* erklärt werden. Für die Mutation T101I in der Gelenksubdomäne von N2 ist eine Verbesserung der Wechselwirkungen mit Domäne N1 sehr wahrscheinlich. Wie die Kristallstruktur des Wildtypproteins zeigt, ist Thr101 lediglich 4 Å von Val43 und Val45 in N1 entfernt (Lubkowski *et al.*, 1998). Ile101 in der stabilisierten Variante bildet vermutlich günstige hydrophobe Interdomänenwechselwirkungen mit Val43 und Val45 aus (Abb. 10a in Teilarbeit D).

Für die Q129H-Mutation kann ebenfalls von verbesserten bzw. zusätzlichen hydrophoben Kontakten ausgegangen werden (Abb. 10c in Teilarbeit D). Im Wildtypprotein ragt Gln129 in eine Tasche, die unter anderem von Tyr151, Tyr166 und Tyr168 gebildet wird. His129 in der Variante löscht die Fluoreszenz dieser Tyrosinreste, was darauf hindeutet, dass das Histidin tatsächlich mit diesen Aminosäureseitenketten wechselwirkt.

Auch Thr13 in Domäne N1 befindet sich in einer hydrophoben Tasche, die von drei aromatischen Aminosäuren, Phe17, Tyr33 und Trp38, ausgekleidet wird. Auch bei der Mutation des Threonin zu Isoleucin kommt es vermutlich zur Optimierung hydrophober Kontakte.

Wie die D209Y-Mutation in der Gelenksubdomäne stabilisierend wirkt, lässt sich auf der Grundlage der Kristallstruktur nicht einfach beantworten. Asp209 ist relativ stark lösungsmittelexponiert. Eventuell wirkt der Austausch gegen Tyr stabilisierend durch eine Verringerung negativer Ladungsdichte in dieser Oberflächenregion von G3P*.

Insgesamt ist festzustellen, dass mindestens drei der vier selektierten Mutationen in G3P* über die Verbesserung hydrophober Wechselwirkungen stabilisierend wirken. Anders als bei der *in vitro*-Evolution des Kälteschockproteins *Bs*-CspB spielte hier die Optimierung der elektrostatischen Interaktionen an der Proteinoberfläche keine entscheidende Rolle. Eventuell ist das entsprechende Stabilisierungspotential bei G3P* aufgrund einer bereits relativ gleichmäßigen Ladungsverteilung geringer. Wie Berechnungen der Coulomb´schen Wechselwirkungen zeigen, weist die labilere N2-Domäne in ihrem globulären Teil keine Bereiche mit hoher negativer oder positiver Oberflächenladungsdichte auf (Abb. 10). Etwas gehäufter treten negative Ladungen an der Oberfläche von Domäne N1 auf. Ein großer Teil davon ist allerdings entweder in die Interaktionen mit Domäne N2 oder dem Rezeptorprotein TolA von *E. coli* involviert (Riechmann & Holliger, 1997, Lubkowski *et al.*, 1998), so dass ein Austausch ohne Beeinträchtigung der *in vivo*-Funktion von G3P nicht möglich ist.

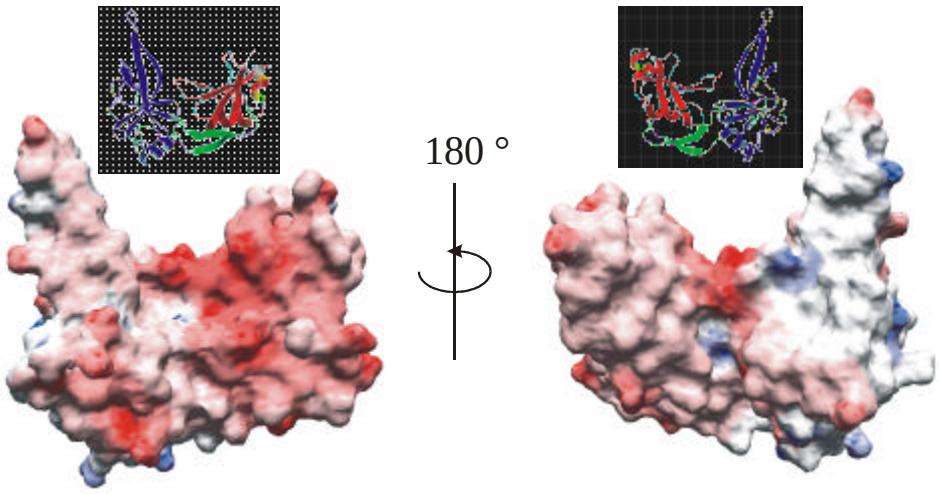


Abbildung 10. Oberflächenladungen des G3P*-Wildtypproteins. Die Oberflächen wurden mit *Swiss-pdb viewer* (Guex & Peitsch, 1997) generiert unter Berücksichtigung der geladenen Aminosäuren Asp, Glu, Arg und Lys sowie der N- und C-Termini. Potentiale von -10 kT/e oder weniger sind dabei rot, neutrale Potentiale (0 kT/e) weiß und Potentiale von + 10 kT/e oder mehr blau gefärbt. Um die Lage der beiden Domänen N1 und N2 im Molekül zu verdeutlichen, sind Tertiärstrukturabbildungen von G3P* gezeigt, in denen die Domäne N1 rot, der globuläre Teil von N2 blau und die Gelenksubdomäne von N2 grün gefärbt sind. Diese Abbildungen wurden mit dem Programm MolMol (Koradi *et al.*, 1996) erstellt. Alle Darstellungen basieren auf den Strukturdaten von Holliger *et al.* (1999).

Stabilitäts-Strukturvergleiche wie hier werden generell verwendet, um die Auswirkungen von Mutationen auf molekularer Ebene zu verstehen. Sie können natürlich nicht beweisen, dass die vorgeschlagenen Interaktionen tatsächlich existieren. Hierzu wären zusätzliche Kristallstrukturen der Varianten, sowie Doppel- und Dreifachmutationszyklen an den Wechselwirkungsstellen nötig. Darüber hinaus können Auswirkungen von Mutationen auf den entfalteten Zustand als Ursache von Änderungen in ΔG_D nie vollständig ausgeschlossen werden (Pace *et al.*, 2000).

Die mit Proside selektierten Mutationen reflektieren die Prinzipien, die der Stabilität von Zweidomänenproteinen allgemein zugrunde liegen. Neben den Interaktionen in den individuellen Domänen tragen die Interdomänenwechselwirkungen ganz entscheidend zur Stabilität dieser Proteine bei. Häufig sind die Wechselwirkungen zwischen den Domänen sehr stark und, wie bei Domäne N2 zu beobachten, essentiell für den nativen Zustand der labileren Domäne, so dass die Domänendissoziation mit der Entfaltung gekoppelt ist (Privalov, 1982, Wenk & Jaenicke, 1999). So führen verbesserte Domäneninteraktionen zur Stabilisierung von Zweidomänenproteinen, was mit den beiden Mutationen T101I und D209Y in der Gelenksubdomäne von N2 des G3P* gezeigt werden konnte. Bei einer Kopplung von Domänendissoziation und Entfaltung stellt die freie Enthalpie der Entfaltung ΔG_D die Summe der Enthalpien beider Prozesse dar. In diesem Fall ist eine Stabilisierung des Gesamtproteins nicht nur durch verbesserte Wechselwirkungen zwischen den Domänen, sondern auch durch die intrinsische Stabilisierung der labileren Domäne selbst möglich. Die Q129H-Mutation im globulären Teil von N2 bestätigt dies.

2.3.2 Die Domänenassoziation als letzter Schritt der Rückfaltung von G3P* führt zur kinetischen Kopplung von N1 und N2

Eindomänenproteine zeichnen sich häufig durch sehr schnelle Faltungsreaktionen nach dem einfachen Zweizustandsmodell aus. Die meisten Proteine sind allerdings aus mehreren Domänen aufgebaut, welche strukturelle, funktionale oder separat faltende Einheiten bilden (Garel, 1992; Jaenicke, 1999; Netzer & Hartl, 1997). Die Faltungsmechanismen von Mehrdomänenproteinen gestalten sich daher in der Regel wesentlich komplexer als die der Eindomänenproteine. Zusätzlich zur Faltung der individuellen Domänen stellt deren Assoziation einen entscheidenden Schritt auf dem Weg zum nativen Zustand dar. Wenn die Einzeldomänen für sich stabil sind, können sie die Interdomänenwechselwirkungen nach ihrer Faltung ausbilden. Wenn nicht, muss die Faltung der Domänen direkt mit ihrer Assoziation gekoppelt sein (Rudolph *et al.*, 1990; Tsunenaga *et al.*, 1987). Wie aus den Gleichgewichtsanalysen bekannt, geht bei G3P* die Domänendissoziation mit der Entfaltung der labileren Domäne N2 einher. An G3P* lässt sich daher sehr gut untersuchen, ob und in welcher Weise Domänenfaltung und Assoziation bzw. Dissoziation und Entfaltung kinetisch gekoppelt sind.

Darüberhinaus ist die Aufklärung des G3P*-Faltungsmechanismus auch wichtig, um den Vorgang der Infektion von *E. coli* durch filamentöse Phagen besser zu verstehen. Die Bindung des G3P an den F-Pilus von *E. coli* induziert die Dissoziation von N1 und N2. Erst durch diese Trennung wird die Bindungsstelle für den Corezeptor TolA auf Domäne N1 exponiert. Sie ist im assoziierten Zustand maskiert, weil ein Teil der für die Bindung von TolA verantwortlichen Aminosäuren durch Wechselwirkungen mit Domäne N2 abgeschirmt ist (Lubkowski *et al.*, 1999).

Die Faltungsreaktionen der individuellen Domänen in G3P* konnten selektiv gemessen werden, da die ungleichmäßig im Protein verteilten Trp- und Tyr-Reste als spezifische Sonden für die Faltungszustände von N1 und N2 genutzt werden konnten (Teilarbeit D). Zur Analyse des Faltungsmechanismus wurde die stabilisierte Vierfachmutante G3P*(T13I/T101I/Q129H/D209Y) (bezeichnet als IIHY-G3P*) verwendet, da aufgrund ihrer höheren Stabilität teilgefaltete Intermediate besser charakterisierbar waren. Wie ein Vergleich ausgewählter Faltungsreaktionen dieser Variante mit denen des G3P*-Wildtypproteins ergab, hatten die stabilisierenden Mutationen keinen Einfluss auf den Mechanismus der Faltung. Die mit IIHY-G3P* gewonnenen Erkenntnisse sind also auf G3P* übertragbar.

Die Rückfaltung von IIHY-G3P* ausgehend vom vollständig denaturierten Protein zeigt stark unterschiedliche Raten für die Domänen N1 und N2. Während die labilere N2-Domäne in zwei Phasen mit Zeitkonstanten von 7 s und 42 s (bei 0,5 M GdmCl) faltet, ist die Kinetik der Faltung von Domäne N1 monoexponentiell und mit 9,4 ms etwa 500-mal schneller (Abb. 11B). Überraschenderweise konnten entsprechende Unterschiede bei der Entfaltung nicht beobachten werden. Für beide Domänen wurden identische, langsame Entfaltungskinetiken mit einer Zeitkonstante von 71 s (bei 4,5 M GdmCl) detektiert, was auf einen gemeinsamen, ratenlimitierenden Schritt hindeutet (Abb. 11A). Die Entfaltung der isolierten Domäne N1(T13I) ist hingegen sehr schnell (t = 28 ms). Die langsame Entfaltung der N1-Domäne ausgehend von nativem IIHY-G3P* hängt also mit der Gegenwart der gefalteten Domäne N2 zusammen. In der Tat zeigte sich auch für Domäne N1 eine sehr schnelle Entfaltung (t = 28 ms), wenn die Domäne N2 entfaltet vorliegt. Dies wurde in einem *stopped flow*-Doppelmischexperiment erreicht, indem denaturiertes

IIHY-G3P* vor der erneuten Entfaltung nur so kurz zurückgefaltet wurde, dass sich Domäne N1, nicht aber Domäne N2 falten konnte. Der geschwindigkeitsbestimmende Schritt der Entfaltung von G3P* ist somit entweder die Domänendissoziation selbst oder eine ihr vorgelagerte langsame Reaktion innerhalb der N2-Domäne.

Die apparenten Entfaltung- und Rückfaltungsraten als Funktion der GdmCl-Konzentration stimmen für die isolierte Domäne N1(T13I) und N1 verknüpft mit der denaturiert gehaltenen N2-Domäne in IIHY-G3P* überein (Abb. 13B). Damit wurde bestätigt, dass Domäne N1 eine unabhängige Faltungseinheit innerhalb von G3P* darstellt. N1 faltet und entfaltet im Millisekundenbereich, wird aber in seiner Entfaltung durch die Interaktionen mit der gefalteten Domäne N2 bis zu 150000-fach verlangsamt (bei 2,5 M GdmCl).

Die aus den Fluoreszenzamplituden der Kinetiken bei verschiedenen GdmCl-Konzentrationen abgeleiteten Stabilitäts- und Kooperativitätsparameter für Domäne N1 ($[GdmCl]_M$ = 3,4 M, m = 6,3 kJ·mol^{-1}·M^{-1}) entsprechen innerhalb der Fehlergrenzen den aus den Gleichgewichtsexperimenten gewonnenen Werten ($[GdmCl]_M$ = 3,3 M, m = 8,9 kJ·mol^{-1}·M^{-1}), so dass sich die Faltung von Domäne N1 in ihrer isolierten bzw. nicht-assoziierten Form mit dem Zweizustandsmodell beschreiben lässt.

G3P besitzt keine katalytische Aktivität. Ein direkter Nachweis, dass die anhand der Fluoreszenzänderungen detektierten Faltungsreaktionen tatsächlich zur Bildung des vollständig nativen Proteins führten, war daher nicht möglich. Stattdessen wurde hierfür ein Zweischritt-Entfaltungstest für native Moleküle genutzt (Schmid, 1983, Schmid, 1986). Die Methode basiert darauf, dass Moleküle, die bei ihrer Rückfaltung den nativen Zustand erreicht und somit die höchste Aktivierungsenergiebarriere bereits überwunden haben, langsam entfalten. Faltungsintermediate dagegen, die den hochenergetischen Übergangszustand noch nicht durchlaufen haben, entfalten wesentlich schneller. Die Amplitude der langsamen Entfaltungsreaktion ist also ein Maß für die Zahl der nativen Moleküle, die nach einer bestimmten Rückfaltungsdauer vorhanden waren. Die Zunahme dieser Amplitude als Funktion der Rückfaltungszeit ergibt somit die Kinetik der Bildung des vollständig gefalteten Proteins. Die entsprechenden Doppelmischexperimente ergaben, dass die mittels Tyr-Fluoreszenz detektierbaren Faltungsreaktionen der Domäne N2 lediglich zu Intermediaten führen. Die zweiphasigen Kinetiken spiegeln wahrscheinlich die Faltung des globulären Teils und der Gelenksubdomäne von N2 sowie die Bildung loser Kontakte mit Domäne N1 vor der eigentlichen Domänenassoziation wider. Die Rückfaltungsreaktion zum nativen IIHY-G3P* ausgehend von diesen Intermediaten ist etwa 100-mal langsamer (bei 0,5 M GdmCl) und fluoreszenzspektroskopisch nicht sichtbar. Sie zeigt eine Zeitkonstante von t = 6200 s und ließ sich sowohl über die Amplitudenzunahme der langsamen Entfaltung von Domäne N2, als auch über die Amplitudenzunahme der langsamen Entfaltung von Domäne N1 detektieren. Das heißt, dieser letzte Schritt der Rückfaltung ist für die kinetische Kopplung der Entfaltungsreaktionen von N2 und N1 verantwortlich und daher der Assoziation der beiden Domänen zuzuordnen.

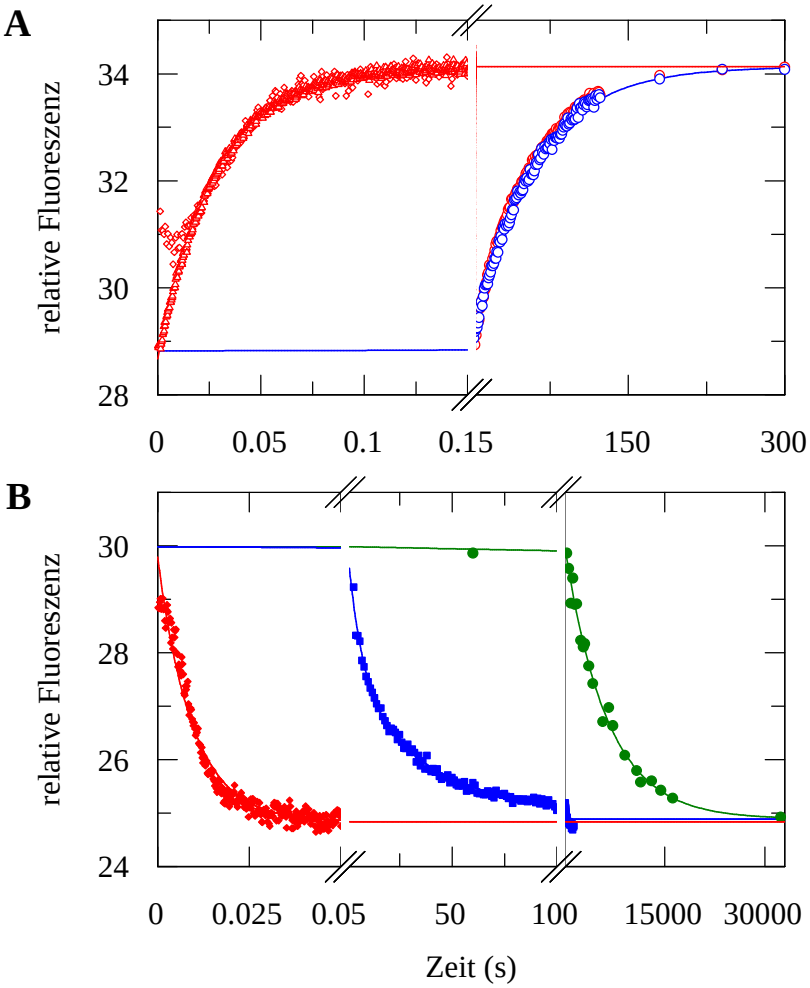


Abbildung 11. Entfaltungs- und Rückfaltungskinetiken von IIHY-G3P*. **(A)** Übereinstimmende Entfaltungskinetiken (t = 71 s) für die Domänen N2 (○) und N1 (○) in IIHY-G3P* bei 4,5 M GdmCl ausgehend vom nativen Protein. (◇) Entfaltungskinetik von Domäne N1 in IIHY-G3P* (t = 28 ms) bei 4.5 M GdmCl ausgehend vom kurzzeitig zurückgefalteten Protein (1 s-Rückfaltungspuls in 0,5 M GdmCl, Domäne N2 bleibt dabei entfaltet). (△) Entfaltungskinetik der isolierten Domäne N1(T13I) (t = 28 ms) bei 4.5 M GdmCl. **(B)** Rückfaltungskinetiken von (◆) Domäne N1 (t = 9,4 ms) und (■) Domäne N2 (t_1 = 7 s, t_2 = 42 s) in IIHY-G3P* sowie (●) Kinetik der langsamen Domänenassoziation bei 0,5 M GdmCl ausgehend vom vollständig denaturierten Protein. Die durchgezogenen Linien zeigen die Analysen der Daten als mono- und biexponentielle Funktionen. Für eine bessere Vergleichbarkeit wurden die Reaktionskinetiken in (A) und (B) so normiert, dass Start- und Endwerte übereinstimmen. Alle Messungen erfolgten wie in Teilarbeit E beschrieben.

Über Doppelmischexperimente, bei denen die Rückfaltung der Domänen N1 und N2, jedoch nicht deren langsame Assoziation zugelassen wurde, konnten die Entfaltungskinetiken der Domäne N2 in ihrer nicht-assoziierten Form bestimmt werden. Zusammen mit den kinetischen Daten für die schnelle Rückfaltungphase von N2 ergab sich ein komplettes Chevron der apparenten Faltungsraten l für Domäne N2 vor Assoziation mit N1. Es erlaubt Rückschlüsse auf die Stabilität dieser Domäne in Abwesenheit der Wechselwirkungen mit Domäne N1 (s. Abb. 13B). Der aus den Kinetiken berechnete Mittelpunkt des Übergangs liegt bei $[GdmCl]_M$ = 1,6 M und damit 0,9 M niedriger als im Gleichgewichtsübergang (Teilarbeit D). Dies verdeutlicht den hohen Beitrag der Interdomänenwechselwirkungen zur Stabilität von Domäne N2 im nativen Protein. Aus den Kooperativitätsparametern des Entfaltungsastes (m_{NU}) und des Rückfaltungsastes (m_{UN}) in der Chevronauftragung wurde eine Kooperativität des kinetischen Entfaltungsübergangs

von $m = 18{,}7\ \mathrm{kJ \cdot mol^{-1} \cdot M^{-1}}$ bestimmt, die vergleichbar ist mit dem im Gleichgewichtsexperiment bestimmten Wert ($m = 14{,}9\ \mathrm{kJ \cdot mol^{-1} \cdot M^{-1}}$). Diese hohe Kooperativität deutet darauf hin, dass sich der globuläre Teil von N2 vermutlich gemeinsam mit der Gelenksubdomäne faltet. Beide Subdomänen bilden also auch im Zeitverlauf der Rückfaltung eine kooperative Einheit, und nicht nur im Gleichgewichtsexperiment, in dem die Entfaltung von N2 mit der Domänendissoziation thermodynamisch gekoppelt ist. Demnach stellt auch die Domäne N2 eine separate Faltungseinheit dar, die in Isolierung stabil ist und sich vor der Assoziation mit Domäne N1 ausbildet. Bei der Denaturierung von G3P* ist die Domänendissoziation geschwindigkeitsbestimmend, und die Entfaltungsreaktionen von N1 und N2 sind kinetisch gekoppelt. Dagegen sind im Gleichgewichtsexperiment die Domänendissoziation und die Entfaltung von N2 thermodynamisch gekoppelt, da die Interdomänenwechselwirkungen einen sehr großen Beitrag zur Stabilität der N2-Domäne liefern. Diese Wechselwirkungen werden erst bei relativ hohen Konzentrationen an Denaturierungsmittel getrennt, bei denen die N2-Domäne in Isolierung bereits vollständig entfaltet ist. Es existiert daher kein Gleichgewichtsintermediat, in dem beide Domänen gefaltet, jedoch nicht miteinander assoziiert sind.

Basierend auf den Kristallstrukturen von G3P* wurde für die Gelenksubdomäne von N2 eine Scharnierfunktion bei der Trennung der Domänen im Verlauf des Infektionsvorgangs postuliert (Holliger *et al.*, 1999). Einen ersten Hinweis darauf, dass in der Tat Konformationsänderungen in der Subdomäne für die Domänenassoziation bzw. -dissoziation während der Faltungsreaktionen verantwortlich sein könnten, brachte die Untersuchung der IIY-G3P*-Variante. In dieser Variante befindet sich an Position 129 im globulären Teil von N2 ein Glutamin wie im Wildtypprotein, wodurch Domäne N2 etwa 7 kJ/mol instabiler ist als in IIHY-G3P*. Wie für eine Destabilisierung des globulären Teils erwartet, werden die Rückfaltungsreaktionen der N2-Domäne etwa 5-fach verlangsamt und die Entfaltung ungefähr 3-fach beschleunigt. Nicht beeinflusst wird dagegen die Domänenassoziation im letzten Rückfaltungsschritt, ihre Zeitkonstante von $t = 6700$ s ist nahezu identisch mit der von IIHY-G3P* ($t = 6200$ s). Somit ist die Kinetik der Domänenassoziation unabhängig von der intrinsischen Stabilität der individuellen Domänen.

2.3.3 Die Kinetik der Domänenassoziation in G3P wird bestimmt durch eine sehr langsame Prolylisomerisierung in der Gelenksubdomäne

Konformationelle Faltungsreaktionen sind in der Regel stark von der Denaturierungsmittelkonzentration abhängig. Je ungünstiger die Bedingungen für die Rückfaltung in den nativen Zustand werden, desto langsamer wird sie (Tanford, 1968). Die Domänenassoziation in IIHY-G3P* zeigt jedoch keine derartige GdmCl-Abhängigkeit, und ihre Zeitkonstante wurde für verschiedene Denaturierungsmittelkonzentrationen mit $t = 6200$ s bestimmt (Abb. 13B, Abb.2 in Teilarbeit F). Solch denaturierungsmittelunabhängige Raten sind ein typisches Kennzeichen von Faltungsreaktionen, die in ihrer Geschwindigkeit durch Prolylisomerisierungen limitiert sind (Balbach & Schmid, 2001). Identifizieren lassen sich derartige Reaktionen über die Katalysierbarkeit durch Prolylisomerasen (Schmid, 1993), vorausgesetzt das entsprechende Prolin ist während der Faltung zugänglich. Tatsächlich konnte die langsame Rückfaltungsreaktion in IIHY-G3P* durch Cyclophilin katalysiert werden, und die Zeitkonstante nahm in Gegenwart von 1 µM dieses Enzyms von 6300 s auf 1100 s ab. Daraus ist zu schließen, dass für die sehr langsame Domänenassoziation die Isomerisierung einer Prolyl-Peptidbindung verantwortlich ist.

Für die Untersuchung der entsprechenden *cis*⇌*trans*-Isomerisierung im denaturierten Protein waren Dreifachmischexperimente nötig, da bei G3P* der Prolin-limitiere Rückfaltungsschritt spektroskopisch nicht detektierbar ist. In diesen Experimenten wurde IIHY-G3P* für unterschiedlich lange Zeiten entfaltet, kurzzeitig rückgefaltet und dann erneut entfaltet. Nur diejenigen Moleküle, die nach der ersten Entfaltung noch ein natives Prolin-Isomer aufwiesen, konnten im zweiten Schritt schnell zum vollständig gefalteten Zustand mit assoziierten Domänen zurückkehren und so im dritten Schritt die für natives G3P* typische langsame Entfaltungsreaktion zeigen. Demnach war es möglich, anhand der Abnahme der Amplitude der langsamen Entfaltungsreaktion im dritten Schritt als Funktion der Entfaltungsdauer im ersten Schritt die Kinetik der *cis*⇌*trans*-Isomerisierung im denaturierten Protein zu verfolgen (s. Teilarbeit F). Für die Isomerisierung wurde eine apparante Rate von $k_{app} = 0{,}0029\ s^{-1}$ beobachtet, die damit deutlich kleiner ist als die Rate der konformationellen Entfaltung von IIHY-G3P* unter gleichen Bedingungen ($l_{app} = 0{,}033\ s^{-1}$ bei 5,0 M GdmCl). Eine Kopplung zwischen konformationellen Entfaltungsreaktionen und der Prolylisomerisierung kann daher nahezu ausgeschlossen werden. Aus der apparenten Isomerisierungsrate k_{app} und der über die Rückfaltung bestimmten mikroskopischen Rate der *trans*→*cis*-Isomerisierung $k_{tc} = 0{,}00016\ s^{-1}$ ergibt sich für die *cis*→*trans*-Isomerisierung ein k_{ct} von $0{,}00274\ s^{-1}$. Die abgeleitete Gleichgewichtskonstante K_{tc} beträgt 0,06, das heißt, bei 94 % aller Moleküle im denaturierten Zustand ist das für die Faltungsgeschwindigkeit entscheidende Prolin in der nicht-nativen *trans*-Konformation. Im Vergleich zu bekannten Prolylisomerisierungen in Peptiden (Stein, 1993) und Proteinen (Cook *et al.*, 1979, Mayr *et al.*, 1996, Schreiber & Fersht, 1993, Kamen & Woody, 2002) ist die in G3P* etwa 100-fach langsamer. Aufgrund der GdmCl-unabhängigen Domänenassoziation in der Rückfaltung ist davon auszugehen, dass diese sehr langsame *trans*→*cis*-Isomersisierung nicht durch eine Kopplung mit konformationellen Faltungsreaktionen bedingt ist, wie sie z.B. bei RNaseT1 auftritt. Vielmehr scheinen die Raten der *cis*/*trans*-Interkonversion tatsächlich von der lokalen Sequenz um das entsprechende *cis*-Prolin bestimmt zu werden.

G3P* enthält im nativen Zustand zwei *cis*-Proline, Pro161 im globulären Teil von Domäne N2 und Pro213 in der Gelenksubdomäne (Abb. 12; Lubkowski *et al.*, 1998). Die kinetischen Experimente mit IIHY-G3P* und der destabilisierten Variante IIY-G3P* hatten bereits auf die Bedeutung der Gelenksubdomäne für die Domänenassoziation hingedeutet (Teilarbeit E). Pro213 zeichnet sich zudem durch eine besondere Sequenzumgebung aus. Die vorhergehende Aminosäure ist Gln212, und gefolgt wird die *cis*-Bindung durch ein weiteres Prolin, Pro214. Aus Untersuchungen von *cis*⇌*trans*-Isomerisierungen in kurzen Peptiden ist bekannt, dass Gln-Pro-Bindungen im Vergleich zu Ala-Pro etwa dreimal langsamer isomerisieren, und dass ein Prolin als nachfolgende Aminosäure die Isomerisierung von Ala-Pro ebenfalls um den Faktor drei verlangsamt (Reimer *et al.*, 1998). Daraufhin wurde das *cis*-Pro213 als verantwortlich für die ratenlimitierende Isomerisierung im letzten Faltungsschritt von G3P* angenommen und durch Glycin ersetzt. Die P213G-Mutation destabilisiert Domäne N2 im Gleichgewichtsübergang um 8 kJ/mol und beschleunigt die Entfaltung ausgehend vom nativen Protein rund 20-fach. Sie hat allerdings keinen Effekt auf die fluoreszenzdetektierten Rückfaltungsreaktionen der N2-Domäne. Offenbar beeinflusst der P213G-Austausch nicht die intrinsische Stabilität der nicht-assoziierten N2-Domäne, sondern nur deren Wechselwirkungen mit Domäne N1. Diese Wechselwirkungen werden erst sehr spät bei der Faltung von G3P* ausgebildet und zum größten Teil durch die Gelenksubdomäne vermittelt.

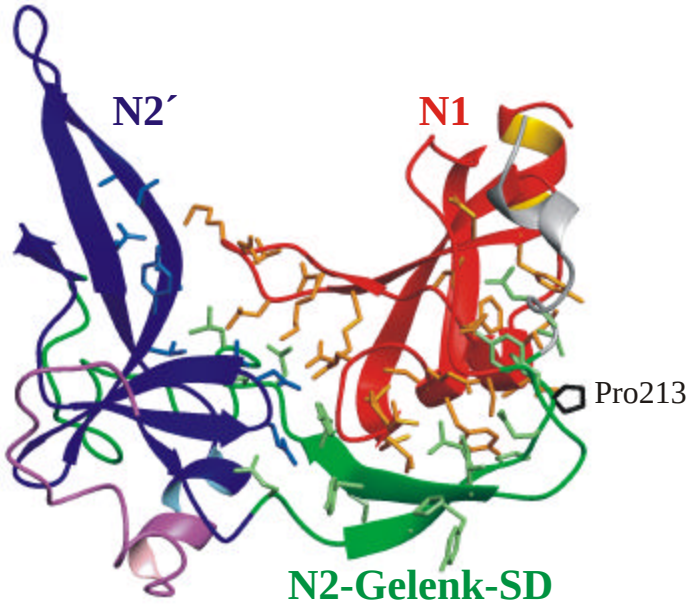


Abbildung 12. Tertiärstrukturdarstellung von G3P*, basierend auf den Strukturdaten von Holliger *et al.* (1999). Domäne N1 ist rot, der globuläre Teil von N2 (N2´) blau und die Gelenksubdomäne von N2 grün gefärbt. Die Seitenketten der an den Interdomänenwechselwirkungen beteiligten Aminosäuren in N1 und N2 (Lubkowski *et al.*, 1998) sind als Stäbchen dargestellt. Das *cis*-Pro213 in der Gelenksubdomäne, das für die extrem langsame Domänenassoziation bei der Rückfaltung verantwortlich ist, ist markiert. Violett gefärbt ist die Schleifenregion in Domäne N2, die den größten Teil der Wechselwirkungen mit dem F-Pilus vermittelt (Aminosäuren 178 – 199, Deng & Perham, 2002).

Die Isomerisierung an Pro213 ist tatsächlich geschwindigkeitsbestimmend für die Domänenassoziation von G3P*. Den Beweis dafür brachte der Effekt der P213G-Mutation auf die Kinetik dieser langsamsten Rückfaltungsreaktion. Der P213G-Austausch verringerte die Zeitkonstante der Assoziation beider Domänen von 6300 s auf 240 s. Durch diese fast 30-fache Beschleunigung wurde außerdem sichtbar, dass die Assoziationsreaktion erst mit einer Verzögerung einsetzt, deren Ausmaß mit der Faltungsrate von N2 korreliert. Dies deutet darauf hin, dass die Domänen in G3P* erst assoziieren können, wenn die Rückfaltung von Domäne N2 stattgefunden hat. Domänenfaltung und -assoziation stellen demzufolge getrennte Schritte in einem sequentiellen Faltungsmechanismus dar (Abb. 13A).

Die mittels der Dreifachsprungexperimente messbare *cis*⇌*trans*-Isomerisierung im denaturierten Protein wird durch den P213G-Austausch etwa 70-fach beschleunigt (t_{app} = 5.6 s), so dass sich mikroskopische Isomerisierungsraten von k_{tc} = 0,0042 s^{-1} und k_{ct} = 0,173 s^{-1} ergeben. Die daraus resultierende Gleichgewichtskonstante K_{tc} = 0,024 sagt aus, dass 97,6 % aller Moleküle der P213G-Mutante im denaturierten Zustand das nicht-native *trans*-Isomer enthalten. Unklar bleibt allerdings, um welche *cis*⇌*trans*-Isomerisierung im Protein es sich handelt. Gegen eine Isomerisierung am zweiten, noch verbliebenen *cis*-Pro161 spricht, dass sich die Rückfaltungsreaktion der P213G-Mutante nicht durch die Prolylisomerase Cyclophilin 18 katalysieren ließ, obwohl Pro161 selbst im nativen Zustand stark exponiert und gut zugänglich ist. Andererseits sind die detektierten Raten für eine Isomerisierung der Nicht-Prolylpeptidbinding Gln212-Gly213 relativ niedrig und der Anteil des *cis*-Isomers im Gleichgewicht des entfalteten Proteins mit 2,4 %

ungewöhnlich hoch (Odefey *et al.*, 1995, Vanhove *et al.*, 1996, Wheeler *et al.*, 1998). Allerdings ist anzumerken, dass nur sehr wenige Referenzdaten zur Kinetik der Isomerisierung von Nicht-Prolylpeptidbindungen existieren. Es kann nicht ausgeschlossen werden, dass die *cis*⇌*trans*-Isomerisierung durch eine Kopplung mit der konformationellen Entfaltung scheinbar verlangsamt wurde. Möglich ist auch der Einfluss einer potentiellen Isomerisierung an Pro214 benachbart zu Gly213 im denaturierten Zustand.

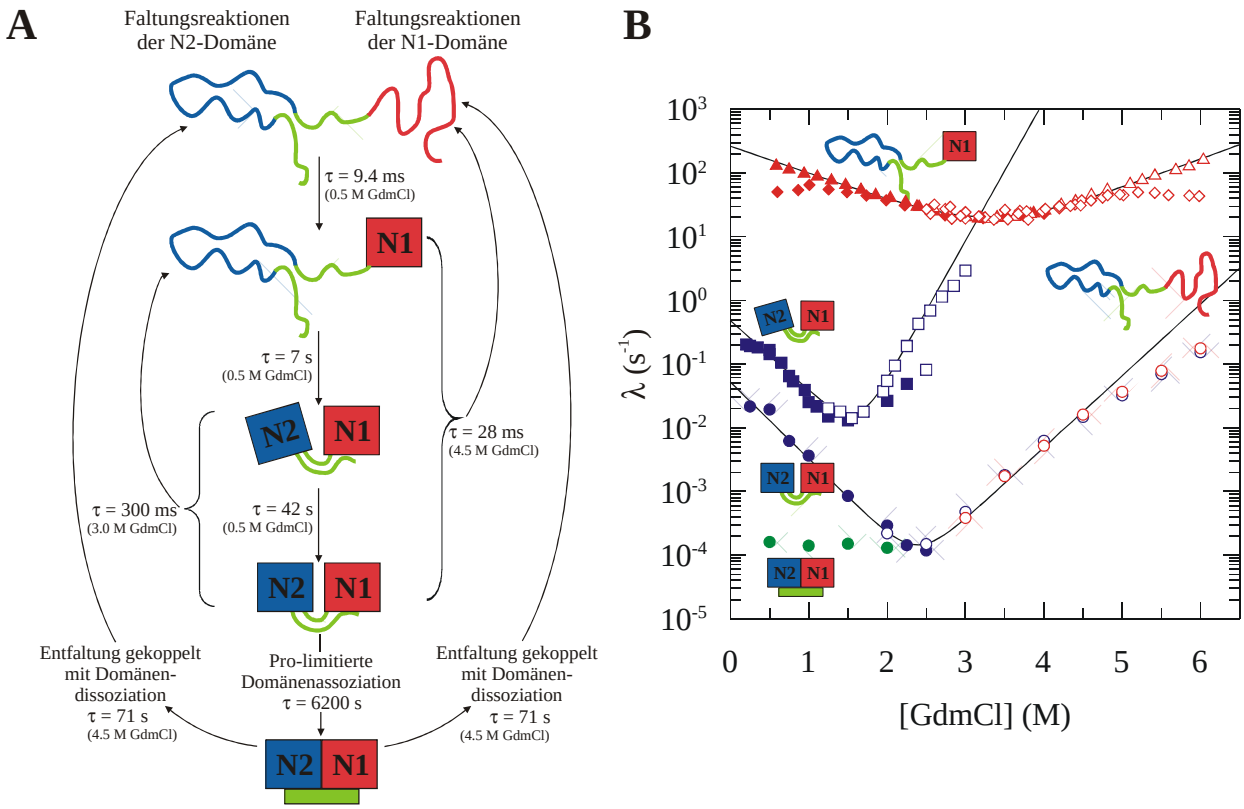


Abbildung 13. (A) Modell des Faltungsmechanismus von IIHY-G3P*. Angegeben sind die Zeitkonstanten *t* für die Rückfaltungsreaktionen der Domänen N1 und N2 bei 0.5 M GdmCl sowie für die Entfaltungsreaktionen bei 4,5 bzw. 3,0 M GdmCl. Der erste Schritt der Rückfaltung zum nativen G3P* ist die sehr schnelle Faltung von Domäne N1 (rotes Quadrat), gefolgt von der Faltung des globulären Teils von N2 (blaues Quadrat). Im letzten Schritt kommt es zur sehr langsamen, Prolin-limitierten Assoziation der Domänen und zur Ausbildung der nativen Konformation der Gelenksubdomäne von N2 (grünes Rechteck). Die Domänenassoziation führt zur kinetischen Kopplung beider Domänen und verlangsamt deren Entfaltung drastisch. **(B)** Apparente Raten *l* der Entfaltungs- und Rückfaltungsreaktionen der Domänen N1 (rot) und N2 (blau) in IIHY-G3P* als Funktion der GdmCl-Konzentration. (○) Entfaltung von Domäne N1 und (○) Entfaltung von Domäne N2 ausgehend vom nativen IIHY-G3P*. (◇) schnelle Entfaltung von N1 und (□) schnelle Entfaltung von N2 ausgehend vom kurzzeitig zurückgefalteten Intermediat mit nichtassoziierten Domänen. (△) schnelle Entfaltung und (▲) schnelle Rückfaltung der isolierten Domäne N1(T13I). (◆) schnelle Rückfaltung von N1 und (■, ●) zweiphasige Rückfaltung von N2 in IIHY-G3P*. (●) GdmCl-unabhängige Raten der Prolylisomerisierung während der Assoziation von N1 und N2.
Zur Verdeutlichung der einzelnen Faltungsreaktionen von IIHY-G3P* sind die entsprechenden Raten mit den in Teilabbildung (A) verwendeten Symbolen der Faltungszustände gekennzeichnet. Die Messung der Faltungsraten erfolgte wie in den Teilarbeiten E und F beschrieben.

2.3.4 Die lokale Sequenzumgebung ist für die sehr langsame *cis*⇌*trans*-Isomerisierung an Pro213 verantwortlich

Um nachzuweisen, dass tatsächlich die lokale Sequenzumgebung des *cis*-Pro213 für die sehr niedrigen Isomerisierungsraten bei der Domänenassoziation in G3P* verantwortlich ist, wurde ein entsprechendes Pentapeptid Ac-AQPPV-OH synthetisiert und dessen Raten der Prolylisomerisierung mittels 1D ^{1}H-NMR-Spektroskopie bestimmt. Für eine bessere Zuordnung der Aminosäureresonanzen war das Pro211 der G3P*-Sequenz durch Ala1 im Peptid ersetzt worden. Es ist bekannt, dass sich der Anteil der *cis*-Isomere in Peptiden von 10-12 % in wässrigen Lösungsmitteln auf bis zu 30-60 % in trockenen TFE/LiCl-Mischungen steigern lässt (Kofron *et al.*, 1991). Durch einen Sprung aus TFE/LiCl zurück in wässrigen Puffer können somit die reversiblen Isomerisierungsreaktionen bei der Nachstellung des *cis*/*trans*-Gleichgewichts anhand der Signale in zeitabhängig aufgenommenen ^{1}H-NMR-Spektren untersucht werden (Grathwohl & Wüthrich, 1981).

Das in TFE mit 0,6 M LiCl gelöste Peptid Ac-AQPPV-OH wurde im NMR-Spektrometer (Bruker DRX500) mit 20 mM Kaliumphosphatpuffer, pH 5,0 zehnfach verdünnt und die Kinetik der Isomerisierungsreaktionen über 21 min in 256 1D-Spektren (jeweils 4 Scans pro Spektrum) bei 25 °C verfolgt. Die Zuordnung der Resonanzen zu den Protonen der einzelnen Aminosäuren im Peptid erfolgte anhand eines DQF-COSY-Spektrums. Da das Peptid zwei Proline enthält, gibt es vier mögliche Isomere: *cis*/*cis, cis*/*trans, trans*/*cis* und *trans*/*trans*. Lediglich die Resonanzen des *trans*/*trans*-Isomers konnten mit Hilfe des 2D-Spektrums eindeutig zugeordnet werden, alle anderen Isomere waren dafür zu gering konzentriert.

Die isomerspezifischen Signale der beiden Prolinreste hatten in den 1D-Spektren zu niedrige Intensitäten für eine kinetische Auswertung. Sehr gut ließ sich allerdings die Resonanz des Amidprotons von Gln2 im *cis*-Isomer (*cis*/*cis* oder *cis*/*trans*) analysieren, die mit dem Signal des Amidprotons von Ala1 im *trans*/*trans*-Isomer bei 8,3 ppm zusammenfiel. Der zeitliche Verlauf des Integrals dieses Signals in den einzelnen 1D-Spektren zeigte eine biexponentielle Kinetik (Abb. 14). Deren erste Phase mit t = 23 s spiegelt möglicherweise die Isomerisierung der Pro3-Pro4 Peptidbindung wider. Die zweite Phase mit einer Zeitkonstante von t = 395 s stimmt sehr gut mit der *cis*⇌*trans*-Isomerisierung von Pro213 im denaturierten IIHY-G3P* (t = 345 s) überein, wie sie in den Dreifachsprungexperimenten fluoreszenzspektroskopisch detektiert werden konnte. Sie ist daher vermutlich der Isomerisierung der Gln2-Pro3-Bindung im Pentapeptid zuzuordnen. Die sehr gute Übereinstimmung der Raten für das Peptid und das Protein kann als deutliches Indiz dafür gesehen werden, dass die langsame Isomersisierung von Pro213 in G3P* tatsächlich auf die lokale Sequenzumgebung zurückzuführen ist. Konformationelle Faltungsreaktionen oder die Ausbildung der Interdomänenwechselwirkungen haben keinen Einfluss auf die Kinetik der Domänenassoziation in G3P*.

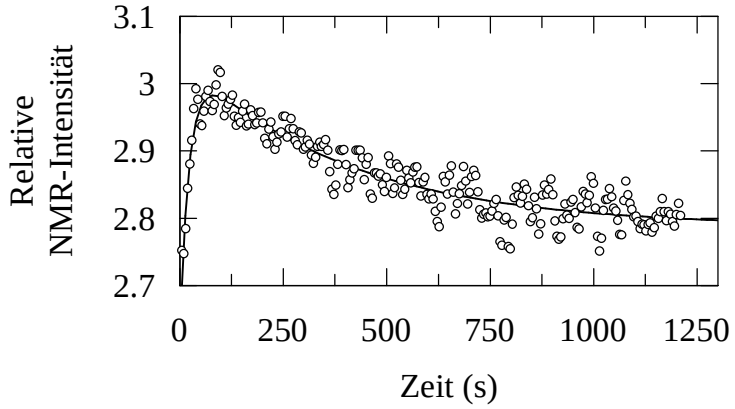


Abbildung 14. Zweiphasige Kinetik der *cis*→*trans*-Prolylisomerisierungen im Pentapeptid Ac-AQPPV-OH. Gezeigt ist der zeitliche Verlauf des Integrals des NMR-Signals bei 8,3 ppm, das die Resonanzen des Amidprotons von Ala1 im *trans*/*trans*-Isomer und des Amidprotons von Gln2 im *cis*/x-Isomer vereint. Die schnelle Phase (t = 23 s) reflektiert vermutlich die *cis*→*trans*-Isomerisierung der Pro3-Pro4-Bindung, während die langsame Phase (t = 395 s) die Isomerisierung der Gln2-Pro3-Bindung widerspiegelt. Die Isomerisierungskinetiken wurden gemessen in einem Bruker DRX500 NMR-Spektrometer bei 25 °C nach 10-facher Verdünnung des in TFE mit 0,6 M LiCl gelösten Pentapeptids. Die Endkonzentration war 5 mM Peptid in 20 mM Kaliumphosphat, pH 5,0, 60 mM LiCl, 10 % TFE, 10 % D_2O.

2.3.5 Ein Netz von Wasserstoffbrückenbindungen um Pro213 als Vermittler zwischen Prolylisomerisierung und Domänenassoziation

Wie lässt sich die beobachtete Kopplung zwischen der sehr langsamen *cis*⇌*trans*-Isomerisierung an Pro213 und der Assoziation der Domänen strukturell erklären? Die Kristallstruktur von G3P* zeigt, dass die beiden Stränge der Gelenksubdomäne von N2 vielfältige Wechselwirkungen mit Domäne N1 eingehen (Abb. 12). Die Region um das *cis*-Pro213 (Leu210 - Val215) sowie das benachbarte Segment zwischen Tyr92 und Pro98 zeichnen sich durch eine Anhäufung von Pro-Resten (Pro96, Pro98, Pro211, Pro213, Pro214) aus. Daraus resultieren geringe Flexibilitäten, wie aus niedrigen B-Faktoren in der Kristallstruktur ersichtlich (Lubkowski *et al.*, 1998, Holliger *et al.*, 1999). Auffällig ist auch die Orientierung der Seitenkette von Gln212, welches der *cis*-Peptidbinding vorausgeht. Gln212 bildet zwei Wasserstoffbrückenbindungen mit Gln52 in Domäne N1, dessen N^{ε} und O^{ε} lediglich 2,9 bzw. 3,0 Å entfernt sind. Eine dritte Bindung existiert zum Rückgratsauerstoff von Pro211, so dass die Konformation dieser Region um Pro213 zusätzlich fixiert wird. Flankiert werden diese Wasserstoffbrücken von drei weiteren Kontakten zwischen der Gelenksubdomäne von N2 und Domäne N1: Tyr92 ist wasserstoffverbrückt mit Thr18, und Val215 sowie Pro211 bilden Wasserstoffbrücken mit Ser16 aus (Abb. 7b in Teilarbeit F). Dieses Netzwerk von Wasserstoffbrücken ist vermutlich für die geordnete Interaktion zwischen N1 und N2 im nativen G3P* sehr wichtig, und es kann nur ausgebildet werden, wenn Pro213 in der *cis*-Konformation vorliegt. Möglicherweise bestimmt so die *trans*→*cis*-Isomerisierung an Pro213 eine Relativbewegung der beiden nahezu vollständig gefalteten Domänen und letztlich deren Assoziation bzw. Einrasten zum nativen Protein.

Ein Strukturvergleich der Gen-3-Proteine der Phagen M13 und fd deutete ebenfalls auf eine Bewegung von Domäne N2 relativ zu N1 hin, die um ein „Gelenk“ bei Gly99 und Ser208 in der Gelenksubdomäne von N2 erfolgen könnte (Holliger *et al.*, 1999) und eventuell Relevanz für die Funktion *in vivo* hat. Bei der Infektion von *E. coli* wird die Dissoziation von N1 und N2 dadurch induziert, dass die Spitze des F-Pilus an der Außenseite der

N2-Domäne bindet (s. Abb. 12; Deng & Perham, 2002). Diese Trennung der gefalteten Domänen ist mechanistisch noch nicht verstanden, sie beinhaltet eventuell Konformationsänderungen oder die partielle Entfaltung der Gelenksubdomäne. Es bleibt zu untersuchen, welche Rolle die Isomerisierung an Pro213 dabei spielt. Nur wenn N1 und N2 dissoziiert sind, ist die TolA-Bindungsstelle auf Domäne N1 exponiert. Dieser Zustand muss demnach im Verlauf des Infektionsvorgangs erhalten bleiben, bis der Pilus zurückgezogen und N1 in räumliche Nähe zu TolA auf der Zelloberfläche gebracht worden ist. Eine sehr langsame Reassoziation der Domänen, die durch die *trans*→*cis*-Isomerisierung an Pro213 kinetisch bestimmt wird, könnte also den dissoziierten Zustand für längere Zeit aufrechterhalten. Möglicherweise wirkt damit Pro213 im Gen-3-Protein als konformationeller Schalter bei der Infektion von *E. coli*.

3 Abkürzungen

Ac-AQPPV-OH	Pentapeptid mit acetyliertem N-Terminus und freier terminaler Carboxylgruppe
Bc-Csp	Kälteschockprotein aus *Bacillus caldolyticus*
Bs-CspB	Kälteschockprotein aus *Bacillus subtilis*
Csp	Kälteschockprotein
ΔC_p	Wärmekapazitätsänderung während der Entfaltung eines Proteins bei konstantem Druck
C_p^{E}	Wärmekapazität der Entfaltungsreaktion
CT	C-terminale Domäne des Gen-3-Proteins
D	Denaturierungsmittel
DQF-COSY	*double quantum filtered-correlated spectroscopy*
DSC	*differential scanning calorimetry*
ε_r	relative Dielektrizitätskonstante
ΔG_D	freie Enthalpie der Entfaltungreaktion
$\Delta\Delta G_D$	Differenz der freien Enthalpien der Entfaltung verschiedener Proteinvarianten
$\Delta G_D^{H_2O}$	freie Enthalpie der Entfaltung in Abwesenheit des Denaturierungsmittels
$\Delta\Delta G_D^{2M}$	Differenz der freien Enthalpien der Entfaltung verschiedener Proteinvarianten bei 2 M GdmCl
$\Delta\Delta G_D^{70°C}$	Differenz der freien Enthalpien der Entfaltung verschiedener Proteinvarianten bei 70 °C
GdmCl	Guanidiniumchlorid
$[GdmCl]_M$	Mittelpunkt des GdmCl-induzierten Entfaltungsübergangs
G3P	Gen-3-Protein filamentöser Phagen
G3P*	N-terminales Fragment von G3P bestehend aus den Domänen N1 und N2
IIHY-G3P*	G3P* mit den Mutationen T13I, T101I, Q129H und D209Y
IIY-G3P*	G3P* mit den Mutationen T13I, T101I und D209Y
G-5	stabilste Variante der Proside-Selektion in Gegenwart von GdmCl, CspB L2Y/E3I/A46Q/T64L/E66L/A67P
ΔH_D	Enthalpie der Entfaltungsreaktion

ΔH_{vH}	van´t Hoff-Enthalpie der Entfaltung
k_{app}	apparente Geschwindigkeitskonstante einer Reaktion
k_{tc}	Geschwindigkeitskonstante der *trans*→*cis*-Isomerisierung
k_{ct}	Geschwindigkeitskonstante der *cis*→*trans*-Isomerisierung
K_{tc}	Gleichgewichtskonstante der *cis*⇌*trans*-Isomerisierung
k_{NU}	mikroskopische Geschwindigkeitskonstante einer Entfaltungsreaktion
k_{UN}	mikroskopische Geschwindigkeitskonstante einer Rückfaltungsreaktion
l	apparente Ratenkonstante (= $k_{NU} + k_{UN}$)
m	Kooperativitätsparameter der Entfaltungsreaktion ($\delta\Delta G/\delta[D]$)
m_{NU}	kinetischer *m*-Wert der Entfaltungsreaktion ($\delta \ln k_{NU}/\delta[D]$)
m_{UN}	kinetischer *m*-Wert der Rückfaltungsreaktion ($\delta \ln k_{UN}/\delta[D]$)
N	nativer Zustand eines Proteins
N1	Domäne N1 des Gen-3-Proteins
N2	Domäne N2 des Gen-3-Proteins
Na-Cacodylat	Natriumsalz der Dimethylarsinsäure
NMR	*nuclear magnetic resonance*
ΔS_D	Entropie der Entfaltungsreaktion
ss-DNA	Einzelstrang-DNA
t	Zeitkonstante einer Reaktion
t_{app}	apparente Zeitkonstante einer Reaktion
t_M	Mittelpunkt des thermischen Entfaltungsübergangs
TFE	Trifluorethanol
Tm-Csp	Kälteschockprotein aus *Thermotoga maritima*
T-10/10	stabilste Variante der Proside-Selektion bei erhöhter Temperatur, CspB L2R/E3V/A46L/T64R/E66V/A67G
U	entfalteter Zustand eines Protein

4 Literaturverzeichnis

Akke, M. & Forsen, S. (1990). Protein stability and electrostatic interactions between solvent exposed charged side chains. *Proteins* **8**(1), 23-29.

Amitrano, C., Peliti, L. & Saber, M. (1989). Population dynamics in a spin-glass model of chemical evolution. *J Mol Evol* **29**(6), 513-25.

Bae, W., Xia, B., Inouye, M. & Severinov, K. (2000). Escherichia coli CspA-family RNA chaperones are transcription antiterminators. *Proc Natl Acad Sci U S A* **97**(14), 7784-9.

Balbach, J. & Schmid, F. X. (2000). Proline isomerization and its catalysis in protein folding. In *Mechanisms of Protein Folding* (Pain, R. H., ed.), pp. 212-237. Oxford University Press, Oxford.

Baldwin, R. L. (1996). How Hofmeister ion interactions affect protein stability. *Biophys. J.* **71**(4), 2056-2063.

Benner, S. A. & Ellington, A. D. (1988). Interpreting the behavior of enzymes: purpose or pedigree? *Crit.Rev.Biochem.Molec.Biol.* **23**, 369-426.

Benson, D. E., Wisz, M. S. & Hellinga, H. W. (2000). Rational design of nascent metalloenzymes. *Proc Natl Acad Sci U S A* **97**(12), 6292-7.

Boeke, J. D. & Model, P. (1982). A prokaryotic membrane anchor sequence: carboxyl terminus of bacteriophage f1 gene III protein retains it in the membrane. *Proc Natl Acad Sci U S A* **79**(17), 5200-5204.

Brandts, J. F., Hu, C. Q. & Lin, L.-N. (1989). A simple model for proteins with interacting domains. Application to scanning calorimetry data. *Biochemistry* **28**, 8588-8596.

Brown, B. M. & Sauer, R. T. (1999). Tolerance of Arc repressor to multiple-alanine substitutions. *Proc Natl Acad Sci U S A* **96**(5), 1983-8.

Cambillau, C. & Claverie, J. M. (2000). Structural and genomic correlates of hyperthermostability. *J Biol Chem* **275**(42), 32383-6.

Clarke, L. & Carbon, J. (1976). A colony bank containing synthetic Col El hybrid plasmids representative of the entire E. coli genome. *Cell* **9**(1), 91-9.

Click, E. M. & Webster, R. E. (1997). Filamentous phage infection: required interactions with the TolA protein. *J Bacteriol* **179**(20), 6464-6471.

Cook, K. H., Schmid, F. X. & Baldwin, R. L. (1979). Role of proline isomerization in folding of ribonuclease A at low temperatures. *Proc. Natl. Acad. Sci. U.S.A.* **76**, 6157-6161.

Dahiyat, B. I., Gordon, D. B. & Mayo, S. L. (1997). Automated design of the surface positions of protein helices. *Protein Sci* **6**(6), 1333-7.

Dahiyat, B. I. & Mayo, S. L. (1997). De novo protein design: fully automated sequence selection. *Science* **278**(5335), 82-7.

Dams, T., Auerbach, G., Bader, G., Jacob, U., Ploom, T., Huber, R. & Jaenicke, R. (2000). The crystal structure of dihydrofolate reductase from Thermotoga maritima: molecular features of thermostability. *J Mol Biol* **297**(3), 659-72.

de Bakker, P. I., Hunenberger, P. H. & McCammon, J. A. (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.

Delbrück, H., Mueller, U., Perl, D., Schmid, F. X. & Heinemann, U. (2001). Crystal structures of mutant forms of the *Bacillus caldolyticus* cold shock protein differing in thermal stability. *J. Mol. Biol.* **313**, 359-369.

Deng, L. W. & Perham, R. N. (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.

Dill, K. A. (1990). Dominant forces in protein folding. *Biochemistry* **29**(31), 7133-55.

Dill, K. A., Bromberg, S., Yue, K., Fiebig, K. M., Yee, D. P., Thomas, P. D. & Chan, H. S. (1995). Principles of protein folding--a perspective from simple exact models. *Protein Sci* **4**(4), 561-602.

Dobzhansky, T. (1973). Nothing in Biology Makes Sense Except in the Light of Evolution. *American Biology Teacher* **35**, 125-129.

Dower, W. J. & Mattheakis, L. C. (2002). In vitro selection as a powerful tool for the applied evolution of proteins and peptides. *Curr Opin Chem Biol* **6**(3), 390-8.

Dunn, I. S. (1996). Phage display of proteins. *Curr Opin Biotechnol* **7**(5), 547-53.

Ehrlich, S. D., Bierne, H., d'Alencon, E., Vilette, D., Petranovic, M., Noirot, P. & Michel, B. (1993). Mechanisms of illegitimate recombination. *Gene* **135**(1-2), 161-6.

Elcock, A. H. (1998). The stability of salt bridges at high temperatures: implications for hyperthermophilic proteins. *J Mol Biol* **284**(2), 489-502.

Frauenfelder, H., Sligar, S. G. & Wolynes, P. G. (1991). The energy landscapes and motions of proteins. *Science* **254**(5038), 1598-603.

Gao, J., Kuczera, K., Tidor, B. & Karplus, M. (1989). Hidden thermodynamics of mutant proteins: a molecular dynamics analysis. *Science* **244**(4908), 1069-72.

Garel, J.-R. (1992). Folding of large proteins:multidomain and multisubunit proteins. In *Protein Folding* 1 edit. (Creighton, T. E., ed.), pp. 405-454. Freeman, New York.

Georgiou, G., Stathopoulos, C., Daugherty, P. S., Nayak, A. R., Iverson, B. L. & Curtiss, R., 3rd. (1997). Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines. *Nat Biotechnol* **15**(1), 29-34.

Gershenson, A. & Arnold, F. H. (2000). Enzyme stabilization by directed evolution. *Genet Eng* **22**, 55-76.

Gilson, M. K. & Honig, B. H. (1986). The dielectric constant of a folded protein. *Biopolymers* **25**(11), 2097-119.

Giver, L., Gershenson, A., Freskgard, P. O. & Arnold, F. H. (1998). Directed evolution of a thermostable esterase. *Proc Natl Acad Sci U S A* **95**(22), 12809-13.

Glaser-Wuttke, G., Keppner, J. & Rasched, I. (1989). Pore-forming properties of the adsorption protein of filamentous phage fd. *Biochim Biophys Acta* **985**(3), 239-47.

Grathwohl, C. & Wüthrich, K. (1981). NMR studies of the rates of proline cis-trans isomerization in oligopeptides. *Biopolymers* **20**, 2623-2633.

Graumann, P. & Marahiel, M. A. (1996). Some like it cold: Response of microorganisms to cold shock. *Arch.Microbiol.* **166**(5), 293-300.

Graumann, P. L. & Marahiel, M. A. (1998). Function and evolution of cold shock proteins, an ancient nucleic acid-binding protein familiy. *Trends Biochem. Sci.* **in press**.

Graumann, P. L. & Marahiel, M. A. (1999). Cold shock response in Bacillus subtilis. *J Mol Microbiol Biotechnol* **1**(2), 203-9.

Grimsley, G. R., Shaw, K. L., Fee, L. R., Alston, R. W., Huyghues-Despointes, B. M., Thurlkill, R. L., Scholtz, J. M. & Pace, C. N. (1999). Increasing protein stability by altering long-range coulombic interactions. *Protein Sci* **8**(9), 1843-1849.

Guex, N. & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**(15), 2714-23.

Hanes, J. & Pluckthun, A. (1997). In vitro selection and evolution of functional proteins by using ribosome display. *Proc Natl Acad Sci U S A* **94**(10), 4937-42.

Hatanaka, H., Tanimura, R., Katoh, S. & Inagaki, F. (1997). Solution structure of ferredoxin from the thermophilic cyanobacterium *Synechococcus elongatus* and its thermostability. *Journal of Molecular Biology* **268**, 922-933.

Hendsch, Z. S., Jonsson, T., Sauer, R. T. & Tidor, B. (1996). Protein stabilization by removal of unsatisfied polar groups: Computational approaches and experimental tests. *Biochemistry* **35**(24), 7621-7625.

Holliger, P., Riechmann, L. & Williams, R. L. (1999). Crystal structure of the two N-terminal domains of g3p from filamentous phage fd at 1.9 Angström: evidence for conformational lability. *J Mol Biol* **288**(4), 649-57.

Honig, B. & Yang, A. S. (1995). Free energy balance in protein folding. *Adv Protein Chem* **46**(6), 27-58.

Horovitz, A. (1996). Double-mutant cycles: a powerful tool for analyzing protein structure and function. *Folding & Design* **1**, R121-R126.

Horovitz, A. & Fersht, A. R. (1992). Co-operative Interactions During Protein Folding. *Journal of Molecular Biology* **224**(3), 733-740.

Jackson, J. R., Sathe, G., Rosenberg, M. & Sweet, R. (1995). In vitro antibody maturation. Improvement of a high affinity, neutralizing antibody against IL-1 beta. *J Immunol* **154**(7), 3310-9.

Jaenicke, R. (1999). Stability and folding of domain proteins. *Progress in Biophysics & Molecular Biology* **71**(2), 155-241.

Jaenicke, R. & Böhm, G. (1998). The stability of proteins in extreme environments. *Curr. Opin. Struct. Biol.* **8**, 738-748.

Jiang, W., Hou, Y. & Inouye, M. (1997). CspA, the major cold-shock protein of Escherichia coli, is an RNA chaperone. *J Biol Chem* **272**(1), 196-202.

Jung, S. & Plückthun, A. (1997). Improving in vivo folding and stability of a single-chain Fv antibody fragment by loop grafting. *Protein Eng* **10**(8), 959-66.

Kamen, D. E. & Woody, R. W. (2002). Folding kinetics of the protein pectate lyase C reveal fast-forming intermediates and slow proline isomerization. *Biochemistry* **41**(14), 4713-23.

Karshikoff, A. & Ladenstein, R. (2001). Ion pairs and the thermotolerance of proteins from hyperthermophiles: a "traffic rule" for hot roads. *Trends Biochem Sci* **26**(9), 550-556.

Kast, P. & Hilvert, D. (1997). 3D structural information as a guide to protein engineering using genetic selection. *Curr Opin Struct Biol* **7**(4), 470-9.

Kauzmann, W. (1959). *Advances in Protein Chemistry* **14**, 1-63.

Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature* **217**(129), 624-6.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*, Cambridge University Press, Cambridge.

Kofron, J. L., Kuzmic, P., Kishore, V., Colonbonilla, E. & Rich, D. H. (1991). Determination of Kinetic Constants for Peptidyl Prolyl Cis- Trans Isomerases by an Improved Spectrophotometric Assay. *Biochemistry* **30**(25), 6127-6134.

Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**(1), 51-5, 29-32.

Korndörfer, I., Steipe, B., Huber, R., Tomschy, A. & Jaenicke, R. (1995). The crystal structure of holo-glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic bacterium Thermotoga mauitima at 2.5 angstrom resolution. *Journal of Molecular Biology* **246**(4), 511-521.

Krebber, C., Spada, S., Desplancq, D., Krebber, A., Ge, L. & Plückthun, A. (1997). Selectively-infective phage (SIP): a mechanistic dissection of a novel in vivo selection for protein-ligand interactions. *J Mol Biol* **268**(3), 607-18.

Kuchner, O. & Arnold, F. H. (1997). Directed evolution of enzyme catalysts. *Trends Biotechnol* **15**(12), 523-30.

Kuzminov, A. (1999). Recombinational repair of DNA damage in Escherichia coli and bacteriophage lambda. *Microbiol Mol Biol Rev* **63**(4), 751-813, table of contents.

Lebbink, J. H., Knapp, S., van der Oost, J., Rice, D., Ladenstein, R. & de Vos, W. M. (1998). Engineering activity and stability of Thermotoga maritima glutamate dehydrogenase. I. Introduction of a six-residue ion-pair network in the hinge region. *J Mol Biol* **280**(2), 287-96.

Lebbink, J. H., Knapp, S., van der Oost, J., Rice, D., Ladenstein, R. & de Vos, W. M. (1999). Engineering activity and stability of Thermotoga maritima glutamate dehydrogenase. II: construction of a 16-residue ion-pair network at the subunit interface. *J Mol Biol* **289**(2), 357-69.

Levinthal, C. J. (1968). Are there pathways for protein folding? *J. Chim. Phys.* **65**, 44-45.

Lindahl, T. (1982). DNA repair enzymes. *Annu Rev Biochem* **51**, 61-87.

Loladze, V. V., Ibarra-Molero, B., Sanchez-Ruiz, J. M. & Makhatadze, G. I. (1999). Engineering a thermostable protein via optimization of charge-charge interactions on the protein surface. *Biochemistry* **38**(50), 16419-23.

Looger, L. L. & Hellinga, H. W. (2001). Generalized dead-end elimination algorithms make large-scale protein side-chain structure prediction tractable: implications for protein design and structural genomics. *J Mol Biol* **307**(1), 429-45.

Lopez, M. M., Yutani, K. & Makhatadze, G. I. (1999). Interactions of the major cold shock protein of Bacillus subtilis CspB with single-stranded DNA templates of different base composition. *J Biol Chem* **274**(47), 33601-8.

Lubkowski, J., Hennecke, F., Plückthun, A. & Wlodawer, A. (1998). The Structural Basis of Phage Display Elucidated By the Crystal Structure of the N-Terminal Domains of G3p. *Nature Structural Biology* **5**(2), 140-147.

Lubkowski, J., Hennecke, F., Plückthun, A. & Wlodawer, A. (1999). Filamentous phage infection: crystal structure of g3p in complex with its coreceptor, the C-terminal domain of TolA. *Structure* **7**(6), 711-722.

Makhatadze, G. I. & Privalov, P. L. (1995). Energetics of protein structure. *Advances in Protein Chemistry* **47**, 307-425.

Malakauskas, S. M. & Mayo, S. L. (1998). Design, structure and stability of a hyperthermophilic protein variant. *Nat Struct Biol* **5**(6), 470-475.

Martin, A. (1998). Evolutive Strategie zur Selektion stabilisierter Varianten von Ribonuklease T1. Diplomarbeit, Universität Bayreuth.

Marvin, D. A. (1998). Filamentous phage structure, infection and assembly. *Curr Opin Struct Biol* **8**(2), 150-158.

Matsumura, M. & Aiba, S. (1985). Screening for thermostable mutant of kanamycin nucleotidyltransferase by the use of a transformation system for a thermophile, Bacillus stearothermophilus. *J Biol Chem* **260**(28), 15298-303.

Maynard Smith, J. (1970). Natural Selection and the Concept of a Protein Space. *Nature* **225**, 563-564.

Mayr, L. M., Odefey, C., Schutkowski, M. & Schmid, F. X. (1996). Kinetic analysis of the unfolding and refolding of ribonuclease T1 by a stopped-flow double-mixing technique. *Biochemistry* **35**, 5550-5561.

Merkler, D. J., Farrington, G. K. & Wedler, F. C. (1981). Protein thermostability. Correlations between calculated macroscopic parameters and growth temperature for closely related thermophilic and mesophilic bacilli. *Int J Pept Protein Res* **18**(5), 430-42.

Miyazaki, K. & Arnold, F. H. (1999). Exploring nonnatural evolutionary pathways by saturation mutagenesis: rapid improvement of protein function. *J Mol Evol* **49**(6), 716-20.

Miyazaki, K., Wintrode, P. L., Grayling, R. A., Rubingh, D. N. & Arnold, F. H. (2000). Directed evolution study of temperature adaptation in a psychrophilic enzyme. *J Mol Biol* **297**(4), 1015-26.

Mrabet, N. T., Van den Broeck, A., Van den brande, I., Stanssens, P., Laroche, Y., Lambeir, A. M., Matthijssens, G., Jenkins, J., Chiadmi, M., van Tilbeurgh, H. & et al. (1992). Arginine residues as stabilizing elements in proteins. *Biochemistry* **31**(8), 2239-53.

Müller, U., Perl, D., Schmid, F. X. & Heinemann, U. (2000). Thermal stability and atomic-resolution crystal structure of the *Bacillus caldolyticus* cold shock protein. *Journal of Molecular Biology* **297**(4), 975-988.

Narinx, E., Baise, E. & Gerday, C. (1997). Subtilisin from psychrophilic antarctic bacteria: characterization and site-directed mutagenesis of residues possibly involved in the adaptation to cold. *Protein Eng* **10**(11), 1271-9.

Netzer, W. J. & Hartl, F. U. (1997). Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature* **388**(6640), 343-349.

Newkirk, K., Feng, W., Jiang, W., Tejero, R., Emerson, S. D., Inouye, M. & Montelione, G. T. (1994). Solution NMR structure of the major cold shock protein (CspA) from *Escherichia coli*: Identification of a binding epitope for DNA. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 5114-5118.

Odefey, C., Mayr, L. M. & Schmid, F. X. (1995). Non-prolyl *cis-trans* peptide bond isomerization as a rate-determining step in protein unfolding and refolding. *Journal of Molecular Biology* **245**, 69-78.

Oshima, T. (1994). Stabilization of proteins by evolutionary molecular engineering techniques. *Current Opinion in Structural Biology* **4**(4), 623-628.

Ott, J. & Eckstein, F. (1987). Protection of oligonucleotide primers against degradation by DNA polymerase I. *Biochemistry* **26**(25), 8237-41.

Pace, C. N., Alston, R. W. & Shaw, K. L. (2000). Charge-charge interactions influence the denatured state ensemble and contribute to protein stability. *Protein Science* **9**(7), 1395-1398.

Pace, C. N., Shirley, B. A., McNutt, M. & Gajiwala, K. (1996). Forces contributing to the conformational stability of proteins. *Faseb J* **10**(1), 75-83.

Parsell, D. & Sauer, R. (1989). The structural stability of a protein is an important determinant of its proteolytic susceptibiliy in *Escherichia coli. J. Biol. Chem.* **264**, 7590-7595.

Perl, D. (1997). Stabilität und Faltungsmechanismus des Kälteschockproteins aus dem thermophilen Bakterium Bacillus caldolyticus. Diplomarbeit, Universität Bayreuth.

Perl, D., Holtermann, G. & Schmid, F. X. (2001). Role of the chain termini for the folding transition state of the cold shock protein. *Biochemistry* **40**(51), 15501-11.

Perl, D., Mueller, U., Heinemann, U. & Schmid, F. X. (2000). Two exposed amino acid residues confer thermostability on a cold shock protein. *Nat Struct Biol* **7**(5), 380-383.

Perl, D. & Schmid, F. X. (2001). Electrostatic stabilization of a thermophilic cold shock protein. *J. Mol. Biol.* **313**, 343-357.

Perl, D., Welker, C., Schindler, T., Schröder, K., Marahiel, M. A., Jaenicke, R. & Schmid, F. X. (1998). Conservation of rapid two-state folding in mesophilic, thermophilic, and hyperthermophilic cold shock proteins. *Nature Structural Biology* **5**, 229-235.

Phadtare, S., Inouye, M. & Severinov, K. (2002a). The nucleic acid melting activity of Escherichia coli CspE is critical for transcription antitermination and cold acclimation of cells. *J Biol Chem* **277**(9), 7239-45.

Phadtare, S., Tyagi, S., Inouye, M. & Severinov, K. (2002b). Three amino acids in Escherichia coli CspE surface-exposed aromatic patch are critical for nucleic acid melting activity leading to transcription antitermination and cold acclimation of cells. *J Biol Chem* **277**(48), 46706-11.

Privalov, P. L. (1979). Stability of proteins. *Advances in Protein Chemistry* **33**, 167-241.

Privalov, P. L. (1982). Stability of proteins. Proteins which do not present a single cooperative system. *Advances in Protein Chemistry* **35**, 1-104.

Proba, K., Worn, A., Honegger, A. & Plückthun, A. (1998). Antibody scFv fragments without disulfide bonds made by molecular evolution. *J. Mol. Biol.* **275**(2), 245-53.

Rakonjac, J. & Model, P. (1998). Roles of pIII in filamentous phage assembly. *J Mol Biol* **282**(1), 25-41.

Reid, K. L., Rodriguez, H. M., Hillier, B. J. & Gregoret, L. M. (1998). Stability and Folding Properties of a Model Beta-Sheet Protein, Escherichia Coli Csp A. *Protein Science* **7**(2), 470-479.

Reimer, U., Scherer, G., Drewello, M., Kruber, S., Schutkowski, M. & Fischer, G. (1998). Side-chain effects on peptidyl-prolyl cis/trans isomerisation. *J Mol Biol* **279**(2), 449-60.

Riechmann, L. & Holliger, P. (1997). The C-terminal domain of TolA is the coreceptor for filamentous phage infection of E. coli. *Cell* **90**(2), 351-360.

Roberts, R. W. & Szostak, J. W. (1997). RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proc Natl Acad Sci U S A* **94**(23), 12297-302.

Rudolph, R., Siebendritt, R., Neslauer, G., Sharma, A. K. & Jaenicke, R. (1990). Folding of an all-beta protein: independent domain folding in gamma-II-crystallin from calf eye lens. *Proc.Nat.Acad.Sci.USA* **87**, 4625-4629.

Russell, R. J., Ferguson, J. M., Hough, D. W., Danson, M. J. & Taylor, G. L. (1997). The crystal structure of citrate synthase from the hyperthermophilic archaeon pyrococcus furiosus at 1.9 A resolution. *Biochemistry* **36**(33), 9983-94.

Sanchez-Ruiz, J. M. & Makhatadze, G. I. (2001). To charge or not to charge? *Trends Biotechnol* **19**(4), 132-135.

Schindelin, H., Jiang, W., Inouye, M. & Heinemann, U. (1994). Crystal structure of CspA, the major cold shock protein of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 5119-5123.

Schindelin, H., Marahiel, M. A. & Heinemann, U. (1993). Universal Nucleic Acid-Binding Domain Revealed by Crystal Structure of the *B. subtilis* Major Cold-Shock Protein. *Nature* **364**, 164-168.

Schindler, T., Herrler, M., Marahiel, M. A. & Schmid, F. X. (1995). Extremely rapid folding in the absence of intermediates: the cold-shock protein from Bacillus subtilis. *Nature Structural Biology* **2**, 663-673.

Schmid, F. X. (1983). Mechanism of folding of ribonuclease A. Slow refolding is a sequential reaction via structural intermediates. *Biochemistry* **22**, 4690-4696.

Schmid, F. X. (1986). Fast-folding and slow-folding forms of unfolded proteins. In *Enzyme Structure Part L* 1 edit. (Hirs, C. H. W. & Timasheff, S. N., eds.), Vol. 131, pp. 71-82. Academic Press, New York.

Schmid, F. X. (1993). Prolyl Isomerase - Enzymatic Catalysis of Slow Protein-Folding Reactions. *Annual Review of Biophysics and Biomolecular Structure* **22**, 123-143.

Schnuchel, A., Wiltschek, R., Czisch, M., Herrler, M., Willimsky, G., Graumann, P., Marahiel, M. A. & Holak, T. A. (1993). Structure in solution of the major cold-shock protein from *Bacillus subtilis*. *Nature* **364**, 169-171.

Schreiber, G. & Fersht, A. R. (1993). The refolding of *cis*- and *trans*-peptidylprolyl isomers of Barstar. *Biochemistry* **32**, 11195-11203.

Schreuder, M. P., Mooren, A. T., Toschka, H. Y., Verrips, C. T. & Klis, F. M. (1996). Immobilizing proteins on the surface of yeast cells. *Trends Biotechnol* **14**(4), 115-20.

Shakhnovich, E. I. (1994). Proteins with selected sequences fold into unique native conformation. *Physical Review Letters* **72**(24), 3907-3910.

Sieber, V., Plückthun, A. & Schmid, F. X. (1998). Selecting proteins with improved stability by a phage-based method. *Nature Biotechnology* **16**, 955-960.

Skerra, A. (1992). Phosphorothioate primers improve the amplification of DNA sequences by DNA polymerases with proofreading activity. *Nucleic Acids Res* **20**(14), 3551-4.

Smith, G. P. & Scott, J. K. (1993). Libraries of peptides and proteins displayed on filamentous phage. *Methods Enzymol* **217**, 228-57.

Spada, S., Honegger, A. & Plückthun, A. (1998). Reproducing the Natural Evolution of Protein Structural Features With the Selectively Infective Phage (Sip) Technology - the Kink in the First Strand of Antibody Kappa Domains. *Journal of Molecular Biology* **283**(2), 395-407.

Spector, S., Wang, M., Carp, S. A., Robblee, J., Hendsch, Z. S., Fairman, R., Tidor, B. & Raleigh, D. P. (2000). Rational modification of protein stability by the mutation of charged surface residues. *Biochemistry* **39**(5), 872-9.

Stein, R. L. (1993). Mechanism of enzymatic and nonenzymatic prolyl cis-trans isomerization. *Advances in Protein Chemistry* **44**, 1-24.

Stemmer, W. P. (1994a). DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc Natl Acad Sci U S A* **91**(22), 10747-51.

Stemmer, W. P. (1994b). Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**(6488), 389-91.

Stengele, I., Bross, P., Garces, X., Giray, J. & Rasched, I. (1990). Dissection of functional domains in phage fd adsorption protein. Discrimination between attachment and penetration sites. *J Mol Biol* **212**(1), 143-149.

Street, A. G. & Mayo, S. L. (1999). Computational protein design. *Structure* **7**(5), R105-R109.

Strop, P. & Mayo, S. L. (2000). Contribution of surface salt bridges to protein stability. *Biochemistry* **39**(6), 1251-5.

Tanford, C. (1968). Protein denaturation. Part B. The transition state from native to denatured state. *Advances in Protein Chemistry* **23**, 218-282.

Tanford, C. (1970). Protein denaturation Part C. *Advances in Protein Chemistry* **24**, 1-95.

Tsalkova, T. N. & Privalov, P. L. (1985). Thermodynamic study of domain organization in troponin C and calmodulin. *J Mol Biol* **181**(4), 533-544.

Tsunenaga, M., Goto, Y., Kawata, Y. & Hamaguchi, K. (1987). Unfolding and refolding of a type kappa immunoglobulin light chain and its variable and constant fragments. *Biochemistry* **26**, 6044-6051.

Van den Burg, B., Vriend, G., Veltman, O. R., Venema, G. & Eijsink, V. G. (1998). Engineering an enzyme to resist boiling. *Proc Natl Acad Sci U S A* **95**(5), 2056-60.

Vanhove, M., Raquet, X., Palzkill, T., Pain, R. H. & Frere, J. M. (1996). The rate-limiting step in the folding of the cis-Pro167Thr mutant of TEM-1 beta-lactamase is the trans to cis isomerization of a non-proline peptide bond. *Proteins: Struct. Funct. Genet.* **25**(1), 104-111.

von Hippel, P. H. & Wong, K.-Y. (1965). On the conformational stability of globular proteins. The effects of various electrolytes and nonelectrolytes on the thermal ribonuclease transition. *J. Biol. Chem.* **240**, 3909-3923.

Wassenberg, D., Welker, C. & Jaenicke, R. (1999). Thermodynamics of the unfolding of the cold-shock protein from Thermotoga maritima. *Journal of Molecular Biology* **289**(1), 187-193.

Wells, J. A. (1990). Additivity of mutational effects in proteins. *Biochemistry* **29**(37), 8509-17.

Wenk, M. & Jaenicke, R. (1999). Calorimetric analysis of the Ca^{2+}-binding beta gamma-crystallin homolog protein S from Myxococcus xanthus: Intrinsic stability and mutual stabilization of domains. *Journal of Molecular Biology* **293**(1), 117-124.

Wheeler, K. A., Hawkins, A. R., Pain, R. & Virden, R. (1998). The slow step of folding of Staphylococcus aureus PC1 beta-lactamase involves the collapse of a surface loop rate limited by the trans to cis isomerization of a non-proline peptide bond. *Proteins* **33**(4), 550-7.

Wintrode, P. L. & Arnold, F. H. (2000). Temperature adaptation of enzymes: lessons from laboratory evolution. *Advances in Protein Chemistry* **55**(5), 161-225.

Xiao, L. & Honig, B. (1999). Electrostatic contributions to the stability of hyperthermophilic proteins. *J Mol Biol* **289**(5), 1435-1444.

Yip, K. S. P., Stillman, T. J., Britton, K. L., Artymiuk, P. J., Baker, P. J., Sedelnikova, S. E., Engel, P. C., Pasquo, A., Chiaraluce, R., Consalvi, V., Scandurra, R. & Rice, D. W. (1995). The structure of Pyrococcus furiosus glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. *Structure.* **3**(11), 1147-1158.

Zhang, X., Meining, W., Fischer, M., Bacher, A. & Ladenstein, R. (2001). X-ray structure analysis and crystallographic refinement of lumazine synthase from the hyperthermophile Aquifex aeolicus at 1.6 A resolution: determinants of thermostability revealed from structural comparisons. *J Mol Biol* **306**(5), 1099-114.

Zhao, H. & Arnold, F. H. (1999). Directed evolution converts subtilisin E into a functional equivalent of thermitase. *Protein Eng* **12**(1), 47-53.

5 Publikationsliste

A. Andreas Martin, Franz X. Schmid, and Volker Sieber.
Proside: A Phage-based Method for Selecting Thermostable Proteins.
In *Methods in Molecular Biology* **230**, *Directed Enzyme Evolution: Screening and Selection Methods* (Arnold, R.H. & Georgiou, G., ed.), Humana Press Inc., Totowa, NJ, 57-70 (2003)

B. Andreas Martin, Volker Sieber and Franz X. Schmid.
In-vitro Selection of Highly Stabilized Protein Variants with Optimized Surface.
The Journal of Molecular Biology **309**, 717-726 (2001)

C. Andreas Martin, Insa Kather and Franz X. Schmid.
Origins of the High Stability of an *in vitro*-selected Cold-shock Protein.
The Journal of Molecular Biology **318**, 1341-1349 (2002)

D. Andreas Martin and Franz X. Schmid.
Evolutionary Stabilization of the Gene-3-protein of Phage fd Reveals the Principles that Govern the Thermodynamic Stability of Two-domain Proteins.
The Journal of Molecular Biology **328**, 863-875 (2003)

E. Andreas Martin and Franz X. Schmid.
The Folding Mechanism of a Two-domain Protein: Folding Kinetics and Domain Docking of the Gene-3-protein of Phage fd.
The Journal of Molecular Biology **329**, 599-610 (2003)

F. Andreas Martin and Franz X. Schmid.
A Proline Switch Controls Folding and Domain Interactions in the Gene-3-protein of the Filamentous Phage fd.
The Journal of Molecular Biology **331**, 1131-1140 (2003)

Im Rahmen dieser Arbeit entstand außerdem folgende Publikation, die nicht Teil der Dissertation ist:

G. Maik Jacob, Christophe Saudan, Georg Holtermann, Andreas Martin, Dieter Perl, André Merbach, and Franz X. Schmid.
Water contributes actively to the rapid crossing of a protein unfolding barrier.
The Journal of Molecular Biology **318**, 837-845 (2002)

6 Darstellung des Eigenanteils

A. Die Arbeit beschreibt das von Volker Sieber entwickelte und von mir optimierte Selektionssystem zur gerichteten Evolution von Proteinen. Die Methode wird anhand der von mir durchgeführten Stabilisierungen des Kälteschockproteins *Bs*-CspB vorgestellt. Formuliert wurde die Arbeit von Franz X. Schmid, Volker Sieber und mir.

B. Alle in der Publikation dargestellten Ergebnisse wurden von mir erarbeitet. Zur Selektion der stabilisierten Varianten wurde das von Volker Sieber entwickelte und von mir optimierte Proside-System verwendet. Die Arbeit wurde von Franz X. Schmid und mir formuliert.

C. Die in der Publikation dargestellten Ergebnisse wurden von mir erarbeitet. Insa Kather war im Rahmen eines Mitarbeiterpraktikums mit der Reinigung einzelner Proteinvarianten und der Messung ihrer thermischen Stabilitäten an dieser Arbeit beteiligt. Die Publikation wurde von Franz X. Schmid und mir formuliert.

D. Alle dargestellten Ergebnisse wurden von mir erarbeitet. Die Publikation wurde von Franz X. Schmid und mir formuliert.

E. Alle dargestellten Ergebnisse wurden von mir erarbeitet. Die Publikation wurde von Franz X. Schmid und mir formuliert.

F. Alle dargestellten Ergebnisse wurden von mir erarbeitet. Die Publikation wurde von Franz X. Schmid und mir formuliert.

7 Teilarbeiten

7.1 Teilarbeit A

A

Andreas Martin, Franz X. Schmid, and Volker Sieber.

Proside: A Phage-based Method for Selecting Thermostable Proteins.

In *Methods in Molecular Biology* **230**, *Directed Enzyme Evolution: Screening and Selection Methods* (Arnold, R.H. & Georgiou, G., ed.), Humana Press Inc., Totowa, NJ, 57-70 (2003)

6

Proside

A Phage-Based Method for Selecting Thermostable Proteins

Andreas Martin, Franz X. Schmid, and Volker Sieber

1. Introduction

For many applications proteins must retain their function for a long time and under a wide range of conditions, sometimes at elevated temperatures. However, the stability of natural proteins is often very low, just sufficient to ensure proper functioning under cellular conditions ***(1)***. It is still largely unknown how the stability of a protein is encoded in its sequence, and theoretical approaches to calculate the contributions of individual amino acids to stability are still in their infancy. Therefore, evolutionary methods for increasing protein stability are of great interest. In such approaches, large protein libraries are searched for stabilized variants by a screening or, ideally, a selection technique.

In vivo selection methods for increasing the stability of specific proteins have been successful in cases where the survival of a microorganism could be linked to the function of these proteins ***(2)***. In vitro, proteins with specific binding functions can be selected, e.g., by ribosome display ***(3,4)*** or phage display ***(5)***. Since ligand binding and protein stability are linked thermodynamically, these display techniques can, in principle, also be adapted to select for binders with improved stability ***(6–9)***.

Proside (*pro*tein *s*tability *i*ncreased by *d*irected *e*volution) ***(10)*** is a general method for protein stabilization that does not depend on specific properties of the protein, such as enzymatic activity or ligand binding. This in vitro selection technique links the proteolytic resistance of the protein to be stabilized with a well selectable property, the infectivity of a filamentous phage. Conformational sta-

From: *Methods in Molecular Biology, vol. 230: Directed Enzyme Evolution: Screening and Selection Methods*
Edited by: R. H. Arnold and G. Georgiou © Humana Press Inc., Totowa, NJ

bility and proteolytic resistance are often correlated ***(11)***, since the folded state of a protein is normally much less sensitive to cleavage by proteases than the unfolded state. This holds in particular for proteases such as chymotrypsin with a preference for hydrophobic regions, which are inaccessible in folded proteins.

Infection of *E. coli* by the phage fd is mediated by the gene-3-protein (g3p), which consists of three domains, N1 (68 aa), N2 (131 aa) and CT (150 aa), connected by glycine-rich stretches of 18 and 39 residues, respectively (*see* **Fig. 1A**). The domains of g3p must be tightly linked for the phage to be infective ***(12)***, and this linkage can also be provided by inserted guest proteins ***(13)***. Therefore, in *Proside* a repertoire of sequences coding for the protein to be stabilized is inserted between the N2 and CT domains of g3p, and the resulting pool of phages is subjected to an in vitro proteolysis step (*see* **Fig. 1B**). The tight linkage between these domains and thus phage infectivity is maintained only when the inserted guest protein remains intact. Variants with increased stability and increased resistance towards proteolysis can be enriched in repeated cycles of in vitro proteolysis of the phage library, infection of *E. coli*, and phage propagation. Since the proteolytic selection step is performed in vitro and fd phages are exceptionally stable, the conditions can be varied over a wide range and tailored for the protein to be stabilized. Generally, conditions are used under which the wild type form of the guest protein is partially unfolded, either by adding a denaturant or by increasing the temperature.

The power of the *Proside* method could be demonstrated by the successful selections of stabilized variants of RNaseT1 ***(10)***, the β1-domain of the staphylococcal protein G, and the cold shock protein *Bs*-CspB from the mesophilic bacterium *Bacillus subtilis* ***(14)***. In the case of *Bs*-CspB, six surface exposed positions were randomized by saturation mutagenesis, and many strongly stabilized variants were identified in two selections, one in the presence of the denaturant guanidinium chloride (GdmCl) and the other at elevated temperature. The best variant showed an increase of more than 28°C in the midpoint of the thermal unfolding transition (T_M), and thus surpassed the cold shock protein *Tm*-Csp from the hyperthermophilic organism *Thermotoga maritima* in stability. In addition, *Proside* was used to increase the stability of the host phage protein g3p itself. Random mutagenesis and selection yielded a variant of g3p with an increase in T_M of 13°C for domain N2 and 6°C for domain N1 (Martin et al., manuscript in preparation). The use of this stabilized g3p extends the temperature range for the selection of guest proteins to 60°C.

2. Materials

2.1. Solutions and Media

1. Double-yeast-tryptone medium (dYT): 16 g tryptone, 10 g yeast, 5 g NaCl dissolved in 1 L H_2O; for plates supplemented with 1.5% agar-agar.

Proside *3*

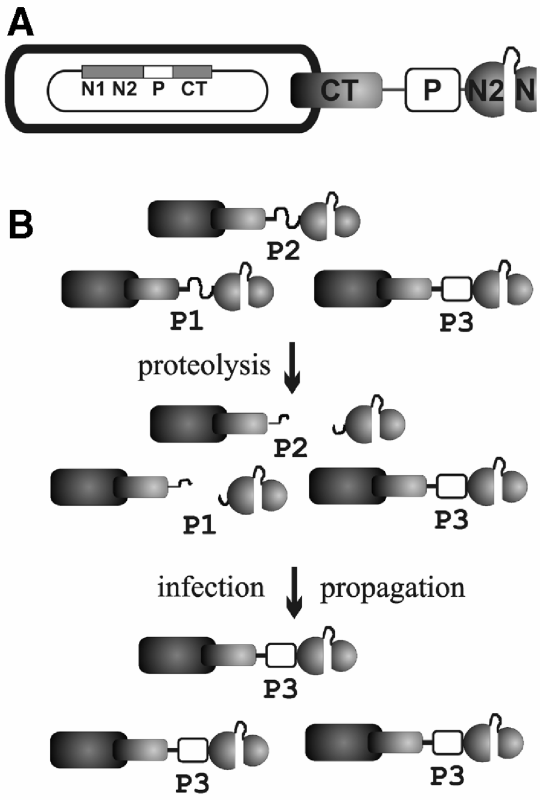


Fig. 1. **(A)** Schematic representation of the filamentous phage fdpro hosting a protein (P) inserted between the C-terminal domain (CT) and the two N-terminal domains (N1,N2) of g3p. **(B)** Outline of the *Proside* selection procedure ***(10)***. Phages hosting different variants of the guest protein (P1, P2, P3) are subjected to an in vitro proteolysis step. Phages with less stable and thus unfolded variants (P1, P2) are cleaved by the protease and lose their infectivity, whereas phages with more stable mutants (P3) remain infectious and can become enriched in repeated selection cycles.

2. Chloramphenicol, 25 mg/mL in ethanol.
3. dYT^{cam} : dYT with 25 µg/mL chloramphenicol.
4. Phage precipitation solution: 30% PEG 8000, 1.5 *M* NaCl, sterilized.
5. Phosphate-buffered saline (PBS)-buffer: 80 g NaCl, 0.2 g KCl, 1.44 g Na_2HPO_4, 0.24 g KH_2PO_4 dissolved in 1 L H_2O, pH 6.9, 0.01% NaN_3, sterilized.
6. Tris-EDTA (TE)-buffer: 10 m*M* Tris-HCl, 1 m*M* Na_2-EDTA, pH 8.0, sterilized.
7. Ethidium bromide, 1% in H_2O.
8. Phenol, pH 7.5, saturated with TE-buffer.
9. Chloroform : isoamyl alcohol, 24:1.
10. 3 *M* sodium acetate, pH 5.2.

11. NAD^+, 10 m*M* in H_2O.
12. ATP, 10 m*M* in H_2O.
13. dNTP mixture, 10 m*M* each dATP, dCTP, dGTP, dTTP.
14. Oligonucleotide primers, 10 μ*M* in H_2O.
15. 100 m*M* potassium phosphate,100 μ*M* $CaCl_2$, pH 8.0.

2.2. Enzymes

1. Restriction endonucleases: *Sfi*I, *Kpn*I, *Hind*III (New England Biolabs, Beverly, MA).
2. T4 DNA Ligase, 20 U/μL (New England Biolabs).
3. T4 Polynucleotide Kinase, 10 U/μL (New England Biolabs).
4. *Vent* DNA polymerase, 2 U/μL (New England Biolabs).
5. *Pfu* DNA polymerase, 3 U/μL (Promega Corporation, Madison, WI).
6. *Taq* DNA polymerase, 5 U/μL (Promega Corporation).
7. *Taq* DNA Ligase, 40 U/μl (New England Biolabs).
8. Chymotrypsin (Roche Diagnostics, Penzberg, Germany).
9. Trypsin (Sigma Chemical Co., St. Louis, MO).

2.3. Miscellaneous

1. Filamentous phage fdpro (send requests to fx.schmid@uni-bayreuth.de).
2. *E. coli* XL1-Blue and *E. coli* ABLE K (Stratagene, La Jolla, CA).
3. Agarose gels, electrophoresis equipment, UV transilluminator, electroporator.
4. DNA molecular weight standards.
5. QIAEX II gel extraction Kit (Qiagen, Hilden, Germany).
6. Flexi Prep kit (Amersham Biosciences, Freiburg, Germany).
7. MF-Millipore membrane filters, pore size 0.025 μm (Millipore GmbH, Schwalbach, Germany).

3. Methods

The method section outlines: 1) how a guest protein is inserted into g3p of the phage fd, 2) how the library of mutants is created, and 3) how stabilized variants of the guest protein are selected from this library. As an example, the selection of stabilized variants of *Bs*-CspB is described.

3.1. Insertion of the Protein to be Stabilized into g3p

Domains N1 and N2 of g3p form a structural and functional entity in which domain CT seems not to be involved ***(15)***. Therefore, foreign proteins can be inserted between N2 and CT (*see* **Fig. 1A**) without significantly changing phage infectivity ***(13)***. It is not known whether there are limits regarding the size of the guest protein or the distance between its N- and C-termini. Disulfide bonds or cysteine residues in the guest protein might interfere with the correct

folding of the g3p and should be eliminated before insertion (*see* **Note 1**). Flexible parts or unstructured linker regions between globular domains could be susceptible to the protease used in the selection step and thus should be cleared of aromatic amino acids (cleavage sites for chymotrypsin) and positively charged residues (cleavage sites for trypsin).

Our modified phage fdpro is derived from fCKCBS ***(13)***, which contains a *cat* gene for the selection of infected cells. The glycine-rich linker between N2 and CT is considerably shortened (*see* **Note 2**) and contains a *Sfi*I, a *Hind*III and a *Kpn*I restriction site for the cloning of the guest protein. The insertion of the protein results in linkers of seven amino acids (PSGAQPA) between N2 and the guest and eight amino acids (ASEA*E*GTP) between the guest and CT, respectively. Twelve base pairs downstream of the C-terminal end of the guest protein (between the *Hind*III and *Kpn*I restriction sites) a TAG amber codon is placed, which is used to eliminate phages coding for the wild type gene by infection of the non-suppressor *E. coli* ABLE K (*see* **Subheading 3.2.**). In the suppressor strain XL1-Blue a Glu residue is incorporated at the amber codon.

Two primers are required for amplifying the guest gene before insertion. The N-terminal primer 5'-TCCGG*GGCCCAGCCGGCC*$(X)_{20}$-3' codes for the *Sfi*I restriction site (italic) and binds to the first 20 nucleotides of the guest gene (X_{20}). The C-terminal primer 5'-CCGG*GGTACC***CTA***AGCTT*CAGAGG $C(X)_{20}$-3' binds to the last 20 bp of the guest gene and contains a *Hind*III restriction site (italic), the amber codon (bold), and a *Kpn*I site (italic) for alternative cloning with *Hind*III or *Kpn*I.

The procedure involves the following steps:

1. Amplify the gene to be inserted in a 100 µL PCR reaction, containing 10 µL Vent buffer, 1 µL 100 m*M* $MgSO_4$, 4 µL of each primer, 3 µL dNTPs, 10 ng template and 1 µL *Vent* DNA polymerase. Run an appropriate PCR program (*see* **Note 3**).
2. Separate the amplified gene using an agarose gel of suitable pore size (e.g., 1.6%).
3. After ethidium bromide staining, excise the bands exhibiting the right size and purify the DNA from the agarose using the QIAEX II gel extraction kit.
4. Determine the concentration of DNA.
5. Purify the double-stranded phage DNA from a 5-mL culture of *E. coli* XL1-Blue infected with the phage fdpro using the Flexi Prep kit.
6. Clone the guest gene into the phage vector via the restriction sites for *Sfi*I and *Hind*III or *Kpn*I, respectively, according to standard protocols ***(16)***.
7. Transform *E. coli* XL1-Blue (*see* **Note 4**) with the ligation product by electroporation and plate the cells on dYT^{cam} agar.
8. Screen the obtained transformants for the right insert by colony-PCR (*see* **Note 5**).
9. Verify the phage variants by sequencing.
10. Grow the positive clones in 5-mL cultures for phage production.
11. Prepare the phages (*see* **Subheading 3.2.1.**).

3.2. Creation of the Phage Library

Having a very large library of different mutants is a prerequisite for a successful search for stabilized variants. Restriction and ligation of double-stranded DNA fragments is usually not efficient enough to obtain libraries with more than 10^5 variants. Therefore, we developed an alternative approach that consistently yields libraries of more than 10^7 variants. We hybridize the mutated DNA fragments coding for the protein to be stabilized to the (+) DNA strand of the phage hosting the corresponding wild type gene between N2 and CT, synthesize the (–) strand with *Pfu* DNA polymerase in the presence of *Taq* ligase, and use the resulting double stranded DNA chimeras for transformation of *E. coli* XL1-Blue (*see* **Note 6**). Next to the wild type gene the template strand contains an amber codon, which is not present in the mutated fragments. Thus, phages with the wild type protein (coded by the original template strand) are eliminated simply by infection of the non-suppressor strain *E. coli* ABLE K with the phages isolated from the suppressor strain XL1-Blue (*see* **Note 7**). The (–) strand is synthesized in the presence of *Taq* ligase to close the nick that is left when the strand is completed. This covalent closure of the mutated strand reduces the loss of mutations by mismatch repair in *E. coli* (*see* **Note 8**). Compared to Quick Change techniques using double-stranded plasmid DNA as a template, this fill-up reaction with single-stranded DNA is much more efficient. The library size is mainly limited by the transformation efficiency of *E. coli* XL1-Blue and ranges between 10^7–10^8 variants.

3.2.1. Preparation of Filamentous Phages

The phages are prepared from the *E. coli* XL1-Blue culture in the stationary phase by polyethylene glycol precipitation ***(16)*** (*see* **Note 9**).

1. Centrifuge 1.5 mL of the XL1-Blue culture at 13,000 g to pellet the cells, mix 1.3 mL of the supernatant with 250 µL phage precipitation solution and incubate for 10 min at room temperature.
2. Pellet the precipitated phages at 13,000 g for 10 min and remove the supernatant carefully.
3. Resuspend the phages in 50 µL PBS-buffer for storage and infection of *E. coli* or TE-buffer for preparation of single-stranded DNA (*see* **Subheading 3.2.2.**) (*see* **Note 10**). Phage solutions can be stored at 4°C for months without a loss in infectivity or can be frozen at –20°C.

The titers of the prepared phage solutions should be about 10^{10}–10^{11} cfu/mL (for determination of phage titers *see* **Subheading 3.3.1.**).

3.2.2. Preparation of Single-stranded Phage DNA

Single-stranded phage DNA is prepared on a small scale according to a standard protocol ***(16)***:

1. Mix 100 µL of TE-buffered phage solution with 50 µL TE-saturated phenol, vortex for 30 s, incubate for 1 min at room temperature and vortex again for 30 s.
2. After centrifugation at 13,000*g* for 1 min transfer the supernatant into a fresh tube, mix it with 50 µL chloroform/isoamyl alcohol and vortex twice for 30 s.
3. After centrifugation mix the supernatant with 11 µL 3 *M* sodium acetate, pH 5.2, and 289 µL ice-cold ethanol. Precipitate the DNA at room temperature for at least 15 min and centrifuge for 10 min at 13,000*g*.
4. Wash the pellet with 200 µL 70% ethanol and dry the DNA.
5. Redissolve the air-dried DNA pellet in 20-30 µL water.

The yield should be about 2–5 µg per 100 µL phage solution.

3.2.3. Synthesis of Mutated Gene Fragments

Depending on the desired number and distribution of mutations within the gene, the fragments are synthesized by PCR using degenerate oligonucleotides or by error-prone PCR. These fragments must be extended to the g3p linker downstream of the inserted gene to abolish the amber codon present in the phage DNA with wild type insert. The primer for this extension 5'-CCGGGTACC**AGA**AGCTTCAGAGGC-3' replaces the amber codon by a Ser codon (bold) and ends right at the C-terminus of the guest gene. For introduction into the phage genome the extended gene fragments have to be phosphorylated.

1. Extend the mutated fragments in a 100 µL PCR reaction containing 10 µL *Vent* buffer, 4 µL of the linker-primer, about 400 n*M* of mutated fragment, 3 µL dNTPs, 10 ng template (single-stranded DNA of the phage containing the wild type insert) and 1 µL *Vent* DNA polymerase. Run an appropriate PCR program (*see* **Note 3**). Depending on the length of the mutated fragment it is advisable to use a temperature ramp for the annealing step, ending at about 50°C.
2. Purify the PCR products by agarose gel electrophoresis and extraction from the gel using the QIAEX II gel extraction kit. Elute the DNA with 25 µL dH_2O.
3. The phosphorylation reaction contains 24 µL purified gene fragment (about 2–4 µg), 3 µL polynucleotide kinase buffer, 3 µL ATP and 1 µL T4 polynucleotide kinase. Incubate for 2 h at 37°C.

3.2.4. Introduction of Mutated Genes into the Phage Genome

The mutated fragments are hybridized to the (+) DNA strand of the phage hosting the wild type gene, the (–) strand is synthesized by *Pfu* polymerase and the nick is closed by the thermostable *Taq* ligase.

1. The reaction mixture contains 4 µL *Pfu* buffer, 2 µL dNTPs, 4 µL NAD^+, about 2 µg (+) strand phage DNA template (*see* **Subheading 3.2.2.**), 200–400 ng mutated gene fragment, 0.5 µL *Taq* DNA ligase, 1 µL *Pfu* DNA polymerase and H_2O to bring the total volume to 40 µL.
2. The program for the extension consists of 1 min denaturation at 95°C, 1 min hybridization at 65°C, and 1 h polymerization and ligation at 68°C.

3. Check the success of the extension reaction by loading 3 μL of the sample before and after the reaction onto an agarose gel. The bands of single-stranded DNA should ideally be doubled in their intensity. They are mostly shifted from a position of apparent migration (19 kb) to the position of double stranded linear phage DNA (7.6 kb).
4. Desalt the sample by microdialysis on MF-Millipore membrane filters, and use 10 μL for the transformation of 40 μL electrocompetent *E. coli* XL1-Blue.
5. Add 700 μL dYT medium right after the electroporation pulse and incubate the cells for 30 min at 37°C.
6. To determine the number of transformants, prepare a dilution series of 1:10 steps (10 μL diluted in 90 μL dYT) in a microtiter plate and pipet 2 μL of each dilution step onto a dYT^{cam} plate. The total number of transformants should be in the range of 10^7–10^8.
7. Pellet the transformed cells and plate them on dYT^{cam}. After growing at 37°C for 10–12 h resuspend the colonies on the plate in 5 mL dYT^{cam} and shake the culture for further phage production at 37°C for 2–3 h.
8. Prepare the phages by polyethylene glycol precipitation (*see* **Subheading 3.2.1.**).
9. To check the titer of mutated phages (without the amber codon) mix 495 μL of an *E. coli* ABLE K culture (OD 0.8) with 5 μL phage solution, incubate for 30 min at 37°C and prepare a 1:10 dilution series as described above. Only the cells infected with phages without the amber codon will grow normally on dYT^{cam} (*see* **Note 7**).
10. Infect *E. coli* ABLE K with the phage library by adding 50 μL of phage solution to 5 mL of a culture with OD 0.8 (*see* **Note 11**).
11. After shaking for 30 min at 37°C add chloramphenicol to a final concentration of 25 μg/mL and incubate the culture at 37°C for at least 7 h for phage production.
12. Prepare the phages by polyethylene glycol precipitation.

3.3. Proside Selection of Stabilized Variants

Filamentous phages like fd show a very high resistance towards proteolysis by site-specific proteases such as trypsin, chymotrypsin or GluC ***(17–19)***. The wild type phage fd remains fully infective after incubation for 1 h with 2.5 μ*M* chymotrypsin or trypsin at 37°C, pH 8.0 or with 1 μ*M* pepsin at 15°C, pH 4.0 ***(10)***, and also after incubation for 30 min with 0.25 μ*M* chymotrypsin at 20°C in 1.75 *M* GdmCl, pH 8.0 ***(14)***. Our improved phage fdproS with the stabilized g3p retains its infectivity even after incubation for 15 min at 57.5°C in the presence of 25 μ*M* chymotrypsin. Thus, proteases can specifically degrade proteins that are inserted into g3p without cleavage of the phage proteins.

To obtain optimal discrimination between folded and unfolded forms of the inserted protein, the type and concentration of the protease, the temperature, the solvent conditions, and the duration of the proteolysis step can be varied over a wide range, especially when using the phage with stabilized g3p. Moreover, the very slow unfolding kinetics of g3p enables a selection of the guest

proteins at even stronger denaturing conditions when only short pulses of denaturation and proteolysis are used.

For every guest protein appropriate selection conditions must be determined in a preliminary screen. The selection pressure should not be too strong so that small differences in stability can be exploited in the selection. In addition, early in a selection many individual variants are poorly populated in the library and therefore, under too-stringent conditions some of them might be lost, even if they are highly stable. The partial unfolding of the guest protein can be achieved either by denaturants, such as GdmCl or urea, or by increased temperature. We prefer to select at elevated temperature, because temperature is easy to adjust and salt effects on stability (as in GdmCl) are avoided.

3.3.1. Finding Optimal Conditions for the Proteolytic Step

The equilibrium unfolding transition of the wild type form of the guest protein provides a valuable reference for finding suitable selection conditions, although the protease in the *Proside* selection shifts the equilibrium to the unfolded state. Often, a temperature or a denaturant concentration near the midpoint of the unfolding transition are good conditions for the proteolysis step. To determine the stability of phages under specified selection conditions, the retained infectivity of the phages after incubation in the presence of the protease is compared with the infectivity of a control sample without protease.

1. For the proteolysis sample mix 17.5 µL 100 m*M* potassium phosphate, 100 µ*M* $CaCl_2$, pH 8.0, with 5 µL phage solution, and for the control sample mix 20 µL 100 m*M* potassium phosphate, 100 µ*M* $CaCl_2$, pH 8.0, with 5 µL phage solution.
2. Incubate the sample and the control at an appropriate temperature for several minutes to establish the folding equilibrium (the incubation time depends on the unfolding kinetics of the guest protein).
3. Add 2.5 µL protease (e.g., 2.5 µ*M* chymotrypsin) to the proteolysis sample and incubate the sample and the control for 15 min.
4. Dilute 5 µL of the proteolysis sample and the control into 495 µL of an *E. coli* XL1-Blue culture (OD 0.8) and incubate at 37°C for 30 min.
5. Prepare dilution series of 1:10 steps, pipet 2 µL of each dilution on a dYT^{cam} plate and determine the number of colonies after growing for 10–12 h.

For a successful selection of stabilized variants the conditions should be optimized such that the retained infectivity of phages with the wild-type protein inserted lies between 1% and 0.01%.

3.3.2. Selection of the Phage Library

The selection procedure consists of repeated cycles of proteolysis, infection of *E. coli* XL1-Blue, phage propagation and preparation. The number of cycles that is necessary to enrich the most stable variants depends strongly on the library size, the applied selection pressure and the stability of variants. The

selection pressure should be increased when, after several cycles, the retained infectivity of the phage pool approaches 50–100%.

1. Mix 35 μL 100 m*M* potassium phosphate, 100 μ*M* $CaCl_2$, pH 8.0, with 10 μL phage solution (about 10^8–10^9 phage particles).
2. Perform the selection step under the conditions identified in **Subheading 3.3.1.**
3. Use the whole sample to infect a 5 mL culture of *E. coli* XL1-Blue (OD 0.8).
4. After incubation for 30 min at 37°C add chloramphenicol to a final concentration of 25 μg/mL and shake for at least 7 h for phage production (*see* **Note 12**).
5. Prepare the phages as described above and use them for the next selection cycle.
6. For every selection cycle prepare in addition two samples of 25 μL, one with and the other without the protease (*see* **Subheading 3.3.1.**) to determine the retained infectivity of phages in the library and to follow the development of the selection.
7. Check the library for the presence of recombinants with shortened inserts after every second or third selection cycle or when the retained infectivity increases dramatically. Determine the insert size of 10–20 isolated clones by colony-PCR (*see* **Note 5**). If recombinants are present, prepare the single stranded DNA from the whole library (*see* **Subheading 3.2.1.**) and use it as a template for PCR amplification of the guest inserts with the two linker primers described in **Note 5**, for PCR conditions see *see* **Subheading 3.1.** Isolate the full length inserts by agarose gel electrophoresis and extraction from the gel, and re-introduce the fragments into the phage genome as described in **Subheading 3.2.4**.
8. If there is no further increase in phage infectivity after proteolysis isolate individual phage clones by plating an appropriate dilution of the infected cells on dYT^{cam}, pick single colonies, and grow the bacteria for phage production.
9. Screen the isolated phages for those with the highest proteolytic stability and determine the sequences of their guest proteins.
10. Clone the variants into a suitable expression vector.

3.4. Example: Stabilization of Bs-CspB (14)

To explore the role of six surface exposed positions (2, 3, 46, 64, 66, 67) for the stability of *Bs*-CspB we constructed a library of variants by using PCR and degenerate oligonucleotides allowing all 20 amino acids at the six sites. The first selection of this library was for stability towards GdmCl-induced unfolding. The phages were incubated in 1.5 *M* GdmCl, 100 m*M* potassium phosphate, pH 8.0, at 25°C (which corresponds to the conditions at the transition midpoint of wild type *Bs*-CspB), and selected by exposing them to 0.25 μ*M* chymotrypsin for 30 min. Only 0.006% of the phages with the wild type *Bs*-CspB insert remained infectious under these conditions. The course of the selection that consisted of six cycles is shown in **Fig. 2**. After the third round the resistance towards proteolysis had already reached 65 %, and therefore the denaturant concentration was increased to 1.75 *M* GdmCl. After the sixth round five different stabilized variants of *Bs*-CspB were identified; they were

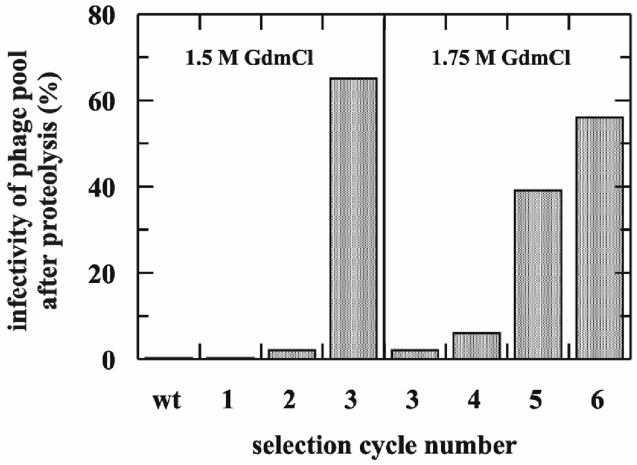


Fig. 2. Selection cycles of the *Bs*-CspB library in the presence of GdmCl ***(14)***. The relative infectivities of the phage pool after the various proteolytic selection cycles are shown. Selection rounds 1–3 were performed in 1.5 *M* GdmCl with 0.25 μ*M* chymotrypsin. Under these conditions phages hosting the wild type *Bs*-CspB (wt) retained 0.006% of their infectivity. For rounds 4–6 the GdmCl concentration was increased to 1.75 *M*. The infectivities after round 3 are shown for both 1.5 and 1.75 *M* GdmCl.

overexpressed and characterized. All of them were much more stable than the wild type protein (T_M = 54°C) with increases in T_M between 13 and 22°C.

The second selection was performed at elevated temperature. In this case the library of *Bs*-CspB was introduced into the phage fdproS with stabilized g3p. In the first six rounds of the selection the phages were exposed to 2.5 μ*M* chymotrypsin at 57.5°C for 15 min. The infectivity of the phage library increased during the first 4 cycles from 3 to 28% and then remained approximately constant (*see* **Fig. 3**). Therefore, we used the library present after the fourth round and enhanced the selection pressure by increasing the chymotrypsin concentration ten-fold to 25 μ*M* and by combining it in tandem with a 15 min proteolysis with 0.2 μ*M* trypsin. At 57.5°C chymotrypsin becomes markedly inactivated within 10 min. Therefore, for selection rounds eight to ten, 25 μ*M* chymotrypsin was added twice for 5 min each. The six variants of *Bs*-CspB that were chosen from the final pool for purification and characterization were all strongly stabilized. They showed T_M values that were 21.9–28.2 degrees higher than the T_M value of the wild type protein.

4. Notes

1. A protein pre-destabilization by eliminating disulfide bonds (e.g., by Cys to Ala mutations) facilitates the selection for higher stability, since much milder condi-

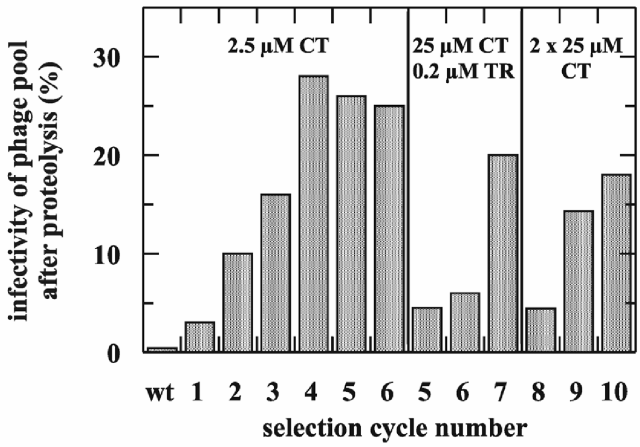


Fig. 3. Selection cycles of the *Bs*-CspB library at elevated temperature ***(14)***. The infectivities of the phage pool after the proteolytic selections at 57.5°C are shown. The first six cycles were performed with 2.5 µ*M* chymotrypsin (CT) for 15 minutes. Phages with wild type insert (wt) retain 0.4% of their infectivity under these conditions. Rounds 5–7 were performed with tandem 15 min proteolyses by 25 µ*M* chymotrypsin and by 0.2 µ*M* trypsin (TR), followed by three rounds with 25 µ*M* chymotrypsin added twice for 5 min each.

tions are sufficient for a comparative selection pressure on the guest protein, and the phage proteins as well as the proteases are less strained (e.g., stabilization of RNase T1 ***[10]***).

2. The linker between N2 and CT was shortened to 15 amino acids to reduce the frequency with which inserted guest sequences are eliminated by recombination. Nevertheless, if particular protein inserts impair phage infectivity (e.g., by non-specific interactions with the domains of g3p) the linker can be again extended to 58 amino acids.
3. The following PCR program gives good yields when amplifying a fragment of 500 bp with *Vent* polymerase: 30 cycles of denaturation at 95°C for 20 s, annealing at 55°C for 20 s and extension at 72°C for 30 s. In the case of annealing problems add 1–2 m*M* $MgSO_4$ or supplement the reaction mixture with 3 U *Taq* DNA polymerase.
4. *E. coli* XL1-Blue carries the *supE44* gene, which allows suppression of amber codons by introduction of a Glu.
5. For colony-PCR prepare a mixture of 10 µL *Taq* buffer, 10 µL 10 m*M* $MgCl_2$, 2 µL dNTPs, 2 µL upstream-linker primer 5'-CCGTCCGGGGCCCAGCCGGCC-3', 2 µL downstream-linker primer 5'-CCGGGTACCAGAAGCTTCAGAGGC-3', 0.5 µL *Taq* DNA polymerase and H_2O to a final volume of 100 µL. Aliquot this mixture to 10 µL per PCR tube and transfer cells from the colonies on the plate with a pipetting tip into the PCR tube. Inoculate a second tube containing

50 µL dYT medium with the same tip to have a small pre-culture available in the case of a positive PCR result. Run a PCR program as described in **Note 3**.

6. A digestion of remaining single-stranded DNA template is not necessary, since it has a very low transformation efficiency and the phages that are encoded by it are eliminated by subsequent infection of the non-suppressor strain *E. coli* ABLE K.
7. In ABLE K the translation stop at the amber codon results in a g3p without the CT domain. Thus, the phages lack g3p, they are not infective and not released from the cells due to defects in the phage assembly ***(20)***. The cells grow very slowly or stop growing at an early stage.
8. The methylation of the synthesized DNA by dam methylase had no effect on the mismatch repair of transfected DNA chimeras in *E. coli* XL1-Blue.
9. We prefer to prepare phages on a small scale (5 mL cultures, precipitation in 1.5 mL), since the quality and reactivity of the single-stranded DNA prepared from these phage solutions is significantly higher than with large-scale preparations (100–500 mL).
10. The phages are more stable when stored for longer times in PBS buffer, but for subsequent preparation of DNA we resuspend the phages in TE buffer to reduce the amount of salt in the DNA sample.
11. In *E. coli* ABLE K the *recA* gene (responsible for homologous recombination) is not knocked out as it is in *E. coli* XL1-Blue. Thus, if the phage construct is susceptible to a complete or partial loss of the inserted guest gene by recombination, the probability for such loss is considerably increased when propagating the phages in the ABLE K strain. We use the infection of ABLE K to reduce the high number of phages with wild type guest protein in the library and to facilitate the enrichment of stabilized, but rare variants. The selection itself also functions without this step, and it can be skipped to prevent recombination. Before the first selection it is, however, necessary to do a infection with the phages prepared from the transformants, because DNA chimeras were transfected and the produced phages can differ in their pheno- and genotype for the guest protein (i.e., phages can present a mutant of the protein on their surface but still contain the gene of the wild type protein and vice versa).
12. Very low phage titers could give problems with the phage propagation after infection of 5 mL liquid cultures. In these cases pellet the cells from the liquid culture after infection and incubation for 30 min at 37°C, and plate them on a dYT^{cam} plate. After growing for 10–12 h isolate the cells from the plate with 5 ml dYT^{cam} and shake the culture for 2–3 h for further phage production.

References

1. Wintrode, P. L. and Arnold, F. H. (2001) Temperature adaptation of enzymes: lessons from laboratory evolution, in *Advances in Protein Chemistry*, **55,** Academic Press, San Diego, pp. 161–225.
2. Matsumura, M. and Aiba, S. (1985) Screening for thermostable mutant of kanamycin nucleotidyltransferase by the use of a transformation system for a thermophile, *Bacillus stearothermophilus*. *J. Biol. Chem.* **260,** 15,298–15,303.

3. Hanes, J. and Plückthun, A. (1997) In vitro selection and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci. USA* **94,** 4937–4942.
4. Plückthun, A., Schaffitzel, C., Hanes, J., and Jermutus, L. (2001) In vitro selection and evolution of proteins, in *Advances in Protein Chemistry*, **55,** Academic Press, San Diego, pp. 1367–1403.
5. Dunn, I. S. (1996) Phage display of proteins. *Curr. Opin. Biotechnol.* **7,** 547–553.
6. Jung, S. and Plückthun, A. (1997) Improving in vivo folding and stability of a single-chain Fv antibody fragment by loop grafting. *Protein Eng.* **10,** 959–966.
7. Proba, K., Worn, A., Honneger, A., and Plückthun, A. (1998) Antibody scFv fragments without disulfide bonds made by molecular evolution. *J. Mol. Biol.* **275,** 245–253.
8. Jackson, J. R., Sathe, G., Rosenberg, M., and Sweet, R. (1995) In vitro antibody maturation. Improvement of a high affinity, neutralizing antibody against IL-1 beta. *J. Immunol.* **154,** 3310–3319.
9. Spada, S., Honneger, A., and Plückthun, A. (1998) Reproducing the natural evolution of protein structural features with the selectively infective phage (SIP) technology. *J. Mol. Biol.* **283,** 395–407.
10. Sieber, V., Plückthun, A., and Schmid, F. X. (1998) Selecting proteins with improved stability by a phage-based method. *Nat. Biotechnol.* **16,** 955–960.
11. Parsell, D. and Sauer, R. (1989) The structural stability of a protein is an important determinant of its proteolytic susceptibility in *Escherichia coli*. *J. Biol. Chem.* **264,** 7590–7595.
12. Stengele, I., Bross, P., Garces, X., Giray, J., and Rasched, I. (1990) Dissection of functional domains in phage fd adsorption protein. Discrimination between attachment and penetration sites. *J. Mol. Biol.* **212,** 143–149.
13. Krebber, C., Spada, S., Desplancq, D., Krebber, A., Ge, L., and Plückthun, A. (1997) Selectively infective phage (SIP): a mechanistic dissection of a novel in vivo selection for protein-ligand interactions. *J. Mol. Biol.* **268,** 607–618.
14. Martin, A., Sieber, V., and Schmid, F. X. (2001) In-vitro selection of highly stabilized protein variants with optimized surface. *J. Mol. Biol.* **309,** 717–726.
15. Holliger, P., Riechmann, L., and Williams, R. L. (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,** 649–657.
16. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning-a laboratory manual*. Cold Spring Harbor Laboratory press, Cold Spring Harbor, NY.
17. Schwind, P., Kramer, H., Kremser, A., Ramsberger, U., and Rasched, I. (1992) Subtilisin removes the surface layer of the phage fd coat. *Eur. J. Biochem.* **210,** 431–436.
18. Kremser, A. and Rasched, I. (1994) The adsorption protein of filamentous phage fd: assignment of its disulfide bridges and identification of the domain incorporated in the coat. *Biochemistry* **33,** 13,954–13,958.
19. Salivar, W. O., Tzagoloff, H., and Pratt, D. (1964) Some physical-chemical and biological properties of the rod-shaped coliphage M13. *Virology* **24,** 359–371.
20. Rakonjac, J. and Model, P. (1998) Roles of pIII in filamentous phage assembly. *J. Mol. Biol.* **282,** 25–41.

7.2 Teilarbeit B

B

Andreas Martin, Volker Sieber and Franz X. Schmid.

In-vitro Selection of Highly Stabilized Protein Variants with Optimized Surface.

The Journal of Molecular Biology **309**, 717-726 (2001)

doi:10.1006/jmbi.2001.4698 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2001) **309**, 717–726

JMB

In-vitro Selection of Highly Stabilized Protein Variants with Optimized Surface

Andreas Martin, Volker Sieber and Franz X. Schmid*

Laboratorium für Biochemie, Universität Bayreuth D-95440 Bayreuth, Germany

*Corresponding author

Thermostable proteins are of prime importance in protein science, but it has remained difficult to develop general strategies for stabilizing a protein. Site-directed mutagenesis based on comparisons with thermophilic homologs is rarely successful because the sequence differences are too numerous and dominated by neutral mutations. Here we used a method of directed evolution to increase the stability of a mesophilic protein, the cold shock protein *Bs*-CspB from *Bacillus subtilis*. It differs from its thermophilic counterpart *Bc*-Csp from *Bacillus caldolyticus* at 12 surface-exposed positions. To elucidate the stabilizing potential of exposed amino acid residues, six of these variant positions were randomized by saturation mutagenesis, the corresponding library of sequences was inserted into the gene-3-protein of the filamentous phage fd, and stabilized variants were selected by the Proside technique. Proside links the increased protease resistance of stabilized protein variants with the infectivity of the phage. Many strongly stabilized variants of *Bs*-CspB were identified in two selections, one in the presence of a denaturant and the other at elevated temperature. Several of them are significantly more stable than the naturally thermostable homolog *Bc*-Csp, and the best variant reaches *Tm*-Csp (the homolog from the hyperthermophile *Thermotoga maritima*) in stability. Remarkably, this variant differs from *Tm*-Csp at five and from *Bc*-Csp at all six randomized positions. This indicates that proteins can be strongly stabilized by many different sets of surface mutations, and Proside selects them efficiently from large libraries. The course of the selection could be directed by the conditions. In an ionic denaturant non-polar surface interactions were optimized, whereas at elevated temperature variants with improved electrostatics were selected, pointing to two different strategies for stabilization at protein surfaces.

Keywords: cold shock protein; protein stability; *in vitro* selection; phage display; electrostatic interactions

Introduction

It is still largely unknown how the stability of a protein is encoded in its sequence and how individual amino acid residues contribute to stability. Long time ago Perutz and co-workers,[1] and, more recently, others[2–5] suggested that favorable ionic interactions of residues at their surfaces can be important for the high stability of thermophilic proteins. Stabilizing of proteins by engineered salt bridges, however, failed in many cases,[6–8] mutational analyses gave ambiguous answers,[9–17] and theoretical calculations[18] suggested that salt bridges destabilize proteins, because the desolvation penalty is higher than the energy that is gained by the electrostatic attraction of opposite charges. Stabilizing proteins by repacking their hydrophobic cores is difficult[19,20] and was successful only for proteins with obvious packing defects,[21] buried charged residues,[22] or cavities that could be filled by amino acids with enlarged side-chains.[23]

Abbreviations used: *Bs*-CspB, *Bc*-Csp, and *Tm*-Csp, cold shock proteins from *Bacillus subtilis*, *Bacillus caldolyticus*, and *Thermotoga maritima*, respectively; CT, bovine α-chymotrypsin; g3p, gene-3-protein of fd phage; GdmCl, guanidinium chloride; t_M, midpoint of a thermal unfolding transition. Proteins from mesophilic and thermophilic organisms are denoted in a short, but slightly incorrect, manner as mesophilic and thermophilic proteins, respectively.

E-mail address of the corresponding author: fx.schmid@uni-bayreuth.de

0022-2836/01/030717–10 $35.00/0

Surface residues can be important for stability, and certain proteins could be stabilized by automated protein design that targeted uncharged residues near the surface,[24] or by screening for second-site suppressors of destabilizing mutations.[25] Electrostatic calculations based on Coulomb's law in combination with experimental work demonstrated that protein stability increased when the distribution of surface charges was changed.[26–29]

Proteins from thermophilic organisms often show increased numbers of charged residues at their surfaces,[30] which is easiest, but not necessarily, explained by assuming that these residues engage in additional ionic interactions. Unfortunately, mesophilic and thermophilic proteins differ usually at many positions, and the determinants that guide the evolution of the protein surface are largely unknown. In addition to stability they include properties such as solubility, molecular interactions, and regulated turnover in the thermophilic organism, as well as neutral genetic drift. Therefore it remains almost impossible to identify those differences that convey extra stability to a thermophilic protein by sequence comparisons only.[30,31]

The thermophilic cold shock protein *Bc*-Csp from *Bacillus caldolyticus* and its mesophilic homolog *Bs*-CspB from *Bacillus subtilis* provide a good example. They differ in sequence at 12 surface positions, but in a systematic mutagenesis study only two residues, Arg3 and Leu66 of *Bc*-Csp (which replace Glu3 and Glu66 of *Bs*-CspB) turned out to be responsible for the difference in stability of 15.8 kJ/mol between the two proteins.[32] In the mesophilic protein the Glu3Arg replacement alone increased stability by 11.5 kJ/mol, and the Glu3Arg and Glu66Leu mutations together stabilized *Bs*-CspB by 14.2 kJ/mol.

This pairwise comparison identified two particular residues of *Bc*-Csp that are important for its stability, but the general potential of these and other surface positions for stabilizing a protein could not be elucidated by site-directed mutagenesis. Directed evolution[21,33–36] offers the means for a wider exploration of the sequence-stability space, provided that an efficient screening or selection system for stabilized proteins is available. For specific proteins stabilized variants have been identified by high-throughput functional screening[37,38] in combinatorial libraries with random mutations.

Here we used a generally applicable selection method for stabilized proteins[35,36,39] and a library of variants of *Bs*-CspB to explore the potential of six surface-exposed positions for stabilizing this protein. We found many strongly stabilized variants in this library with several of them being even more stable than the naturally thermostable homolog *Bc*-Csp.

Results and Discussion

Selection procedure

To explore the role of six out of the 12 positions that are different between *Bs*-CspB and *Bc*-Csp, we constructed a library of variants of *Bs*-CspB by allowing all 20 amino acids at the positions 2, 3, 46, 64, 66 and 67 (Figure 1(a)). The residues at positions 3 and 66 had been found to be critical for the difference in stability between *Bs*-CspB and *Bc*-Csp,[32] the other four positions are located near 3 and 66 in folded *Bs*-CspB. The size of the library is limited by the transformation efficiency of *Escherichia coli*, therefore we randomized only six codons.

This library was searched for stabilized variants using the Proside (*pro*tein *s*tability *i*ncreased by *d*irected *e*volution) method.[36] Proside links the increased protease resistance of stabilized protein variants with the infectivity of the filamentous phage fd. Infection of *E. coli* by the phage is mediated by the gene-3-protein (g3p), which consists of three domains, N1 (68 aa), N2 (131 aa) and C (150 aa), connected by glycine-rich stretches of 18 and 39 residues, respectively. The domains of

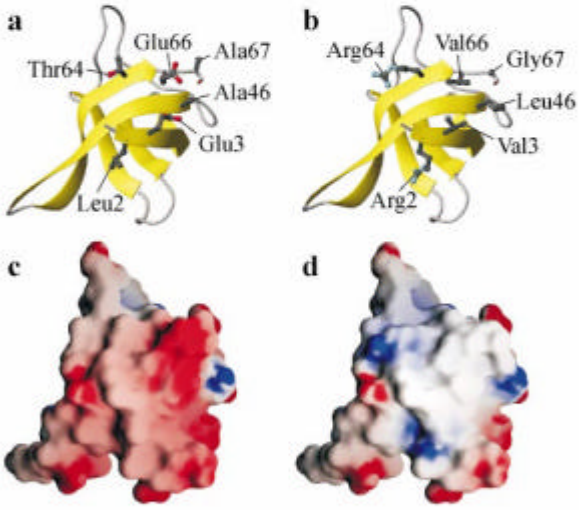


Figure 1. Tertiary structure of *Bs*-CspB. (a) The side-chains of the six residues that were randomized for the selections are shown in ball-and-stick representation (coordinates from ref.[51]). Side-chain nitrogen atoms are colored blue, oxygen atoms red. (b) Structure of the most stable variant (variant T-10/10) after the thermal selection. This structure was modelled by Swiss-pdb viewer[52] based on the structure of the wild-type protein in (a). (c) and (d) Surface charges of wild-type *Bs*-CspB (c) and variant T-10/10 (d), calculated with GRASP.[53] The molecule surfaces were generated using a probe radius of 1.4 Å, and colored using the GRASP electrostatic potential maps, assuming fully charged residues Asp, Glu, Arg and Lys and the terminal carboxy function. Potentials of -15 kT/e or less are shown in red, neutral potential (0 kT/e) is colorless, and potentials of $+15$ kT/e or more are colored blue.

g3p must be tightly linked for the phage to be infective.[40]

In Proside, the repertoire of sequences coding for the protein to be stabilized is inserted between the N2 and C domains of g3p.[36] The tight linkage between these domains is maintained only when the inserted guest protein remains intact (e.g. in the presence of a protease). Variants with increased stability and thus increased resistance towards proteolysis can be enriched in repeated cycles of *in vitro* proteolysis of the phage library, infection of *E. coli*, and phage propagation. Importantly, this selection method is generally applicable, it does not depend on a specific function of the protein to be stabilized.

The proteolytic selection step of Proside is performed *in vitro*, and therefore the conditions can be varied over a wide range, limited only by the stability of the phage itself.[36] Generally, conditions are used under which the wild-type form of the guest protein is partially unfolded.

Selection for stability towards GdmCl-induced unfolding

In the first selection we incubated the phages with the *Bs*-CspB library in 1.5 M GdmCl (pH 8.0), at 25 °C and selected for stabilized variants by exposing them to 0.25 μM bovine α-chymotrypsin (CT) for 30 minutes. Under these conditions only 0.006 % of a reference phage with the wild-type *Bs*-CspB insert survived proteolysis. The course of a selection that consisted of six rounds of proteolysis, propagation of the surviving phages in *E. coli*, and phage purification is shown in Figure 2. Twelve individual clones from the repertoire after the last selection cycle coded for five different variants (Table 1). They were overexpressed in *E. coli*, purified, and their thermal stability determined. All of them were much more stable than wild-type *Bs*-CspB ($t_M = 54.0$ °C) with increases in t_M between 13 and 22 degrees (Figure 3(a)). The most stable variant, which was found in six of the 12 clones, had a t_M of 76.0 °C and was thus almost as stable as the thermophilic protein *Bc*-Csp ($t_M = 76.9$ °C). This variant differed at all six randomized positions from the mesophilic parent *Bs*-CspB, and, more intriguing, it coincided with the thermophilic reference *Bc*-Csp only at Leu66. A leucine was found at position 66 in four out of the five sequences (Table 1), and it was coded by three different triplets, providing strong evidence that this leucine was selected because it increases the stability of the cold shock proteins. In fact, many homologs of *Bs*-CspB have a Leu at position 66.

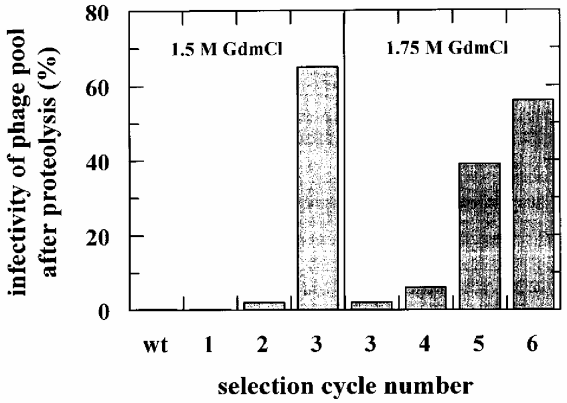


Figure 2. Selection cycles in the presence of GdmCl. The relative infectivities of the phage pool containing the *Bs*-CspB repertoire after the various proteolytic selection cycles are shown. The infectivity in a control incubation without protease is set as 100 %. Selections 1-3 were performed as described in Materials and Methods in 1.5 M GdmCl with 0.25 μM CT. Phages hosting wild-type *Bs*-CspB (wt) retained 0.006 % of their infectivity under these conditions. After the third round the resistance to proteolysis had already reached 65 %, and therefore the denaturant concentration was increased to 1.75 M GdmCl to enhance the selection pressure. The infectivities after round 3 are shown for both 1.5 and 1.75 M GdmCl. After three additional rounds the protease resistance of the phage repertoire had again increased to 56 %.

Table 1. Stabilized variants (G-1 to G-5) of *Bs*-CspB selected after six rounds of Proside in the presence of GdmCl

Protein	Pos. 2	Pos. 3	Pos. 46	Pos 64	Pos. 66	Pos. 67	Sequence frequency	t_M (°C)	$\Delta\Delta G_D$ (kJ/mol)
Bs-CspB	*TTA Leu*	*GAA Glu*	*GCT Ala*	*ACT Thr*	*GAA Glu*	*GCG Ala*	-	54.0	0
G-1	*TTA Leu*	*GAA Glu*	*GCT Ala*	*ACT Thr*	CTG Leu	CCG Pro	1/12	n.d.	n.d.
G-2	ATA Ile	TCC Ser	CAA Gln	CTT Leu	CTC Leu	CCC Pro	1/12	67.1	9.3
G-3	*TTA Leu*	*GAA Glu*	TTA Leu	TAC Tyr	TTG Leu	CAT His	1/12	70.1	11.1
G-4	TAT Tyr	AGA Arg	TTA Leu	*ACT Thr*	*GAA Glu*	*GCG Ala*	3/12	72.5	12.7
G-5	TAT Tyr	ATC Ile	CAA Gln	CTT Leu	CTC Leu	CCC Pro	6/12	76.0	15.1
Bc-Csp	CAA Gln	CGT Arg	GAA Glu	GTC Val	TTA Leu	-	-	76.9	15.6

The selected codons and amino acid residues at the six randomized positions are shown in comparison with the sequences of *Bs*-CspB and *Bc*-Csp, together with the frequency of appearance of every sequence, the melting temperatures (t_M), and the differences in the Gibbs free energy of denaturation at 70 °C ($\Delta\Delta G_D$) relative to wild-type *Bs*-CspB for the purified proteins. The $\Delta\Delta G_D$ values were derived from thermal unfolding transitions (Figure 3(a)) measured by CD at 222.6 nm as described.[48] Coincidences in sequence with wild-type *Bs*-CspB are shown in italic.

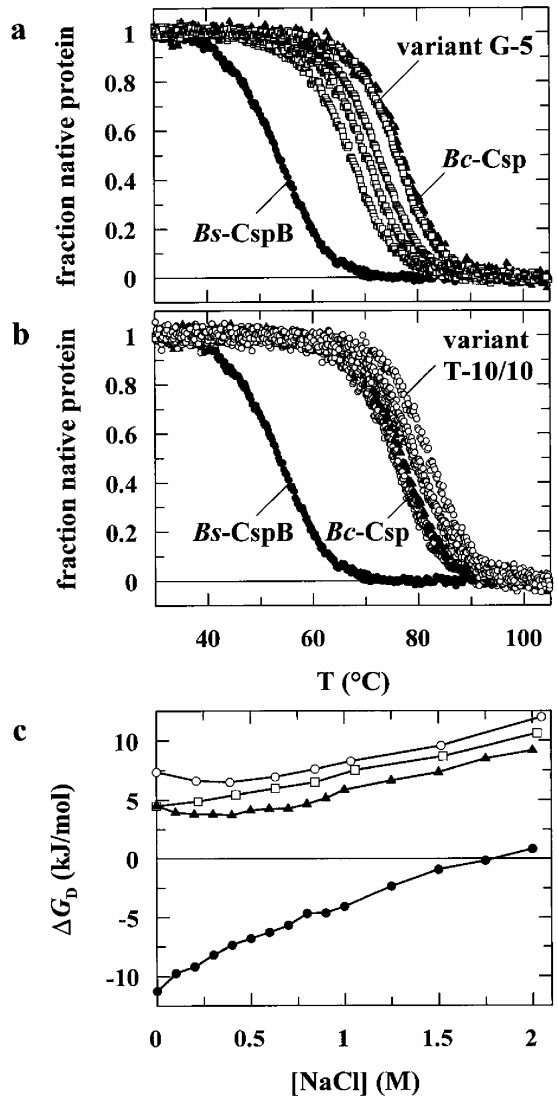


Figure 3. Thermal stability of the selected variants. (a) Variants obtained after the selection in the presence of GdmCl (□). (b) Variants obtained after the thermal selection (○). In (a) and (b) the unfolding transitions of wild-type *Bs*-CspB (●) and *Bc*-Csp (▲) are shown for comparison. The fractions of native protein as obtained after a two-state analysis of the data are shown as a function of temperature. (c) The free energies of denaturation at 70 °C, ΔG_D, of *Bc*-Csp (▲), *Bs*-CspB (●), variant G-5 from the GdmCl-selection (□) and variant T-10/10 from the thermal selection (○) as a function of the NaCl concentration. They were derived from thermal unfolding transitions as shown in (a) and (b).

Very few charged residues were found in this selection. In particular, an Arg at position 3, which is known to stabilize *Bs*-CspB by more than 11 kJ/mol,[32] occurred in only one sequence; in the most stable variant from this selection (G-5, Table 1) an Ile occupies this position. We attribute the lack of charged residues to the presence of moderate concentrations of GdmCl in the selection. This ionic denaturant weakens electrostatic interactions.

None of the selected variants had a Glu66 as in the parent protein *Bs*-CspB, which confirms the conclusion from the directed mutagenesis that Glu66 is destabilizing. Two of the selected variants (G-1 and G-3, Table 1) showed the gene sequence of wild-type *Bs*-CspB at several of the randomized base triplets. They were possibly introduced during construction of the library by recombination with the wild-type gene or by mismatch repair of the transfected heteroduplex DNA in *E. coli*.

Selection for stability at elevated temperature

In the second selection we challenged the phage library by proteolysis at elevated temperature and low ionic strength to avoid screening of potentially stabilizing ionic interactions. We first applied the Proside method to a pool in *E. coli* XL1Red randomly mutated wild-type fd phages without a guest protein to select a variant with an increased proteolytic stability at high temperature. This was necessary, because only 42% of the wild-type phage fd without a guest protein remained infectious after a 15-minute incubation with 0.25 μM CT at 40 °C. The mutated phage found in a selection (at 40-57 °C in the presence of 0.25 μM CT) remained 100% infective after a 15-minute incubation with 0.25 μM CT at 57.5 °C (unpublished results). It was used for the selections at elevated temperature. In the first six rounds of this selection (Figure 4) the phages with the CspB library were exposed to 2.5 μM CT at 57.5 °C for 15 minutes. The survival rate of the phage library increased during the first four rounds to 28% and then remained approximately constant in rounds 5 and

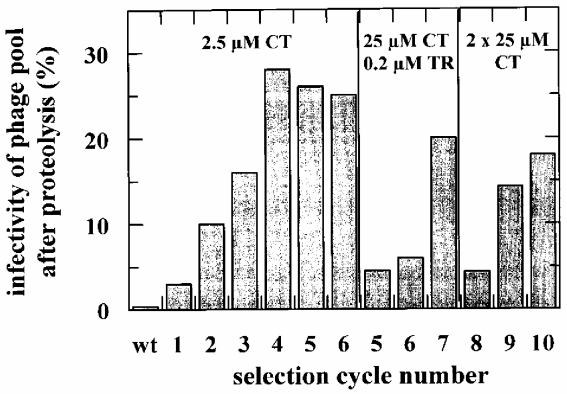


Figure 4. Selection cycles at elevated temperature. The infectivities of the phage pool hosting the *Bs*-CspB library after the proteolytic selections at 57.5 °C are shown. As described in Materials and Methods, the first six cycles of selection were performed with 2.5 μM CT for 15 minutes, followed by three cycles (5-7) of tandem 15-minute proteolyses by 25 μM CT and by 0.2 μM trypsin (TR). In the three final rounds (8-10) 25 μM CT were added twice for five minutes each.

6. Eight individual clones were picked and sequenced after round 6, proteins were purified from five of these and the thermal stabilities determined (Table 2).

As in the GdmCl-selection all the purified variants were much more stable than the mesophilic parent protein (with increases in t_M between 14.7 and 18.6 degrees) but still showed strongly different sequences at the randomized positions. Again, a number of variants showed the sequence of wild-type *Bs*-CspB at several of the randomized positions in the amino-terminal region.

Obviously, the selection had not yet converged. Therefore we took the library present after the fourth round and enhanced the selection pressure by increasing the CT concentration tenfold to 25 μM and by combining it in tandem with a 15 minute proteolysis with trypsin. The additional proteolysis with trypsin not only increased the selection pressure, but also eliminated recombinant phages with shortened CspB inserts. These recombinants contained only small parts of the CspB guest protein, which lacked most of the cleavage sites for chymotrypsin and thus were highly resistant towards this protease. They were, however, susceptible to trypsin. After three rounds under these conditions (Figure 4) four clones were picked and sequenced. The corresponding purified proteins had t_M values 10-20.7 deg. C higher than *Bs*-CspB and still variable residues at the randomized positions (except an Arg64 that was found in all four proteins).

At 57.5 °C CT becomes markedly inactivated within ten minutes and therefore, in selection rounds 8-10, we added 25 μM CT twice for five minutes each. Then 30 clones were analyzed. They showed ten different sequences (Table 2), with five of them being found several times. In this final pool all ten sequences showed a positively charged residue (Arg or Lys) at position 64. Interestingly, the cold shock proteins from the extremophiles *Thermotoga maritima* and *Aquifex aeolicus* also have positively charged residues at this position. At position 2 a Lys or Arg was present in eight sequences. A variety of residues is found at 66, but in eight of the ten sequences they have aliphatic side-chains (Leu, Ile, Val).

The preferences at the remaining three positions are less clear. At position 46 six of the variants carry a Leu, which is coded by three different triplets. Interestingly, a Leu had not been found at position 46 after the selection rounds 6 or 7 (Table 2) and is not present in either parent protein. Seven different amino acids are found at position 3 with a slight preference for Ile, which is present in three variants. An Arg3 is obtained only once, despite the fact that Arg3 is strongly stabilizing when introduced as a single mutation into

Table 2. Stabilized variants of *Bs*-CspB selected at elevated temperature

Protein	Pos. 2	Pos. 3	Pos. 46	Pos 64	Pos. 66	Pos. 67	Sequence frequency	t_M (°C)	$\Delta\Delta G_D$ (kJ/mol)
Bs-CspB	*TTA Leu*	GAA Glu	*GCT Ala*	*ACT Thr*	*GAA Glu*	*GCG Ala*	-	54.0	0
T-6/1	*TTA Leu*	*GAA Glu*	TCT Ser	AGA Arg	ACA Thr	GAT Asp	1/8	n.d.	n.d.
T-6/2	*TTA Leu*	*GAA Glu*	CCC Pro	AAA Lys	CTC Leu	CCC Pro	1/8	n.d.	n.d.
T-6/3	*TTA Leu*	*GAA Glu*	ACG Thr	AGG Arg	TTG Leu	ACA Thr	1/8	n.d.	n.d.
T-6/4	CAT His	CTA Leu	CAG Gln	ATC Ile	CAC His	ACC Thr	1/8	68.7	10.2
T-6/5	*TTA Leu*	*GAA Glu*	ATT Ile	ATT Ile	CGA Arg	CCG Pro	1/8	71.7	12.2
T-6/6	GTA Val	AAA Lys	AAT Asn	AAA Lys	TTA Leu	AAT Asn	1/8	71.7	12.2
T-6/7	*TTA Leu*	*GAA Glu*	AGA Arg	ACA Thr	CTC Leu	CAG Gln	1/8	72.2	12.8
T-6/8	AAG Lys	CTC Leu	CGA Arg	AAA Lys	CTA Leu	CGG Arg	1/8	72.6	12.7
T-7/1	GGA Gly	AAG Lys	TTC Phe	CGC Arg	ACT Thr	TCC Ser	1/4	64.0	7.6
T-7/2	*TTA Leu*	*GAA Glu*	*GCT Ala*	CGT Arg	ATT Ile	GAA Glu	1/4	71.1	11.9
T-7/3	*TTA Leu*	*GAA Glu*	*GCT Ala*	CGC Arg	ATA Ile	CCC Pro	1/4	71.2	12.0
T-7/4	AGT Ser	AAA Lys	ATT Ile	CGT Arg	CTC Leu	GGG Gly	1/4	74.7	14.0
T-10/1	ATA Ile	CAA Gln	CGT Arg	AAG Lys	TTA Leu	GGA Gly	1/30	n.d.	n.d.
T-10/2	GTA Val	TCA Ser	CTC Leu	AAA Lys	CTT Leu	TCC Ser	1/30	n.d.	n.d.
T-10/3	AGG Arg	ACT Thr	CTC Leu	AAA Lys	ATC Ile	ACA Thr	1/30	n.d.	n.d.
T-10/4	AGG Arg	ACT Thr	CTC Leu	AAA Lys	TCA Ser	CAG Gln	1/30	n.d.	n.d.
T-10/5	CGA Arg	ATC Ile	AGC Ser	AAG Lys	TTG Leu	CCC Pro	9/30	75.3	14.5
T-10/6	CGG Arg	TTC Phe	CCC Pro	AGA Arg	ATA Ile	GCG Ala	2/30	75.9	14.8
T-10/7	AAG Lys	AGG Arg	CTA Leu	AAA Lys	TTA Leu	CAA Gln	1/30	79.0	16.1
T-10/8	AAA Lys	ATC Ile	CTC Leu	CGT Arg	CAC His	GCA Ala	7/30	79.4	16.5
T-10/9	AAG Lys	ATC Ile	ACT Thr	AGA Arg	GTA Val	ACT Thr	4/30	79.8	17.1
T-10/10	AGA Arg	GTA Val	TTA Leu	AGA Arg	GTA Val	GGT Gly	3/30	82.2	19.0
Bc-Csp	CAA Gln	CGT Arg	GAA Glu	GTC Val	TTA Leu	-	-	76.9	15.6

The codons and amino acid residues selected at the six randomized positions after six, seven and ten rounds of *in vitro* proteolysis at 57.5 °C are shown for the selected variants (T-*x*/*y*) in comparison with the sequences of *Bs*-CspB and *Bc*-Csp. In addition, the Table lists the frequencies of appearance of every sequence, the melting temperatures (t_M), and the changes in the Gibbs free energy of denaturation at 70 °C ($\Delta\Delta G_D$) relative to wild-type *Bs*-CspB for the purified proteins. As in Table 1 coincidences in sequence with wild-type *Bs*-CspB are italicized.

Bs-CspB.[32] No more wild-type sequences were found in the 30 clones that were analyzed after round 10. At the carboxy-terminal position 67 the selection yielded six different residues, which are coded by ten different triplets. These residues differ greatly in size, but charged as well as large aliphatic and aromatic side-chains have been eliminated from this position during the selection.

The six variants that were overexpressed and purified are all strongly stabilized and show t_M values that are 21.9-28.2 degrees higher than the t_M value of wild-type *Bs*-CspB (Figure 3(b)). Most of them are also more stable than the naturally thermostable *Bc*-Csp, which has a t_M value of 76.9 °C (Figure 3b). The most stable variant shows a melting temperature of 82.2 °C and an increase in ΔG_D of 19.0 kJ/mol (Table 2). Thus it reaches the cold shock protein *Tm*-Csp from the hyperthermophilic organism *Thermotoga maritima* in stability, which shows a t_M value of 82.0 °C at pH 7.0.[41] Nevertheless, at the randomized positions the two proteins have only one residue (Val66) in common. The best variant differs at all six positions from both the mesophilic parent protein *Bs*-CspB and the thermophilic reference *Bc*-Csp, as well as from the most stable variant obtained in the GdmCl-selection ($t_M = 76.0$ °C; Table 1).

Electrostatic contributions to the additional stability

The salt dependence of the stability of a protein provides a signature for electrostatic interactions at the protein surface, because they are screened by counterions. The mesophilic parent protein *Bs*-CspB is destabilized overall by ionic interactions, and therefore its stability increases strongly when these unfavorable interactions are screened by 0-0.5 M NaCl. In contrast, the thermophilic reference protein *Bc*-Csp has an overall stabilizing surface charge distribution, and its stability decreases between 0 and 0.5 M NaCl (Figure 3(c) and ref.[32]).

In the best variant from the GdmCl-selection (variant G-5, Table 1) the electrostatic interactions are partially improved. Its stability still increases between 0 and 0.5 M NaCl, but less strongly than in the case of wild-type *Bs*-CspB (Figure 3(c)), where Glu3 and Glu66 repel each other and thus contribute unfavorable electrostatic interactions.[32] These two negatively charged residues are eliminated in the variants from both the thermal and the GdmCl-selection. The charged denaturant GdmCl attenuates ionic interactions, and therefore the most stable variant from the GdmCl-selection (G-5, Table 1) showed only uncharged, mostly non-polar residues at all six positions (Table 1). It suggests that this variant was optimized for stability at high salt concentration primarily by an improvement of hydrophobic interactions.

The thermal selection was guided by an optimization of both non-polar and electrostatic interactions. T-10/10, the most stable variant ($t_M = 82.2$ °C) shows strongly improved electrostatic properties and resembles the thermophilic reference protein in its salt dependence (Figure 3(c)), although the two proteins differ strongly in the distribution of charged residues at the randomized positions. Negatively charged residues at positions 3 and 66 as in wild-type *Bs*-CspB were avoided in variant T-10/10 (Table 2) and positively charged residues introduced at positions 2 and 64 during the selection. This stabilizes CspB probably by disrupting an unfavorable region of high negative charge density (Figure 1(c) and (d)). The electrostatic interactions can, apparently, be improved in multiple ways, by introducing positively charged residues either at position 3 as in *Bc*-Csp, or at positions 64 and/or 2 as in many of the selected stable variants (Table 2).

The strong stabilization by non-polar residues at the surface, as found in both selections, is surprising. It may originate in part from a destabilization of the unfolded state.

The best variants from the thermal selection (T-10/10) and the GdmCl-selection (G-5) are similarly stable towards unfolding by GdmCl at 25 °C, and both are slightly more stable than *Bc*-Csp

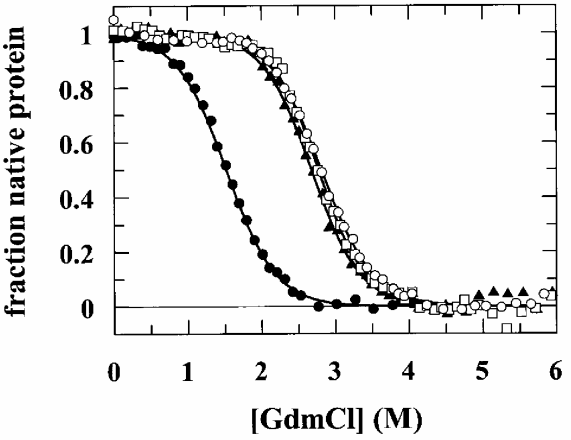


Figure 5. GdmCl-induced equilibrium unfolding transitions of the variants G-5 (□) and T-10/10 (○) at 25 °C. For comparison, the transitions for *Bs*-CspB (●), and *Bc*-Csp (▲) are also shown. These data are taken from ref.[50] The fractions of native protein as obtained after two-state analyses of the data are shown as a function of the GdmCl concentration. The transitions were measured by protein fluorescence at 343 nm (10 nm band width) after excitation at 280 nm (5 nm band width). The protein concentrations were 1.0 μM in 0.1 M Na cacodylate (pH 7.0). The continuous lines represent least-squares fit analyses of the experimental data based on a two-state U⇌N unfolding mechanism.[54] The transition midpoints, cooperativity values (m) and the ΔG_D values at 2.0 M GdmCl, respectively, were 2.77 M GdmCl, −7.8 kJ/(mol M) and 6.0 kJ/mol for variant G-5 and 2.82 M GdmCl, −7.1 kJ/(mol M) and 5.8 kJ/mol for variant T-10/10.

(Figure 5). At 2 M GdmCl the stabilities of T-10/10 and G-5 are 0.7 and 0.9 kJ/mol, respectively, higher than the stability of *Bc*-Csp. This corresponds well with the results obtained in the presence of 2 M NaCl at 70 °C from thermal unfolding transitions (Figure 3(c)). Together, these results confirm that, in fact, the GdmCl selection led to protein variants that are optimized for stability at high salt concentration, whereas the thermal selection led to variants with optimized thermal stability.

Functional properties of the stabilized variants

Cold shock proteins bind to single-stranded nucleic acids, and in the cells they are assumed to interact with DNA and/or mRNA,[42] preferentially with pyrimidine-rich sequences.[43] Binding is accompanied by a strong decrease in the fluorescence of Trp8, which is at the binding site. Comparative fluorescence titrations revealed that the most stable variants from the two selections bind to $(dT)_7$, a heptadeoxynucleotide, with the same affinity as wild-type CspB. The apparent dissociation constants K_D (at 15 °C) were $7.0(\pm 0.7) \times 10^{-9}$ M for wild-type CspB, $7.6(\pm 0.5) \times 10^{-9}$ M for variant G-5 from the GdmCl selection, and $7.1(\pm 0.3) \times 10^{-9}$ M for variant T-10/10 from the thermal selection. Thus, the strong stabilization by mutations of solvent-exposed residues remote from the active site left the nucleic acid binding activity of the protein unchanged.

Conclusions

By using a library of *Bs*-CspB variants with randomized codons for six surface positions and the Proside selection technique, we identified many residue combinations that converted *Bs*-CspB into a thermostable protein. Several of the selected variants were significantly more stable than the thermophilic reference *Bc*-Csp, even though they do not contain an Arg3 and a Leu66. These two residues were found to be responsible for the increased stability of *Bc*-Csp.[32] Evidently, Glu3Arg and Glu66Leu are not the only mutations that can confer thermostability to *Bs*-CspB. The other four randomized positions also carry a strong potential for stabililization, even though the particular sequence differences between *Bs*-CspB and *Bc*-Csp were found to be unimportant for stability. The strategy used by nature to stabilize *Bc*-Csp evidently represents only one of many possibilities. The combinations found by Proside for the six positions could not be deduced from phylogenetic comparisons or from a theoretical analysis of the folded protein only.

Surface positions are much better suited for stabilizing a protein than positions in its interior. Changes in the densely packed cores[44] often create packing defects and are thus destabilizing.[45] Residues at the surface tend to be flexible, show few intra-protein interactions, and thus their contributions to stability often are locally confined and additive.[46] Moreover, the protein surface is in contact with the solvent. Therefore, in Proside, the course of the stabilization can be directed by the conditions used in the selection cycles, and the proteins can be optimized for applications in specific solvents. Surface changes that spare the active site are unlikely to compromise the function of a protein. In fact, the strong increase in stability of surface-optimized *Bs*-CspB was not accompanied by a loss of affinity for nucleic acids, confirming that high thermostability is compatible with activity at low temperatures.[38,47]

Our results are encouraging. They indicate that there is a high potential for stabilizing a protein by making a few changes at its surface. The sites for stabilization can be chosen such that they do not interfere with function, and stabilized variants can be reliably selected from large libraries using Proside.

Together with two closely related techniques[35,39] Proside[36] is the only general evolutionary method for protein stabilization. In all other techniques stabilized variants are selected or screened for by procedures that are based on the specific properties of the particular protein, such as binding to a ligand or a surface, enzymatic activity at elevated temperature, or the growth of an organism under particular conditions. In Proside the selection is independent of such functional properties and thus easily applied to many proteins. The selection step is performed *in vitro*, and therefore the solvent conditions can be varied widely. Proside should thus be useful for stabilizing proteins for applications in biotechnology as well as for exploring fundamental aspects of protein folding.

Materials and Methods

Construction of the phage library

The library of CspB variants was constructed using PCR and oligonucleotides with degenerate base triplets (NNN) at the positions that code for the six amino acids to be randomized. Thus, the theoretical diversity of this library was about 7×10^{10} at the DNA level. To obtain a high yield of recombinant phages the repertoire of mutated DNA fragments comprising *cspB* were hybridized to the (+) DNA strand of phage hosting the wild-type *cspB* gene, completed to double-stranded DNA and transfected into *E. coli* XL1Blue (Stratagene, La Jolla, USA). The template strand contained an amber codon next to the *cspB* gene, which was not present in the mutated fragments. This amber codon was employed to selectively eliminate the phages with wild-type CspB by infecting the non-supressor-strain *E. coli* Able K (Stratagene, La Jolla, USA) with the phages isolated from XL1Blue. The initial library of 2×10^7-3×10^7 transformants seemed to be sufficiently large and to contain many of the possible combinations of codons at the randomized positions. This is indicated by the number of different codons found for frequently selected amino acids (cf. Tables 1 and 2). The eight Leu found at position 46 were coded by three different codons, the 14 Leu

found at 66 were coded by all six Leu codons. Similarly, the five Arg found at position 2 and the ten Arg at position 64 were coded by four different Arg codons each.

Selections in the presence of guanidinium chloride (GdmCl)

Selections 1-3 were performed with about 10^9 phages in 50 μl 1.5 M GdmCl, 100 mM K phosphate (pH 8.0) at 20 °C with 0.25 μM CT for 30 minutes. Phages hosting wild-type *Bs*-CspB retained 0.006 % of their infectivity under these conditions. Selections 4-6 were performed in 1.75 M GdmCl.

Selections at elevated temperature

The first six rounds of selection at 57.5 °C were performed with 2.5 μM CT in 100 mM K phosphate (pH 8.0) for 15 minutes. Then the pool as present after round 4 was exposed to three rounds (5-7) of 15 minutes proteolysis with 25 μM CT followed by a 15 minutes proteolysis with 0.2 μM trypsin. In the three final rounds (8-10) 25 μM CT was added twice for five minutes each. Before adding the protease phages were preincubated for two minutes at 57.5 °C to establish the folding equilibrium.

Protein purification, stability measurements, and K_D measurements

The genes for CspB variants were PCR-amplified with oligonucleotide primers coding for the mutations at the gene termini (comprising positions 2, 3, 64, 66 and 67) and single-stranded phage DNA as template, and cloned into the expression plasmid pET11a (Novagen, Madison, Wisconsin, USA). The proteins were overproduced in *E. coli* BL21(DE3)pLysS (Stratagene, La Jolla, USA). After cell lysis and centrifugation, all variants were found in the supernatant and were purified as described for wild-type *Bc*-Csp[48] and *Bs*-CspB[49] with minor modifications. The purified proteins resembled wild-type *Bs*-CspB in solubility. Thermal unfolding transitions of 4 μM protein were measured in 0.1 M Na cacodylate/HCl (pH 7.0) by circular dichroism at 222.6 nm and analysed as described.[48] GdmCl-induced unfolding transitions were determined as described in ref.[50]

Binding of heptadeoxythymidine to CspB was followed by the 95 % decrease in protein fluorescence at 350 nm (20 nm band width) after excitation at 280 nm (10 nm band width) and analyzed based on the observed stoichiometry of two *Bs*-CspB molecules binding to one molecule of heptadeoxythymidine. To determine the K_D values 1.0 ml of 0.1 μM CspB in 20 mM Na cacodylate, 0.5 mM polyoxyethyl-(8)-decylether (pH 7.5) was titrated at 15 °C with 20 μM heptadeoxythymidine in the same buffer.

Acknowledgments

The stability data of wild-type *Bs*-CspB and wild-type *Bc*-Csp in Figures 3(c) and 4 were kindly provided by Dieter Perl. We thank the members of our group and Terrence Oas for useful discussions and comments on the manuscript.

References

1. Perutz, M. & Raidt, H. (1975). Stereochemical basis of heat stability in bacterial ferredoxins and in haemoglobin A2. *Nature,* **255**, 256-259.
2. Xiao, L. & Honig, B. (1999). Electrostatic contributions to the stability of hyperthermophilic proteins. *J. Mol. Biol.* **289**, 1435-1444.
3. de Bakker, P. I., Hunenberger, P. H. & McCammon, J. A. (1999). Molecular dynamics simulations of the hyperthermophilic protein sac7d from *Sulfolobus acidocaldarius*: contribution of salt bridges to thermostability. *J. Mol. Biol.* **285**, 1811-1830.
4. Elcock, A. H. (1998). The stability of salt bridges at high temperatures: implications for hyperthermophilic proteins. *J. Mol. Biol.* **284**, 489-502.
5. Kumar, S., Ma, B., Tsai, C. J. & Nussinov, R. (2000). Electrostatic strengths of salt bridges in thermophilic and mesophilic glutamate dehydrogenase monomers. *Proteins: Struct. Funct. Genet.* **38**, 368-383.
6. Matthews, B. W. (1991). Mutational analysis of protein stability. *Curr. Opin. Struct. Biol.* **1**, 17-21.
7. Lebbink, J. H., Knapp, S., van der Oost, J., Rice, D., Ladenstein, R. & de Vos, W. M. (1998). Engineering activity and stability of *Thermotoga maritima* glutamate dehydrogenase. I. Introduction of a six-residue ion-pair network in the hinge region. *J. Mol. Biol.* **280**, 287-296.
8. Lebbink, J. H., Knapp, S., van der Oost, J., Rice, D., Ladenstein, R. & de Vos, W. M. (1999). Engineering activity and stability of *Thermotoga maritima* glutamate dehydrogenase. II: construction of a 16-residue ion-pair network at the subunit interface. *J. Mol. Biol.* **289**, 357-369.
9. Meeker, A. K., Garcia-Moreno, B. & Shortle, D. (1996). Contributions of the ionizable amino acids to the stability of staphylococcal nuclease. *Biochemistry,* **35**, 6443-6449.
10. Sali, D., Bycroft, M. & Fersht, A. R. (1991). Surface electrostatic interactions contribute little to stability of barnase. *J. Mol. Biol.* **220**, 779-788.
11. Serrano, L., Horovitz, A., Avron, B., Bycroft, M. & Fersht, A. R. (1990). Estimating the contribution of engineered surface electrostatic interactions to protein stability by using double-mutant cycles. *Biochemistry,* **29**, 9343-9352.
12. Lumb, K. J. & Kim, P. S. (1996). Interhelical salt bridges, coiled-coil stability, and specificity of dimerization - response. *Science,* **271**, 1137-1138.
13. Lavigne, P., Sönnichsen, F. D., Kay, C. M. & Hodges, R. S. (1996). Interhelical salt bridges, coiled-coil stability, and specificity of dimerization. *Science,* **271**, 1136-1138.
14. Takano, K., Tsuchimori, K., Yamagata, Y. & Yutani, K. (2000). Contribution of salt bridges near the surface of a protein to the conformational stability. *Biochemistry,* **39**, 12375-12381.
15. Li, W. T., Shriver, J. W. & Reeve, J. N. (2000). Mutational analysis of differences in thermostability between histones from mesophilic and hyperthermophilic archaea. *J. Bacteriol.* **182**, 812-817.
16. Frankenberg, N., Welker, C. & Jaenicke, R. (1999). Does the elimination of ion pairs affect the thermal stability of cold shock protein from the hyperthermophilic bacterium *Thermotoga maritima*? *FEBS Letters,* **454**, 299-302.
17. Marti, D. N., Jelesarov, I. & Bosshard, H. R. (2000). Interhelical ion pairing in coiled coils: solution structure of a heterodimeric leucine zipper and determi-

nation of pKa values of glu side-chains. *Biochemistry,* **39**, 12804-12818.
18. Hendsch, Z. S. & Tidor, B. (1999). Electrostatic interactions in the GCN4 leucine zipper: substantial contributions arise from intramolecular interactions enhanced on binding. *Protein Sci.* **8**, 1381-1392.
19. Desjarlais, J. R. & Handel, T. M. (1999). Side-chain and backbone flexibility in protein core design. *J. Mol. Biol.* **290**, 305-318.
20. Lazar, G. A., Desjarlais, J. R. & Handel, T. M. (1997). De novo design of the hydrophobic core of ubiquitin. *Protein Sci.* **6**, 1167-1178.
21. Finucane, M. D. & Woolfson, D. N. (1999). Core-directed protein design. II. Rescue Of a multiply mutated and destabilized variant of ubiquitin. *Biochemistry,* **38**, 11613-11623.
22. Waldburger, C. D., Schildbach, J. F. & Sauer, R. T. (1995). Are buried salt bridges important for protein stability and conformational specificity? *Nature Struct. Biol.* **2**, 122-128.
23. Baldwin, E., Xu, J., Hajiseyedjavadi, O., Baase, W. A. & Matthews, B. W. (1996). Thermodynamic and structural compensation in ''size-switch'' core repacking variants of bacteriophage T4 lysozyme. *J. Mol. Biol.* **259**, 542-559.
24. Malakauskas, S. M. & Mayo, S. L. (1998). Design, structure and stability of a hyperthermophilic protein variant. *Nature Struct. Biol.* **5**, 470-475.
25. Pakula, A. A. & Sauer, R. T. (1989). Amino acid substitutions that increase the thermal stability of the lambda Cro protein. *Proteins: Struct. Funct. Genet.* **5**, 202-210.
26. Loladze, V. V., Ibarra-Molero, B., Sanchez-Ruiz, J. M. & Makhatadze, G. I. (1999). Engineering a thermostable protein via optimization of charge-charge interactions on the protein surface. *Biochemistry,* **38**, 16419-16423.
27. Grimsley, G. R., Shaw, K. L., Fee, L. R., Alston, R. W., Huyghues-Despointes, B. M., Thurlkill, R. L., Scholtz, J. M. & Pace, C. N. (1999). Increasing protein stability by altering long-range coulombic interactions. *Protein Sci.* **8**, 1843-1849.
28. Spector, S., Wang, M., Carp, S. A., Robblee, J., Hendsch, Z. S., Fairman, R., Tidor, B. & Raleigh, D. P. (2000). Rational modification of protein stability by the mutation of charged surface residues. *Biochemistry,* **39**, 872-879.
29. 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.
30. Jaenicke, R. & Böhm, G. (1998). The stability of proteins in extreme environments. *Curr. Opin. Struct. Biol.* **8**, 738-748.
31. Wintrode, P. L. & Arnold, F. H. (2000). Temperature adaptation of enzymes: lessons from laboratory evolution. *Advan. Protein Chem.* **55**, 161-225.
32. Perl, D., Mueller, U., Heinemann, U. & Schmid, F. X. (2000). Two exposed amino acid residues confer thermostability on a cold shock protein. *Nature Struct. Biol.* **7**, 380-383.
33. Arnold, F. A. & Volkov, A. A. (1999). Directed evolution of biocatalysts. *Curr. Opin. Chem. Biol.* **3**, 54-59.
34. Tobin, M. B., Gustafsson, C. & Huisman, G. W. (2000). Directed evolution: the 'rational' basis for 'irrational' design. *Curr. Opin. Struct. Biol.* **10**, 421-427.
35. Kristensen, P. & Winter, G. (1998). Proteolytic selection for protein folding using filamentous bacteriophages. *Fold. Des.* **3**, 321-328.
36. Sieber, V., Plückthun, A. & Schmid, F. X. (1998). Selecting proteins with improved stability by a phage-based method. *Nature Biotechnol.* **16**, 955-960.
37. Spiller, B., Gershenson, A., Arnold, F. H. & Stevens, R. C. (1999). A structural view of evolutionary divergence. *Proc. Natl Acad. Sci. USA,* **96**, 12305-12310.
38. Zhao, H. & Arnold, F. H. (1999). Directed evolution converts subtilisin E into a functional equivalent of thermitase. *Protein Eng.* **12**, 47-53.
39. Finucane, M. D., Tuna, M., Lees, J. H. & Woolfson, D. N. (1999). Core-directed protein design. I. An experimental method for selecting stable proteins from combinatorial libraries. *Biochemistry,* **38**, 11604-11612.
40. Stengele, I., Bross, P., Garces, X., Giray, J. & Rasched, I. (1990). Dissection of functional domains in phage fd adsorption protein. Discrimination between attachment and penetration sites. *J. Mol. Biol.* **212**, 143-149.
41. Wassenberg, D., Welker, C. & Jaenicke, R. (1999). Thermodynamics of the unfolding of the cold-shock protein from *Thermotoga maritima*. *J. Mol. Biol.* **289**, 187-193.
42. Graumann, P., Wendrich, T. M., Weber, M. H. W., Schröder, K. & Marahiel, M. A. (1997). A family of cold shock proteins in *Bacillus subtilis* is essential for cellular growth and for efficient protein synthesis at optimal and low temperatures. *Mol. Microbiol.* **25**, 741-756.
43. Lopez, M. M., Yutani, K. & Makhatadze, G. I. (1999). Interactions of the major cold shock protein of *Bacillus subtilis* CspB with single-stranded DNA templates of different base composition. *J. Biol. Chem.* **274**, 33601-33608.
44. Richards, F. M. (1977). Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* **6**, 151-176.
45. Lim, W. A. & Sauer, R. T. (1991). The role of internal packing interactions in determining the structure and stability of a protein. *J. Mol. Biol.* **219**, 359-376.
46. Reidhaar-Olson, J. F. & Sauer, R. T. (1990). Functionally acceptable substitutions in two alpha-helical regions of lambda repressor. *Proteins: Struct. Funct. Genet.* **7**, 306-316.
47. Merz, A., Yee, M. C., Szadkowski, H., Pappenberger, G., Crameri, A., Stemmer, W. P. C., Yanofsky, C. & Kirschner, K. (2000). Improving the catalytic activity of a thermophilic enzyme at low temperatures. *Biochemistry,* **39**, 880-889.
48. Mueller, U., Perl, D., Schmid, F. X. & Heinemann, U. (2000). Thermal stability and atomic-resolution crystal structure of the *Bacillus caldolyticus* cold shock protein. *J. Mol. Biol.* **297**, 975-988.
49. Schindelin, H., Herrler, M., Willimsky, G., Marahiel, M. A. & Heinemann, U. (1992). Overproduction, crystallization, and preliminary X-ray diffraction studies of the major cold shock protein from *Bacillus subtilis*, CspB. *Proteins: Struct. Funct. Genet.* **14**, 120-124.
50. Perl, D., Welker, C., Schindler, T., Schröder, K., Marahiel, M. A., Jaenicke, R. & Schmid, F. X. (1998). Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins. *Nature Struct. Biol.* **5**, 229-235.

51. Schindelin, H., Marahiel, M. A. & Heinemann, U. (1993). Universal nucleic acid-binding domain revealed by crystal structure of the *B. subtilis* major cold-shock protein. *Nature,* **364**, 164-168.
52. Guex, N. & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis,* **18**, 2714-2723.
53. Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281-296.
54. Santoro, M. M. & Bolen, D. W. (1988). Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl α- chymotrypsin using different denaturants. *Biochemistry,* **27**, 8063-8068.

Edited by C. R. Matthews

(Received 1 February 2001; received in revised form 10 April 2001; accepted 13 April 2001)

7.3 Teilarbeit C

C

Andreas Martin, Insa Kather and Franz X. Schmid.

Origins of the High Stability of an *in vitro*-selected Cold-shock Protein.

The Journal of Molecular Biology **318**, 1341-1349 (2002)

doi: 10.1016/S0022-2836(02)00243-7 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2002) **318**, 1341–1349

JMB

Origins of the High Stability of an *in vitro*-selected Cold-shock Protein

Andreas Martin, Insa Kather and Franz X. Schmid*

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

*Corresponding author

In previous work, we had identified stabilized forms of the cold-shock protein *Bs*-CspB from *Bacillus subtilis* in a combinatorial library by an *in vitro* selection procedure. In this library, the sequence positions 2, 3, 46, 64, 66, and 67 had been randomized, because *Bs*-CspB differs from the naturally thermostable homolog *Bc*-Csp from *Bacillus caldolyticus*, among others, at these six positions. For the most stable selected variant, the midpoint of thermal unfolding (t_M) increased by 28.2 deg. C and the Gibbs free energy of unfolding (ΔG_D) by 19 kJ/mol. Here, we analyzed by site-directed mutagenesis how the selected residues contribute individually to this strong stabilization. Val3 and Val66, which replace Glu3 and Glu66 of wild-type *Bs*-CspB, each contribute about 7 kJ/mol to stability, the Thr64Arg substitution contributes 4.5 kJ/mol, and 3.2 kJ/mol originate from the Ala46Leu replacement. Gly67 at the carboxy terminus is unimportant for stability, the Arg selected at position 2 is overall slightly destabilizing but improves the coulombic interactions. The best variant differs from *Bc*-Csp at all six positions; nevertheless, natural and *in vitro* selection followed similar principles. In both cases, negatively charged residues at the adjacent positions 3 and 66 are avoided, and a positively charged residue is introduced into this area of the protein surface. Its exact location is unimportant. It can be at position 3, as in the thermophilic *Bc*-Csp, or at positions 2 or 64, as in the most stable selected variant. These positively charged residues contribute to stability not by engaging in pairwise coulombic interactions with a specific carboxyl group, but by generally improving the charge distribution in this particular region of the protein surface. These coulombic effects contribute significantly to the thermostability of the cold-shock proteins. They are only weakly interdependent and best explained by the presence of a flexible ion network at the protein surface. Our results emphasize that surface positions are very good candidates for optimizing protein stability.

Keywords: cold-shock protein; protein stability; *in vitro* selection; phage display; electrostatic interactions

Introduction

The stability of natural proteins is generally very low, just sufficient to ensure proper functioning under cellular conditions. The reason for this marginal stability is clear. Most spontaneous mutations in proteins are unfavorable and decrease conformational stability. The rare mutations that increase the stability of a protein beyond the requirements of the particular organism are thus not conserved in the absence of selective pressure, and only the minimally required stability of a protein is maintained in evolution.[1] This suggests that there should be ample room for optimizing stability by protein engineering.

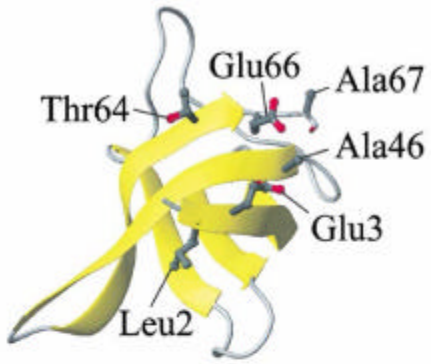
Protein stability can be increased by computer-aided design, which employs knowledge derived from structural and thermodynamic investigations,[2,3] or, alternatively, by evolutionary approaches, in which large protein libraries are searched for stabilized variants by screening or selection techniques.[4–6] In addition to their relevance for optimizing commercial proteins,

Abbreviations used: *Bs*-CspB, cold-shock protein from *Bacillus subtilis*; *Bc*-Csp, cold-shock protein from *Bacillus caldolyticus*; t_M, midpoint of a thermal unfolding transition.

E-mail address of the corresponding author: fx.schmid@uni-bayreuth.de

0022-2836/02/$ - see front matter

a



b

```
                1          10         20         30
Bs-CspB    MLEGKVKWF NSEKGFGFIE VEG-QDDVFVH FSAIQ
CspB Sel   MRVGKVKWF NSEKGFGFIE VEG-QDDVFVH FSAIQ
Bc-Csp     MQRGKVKWF NNEKGYGFIE VEG-GSDVFVH FTAIQ
Tm-Csp      MRGKVKWF DSKKGYGFIT KDE-GGDVFVH WSAIE
Aa-Csp    MSFRGTVKWF SKDKGYGFIT REDTNADVFVH FTDIQ

               35     40         50         60
Bs-CspB    GEGFK TLEEGQAVSF EIVEGNRGPQ AANVTKEA
CspB Sel   GEGFK TLEEGQLVSF EIVEGNRGPQ AANVRKVG
Bc-Csp     GEGFK TLEEGQEVSF EIVQGNRGPQ AANVVKL
Tm-Csp     MEGFK TLKEGQVVEF EIQEGKKGPQ AAHVKVVE
Aa-Csp     MEGFK TLQKGQKVEF DVVEDTKGPR AKNVRVLGE
```

Figure 1. (a) Tertiary structure of *Bs*-CspB.[11] The side-chains of the six amino acid residues that were randomized for selection with *Proside* are shown in ball-and-stick representation. (b) Sequence alignment of the cold-shock proteins *Bs*-CspB from the mesophile *Bacillus subtilis*, the most stable selected *Bs*-CspB variant Sel, *Bc*-Csp from the thermophile *Bacillus caldolyticus*, *Tm*-Csp and *Aa*-Csp from the hyperthermophiles *Thermotoga maritima* and *Aquifex aeolicus*, respectively. The positively charged residues Arg and Lys are colored blue, the negatively charged residues Glu and Asp are colored red. The six positions in *Bs*-CspB chosen for random mutagenesis are underlaid in grey.

both approaches open new avenues to improve our understanding of the principles of protein stability. After a successful design or selection experiment, it is thus important to analyze the nature and the magnitude of the contributions to protein stability of the individual sequence changes on a residue-by-residue basis.

In previous work, we identified stabilized variants in a combinatorial library of the cold-shock protein *Bs*-CspB by using *Proside*, a general selection method for proteins with improved thermodynamic stability[7–9]. *Bs*-CspB is a small *Bacillus subtilis* protein of 67 amino acid residues with a marginal conformational stability. The Gibbs free energy of unfolding, ΔG_D, is about 12 kJ/mol at pH 7.0 and 25 °C, and the midpoint of the thermal unfolding transition, t_M, is at 54.0 °C. *Bs*-CspB differs from the thermophilic homolog *Bc*-Csp from *Bacillus caldolyticus* at 12 surface-exposed positions. Six of these positions (Figure 1(a)) were randomized by saturation mutagenesis, the corresponding library of sequences was inserted into the gene-3 protein of the filamentous phage fd, and stabilized variants were selected by the *Proside* technique.[10] *Proside* links the protease-resistance of a protein variant with the infectivity of the corresponding phage, and thus allows selection for increased protein stability.[7] Many strongly stabilized forms of *Bs*-CspB were found in these selections, and the most stable showed a t_M value of 82.2 °C and an increase in ΔG_D of 19 kJ/mol. It differs from *Bc*-Csp at all six randomized positions (Figure 1(b)).

To elucidate the physical basis of the strongly enhanced stability of the best variant, we introduced the selected residues individually into wild-type *Bs*-CspB by site-directed mutagenesis and measured their contributions to stability. Then the selected residues were combined in various arrangements to examine whether their contributions to stability are additive, and to uncover energetic interactions, in particular between charged groups. Finally, we used the results of this analysis for a further stabilization of *Bs*-CspB. By combining four stabilizing residues, a variant was obtained with an increase in t_M of 31.6 deg. C.

Results

Contributions of individual positions

In the library of *Bs*-CspB variants the positions 2, 3, 46, 64, 66, and 67 had been randomized (Figure 1(a)).[10] These positions are located near each other on the surface of the folded protein, and the distances between the respective C^β atoms, as derived from the crystal structure of the wild-type protein,[11] are shown schematically in Figure 2(a). All variants that were obtained after the *Proside* selection had lost the two Glu residues at positions 3 and 66, and most of them had acquired positively charged residues at positions 2 and 64 (Figure 2(b)). Thus, they no longer display the patch of negative charge density around positions 3 and 66, and their overall net charge is much less negative than in wild-type *Bs*-CspB.

The six residues selected at the randomized positions in this variant were introduced into wild-type *Bs*-CspB, first individually, and then in various combinations. The respective protein variants were produced, and their stabilities were determined from thermal unfolding transitions measured by the change in circular dichroism (CD) at 222.6 nm as described.[10] The stability data of all variants are given in Table 1.

Position 67

For position 67, which is at the carboxy terminus of *Bs*-CspB (Figure 1(a)), no clear preference for a particular residue could be observed in the *Proside* selections. Six different residues were found, all of

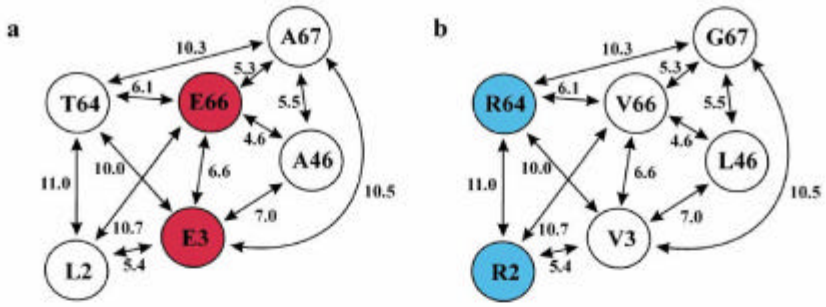


Figure 2. A representation of the six randomized positions at the surface of *Bs*-CspB. The distances between the C^{β} atoms of these six amino acid residues as derived from the crystal structure of the wild-type protein[11] are shown. (a) Residues at positions 2, 3, 46, 64, 66 and 67 in wild-type *Bs*-CspB. The negative charges of Glu3 and Glu66 are indicated by the red color. (b) Respective amino acid residues selected in the most stable variant. The positively charged residues Arg2 and Arg64 are marked in blue. Distances are supposed to be the same as for wild-type *Bs*-CspB in (a).

them small and uncharged.[10] When Ala67 was substituted by a glycine residue (as present in the most stable variant) in a mutant that also contained four other stabilizing mutations (at 2, 3 64, and 66), t_M remained constant at 75.0 °C (compare variants 11 and 12 in Table 1). This suggests that residue 67 is unimportant for the stability of *Bs*-CspB, in good agreement with the finding that its truncation does not affect stability,[12] and with the variability in sequence and length at the carboxy termini of homologous cold-shock proteins (cf. Figure 1).

Position 46

For position 46 an Ala, as in the wild-type protein, was never found in the *Proside* selections. Most of the stabilized variants contained a Leu46. The single Ala46Leu mutation stabilizes *Bs*-CspB by 3.2 kJ/mol (variant 1, Table 1), and this stabilization is even more pronounced in the presence of other substitutions. When, in the best variant, Leu46 is reverted back to Ala, t_M and ΔG_D decrease by 7.2 deg. C and 4.6 kJ/mol, respectively (variants Sel and 12, Table 1), and in a variant containing Val3, Arg64 and Val66 the Ala46Leu substitution stabilizes the protein by 4.9 kJ/mol (variants 9 and HS in Table 1). The latter two variants contain Val at position 66, and it is possible that Leu46 and Val66 interact directly. Note that in wild-type *Bs*-CspB, the C^{β} atoms of Ala46 and Glu66 are only 4.6 Å apart from each other (Figure 2).[11]

Positions 3 and 66

Glu3 and Glu66 are near each other (Figure 2(a)) and lead to a patch of negative charge density in the folded wild-type protein. Thus, they decrease conformational stability, and previous results had

Table 1. Thermodynamic parameters for wild-type *Bs*-CspB, the most stable selected variant (Sel), the variants for analysis of single residue contributions and the designed hyperstable variant (HS)

Variant	Pos. 2	Pos. 3	Pos. 46	Pos. 64	Pos. 66	Pos. 67	t_M (°C)	$-\Delta G_D$ (kJ/mol)	$-\Delta\Delta G_D$ (kJ/mol)	ΔG_D 2 M NaCl (kJ/mol)	$\Delta\Delta G_D$ 2 M NaCl (kJ/mol)	$\Delta\Delta G_D$ coulomb (kJ/mol)
Bs-CspB	Leu	Glu	Ala	Thr	Glu	Ala	54.0	−11.1	0	0.8	0	0
1	Leu	Glu	***Leu***	Thr	Glu	Ala	57.9	−7.9	3.2	3.1	2.3	0.9
2	Leu	***Val***	Ala	Thr	Glu	Ala	64.0	−3.7	7.4	3.6	2.8	4.6
3	Leu	Glu	Ala	Thr	***Val***	Ala	64.4	−3.9	7.2	5.5	4.7	2.5
4	Leu	Glu	Ala	***Arg***	Glu	Ala	59.9	−6.6	4.5	2.8	2.0	2.5
5	***Arg***	Glu	Ala	Thr	Glu	Ala	50.7	−12.6	−1.5	−2.6	−3.4	1.9
6	Leu	***Val***	Ala	Thr	***Val***	Ala	73.2	2.1	13.2	8.4	7.6	5.6
7	Leu	***Val***	Ala	***Arg***	Glu	Ala	69.2	−0.5	10.6	5.1	4.3	6.3
8	Leu	Glu	Ala	***Arg***	***Val***	Ala	68.8	−0.6	10.5	6.9	6.1	4.4
9	Leu	***Val***	Ala	***Arg***	***Val***	Ala	78.1	5.4	16.5	9.8	9.0	7.5
10	***Arg***	***Val***	Ala	Thr	***Val***	Ala	70.1	0.1	11.2	5.4	4.6	6.6
11	***Arg***	***Val***	Ala	***Arg***	***Val***	Ala	75.1	2.6	13.7	7.0	6.2	7.5
12	***Arg***	***Val***	Ala	***Arg***	***Val***	***Gly***	75.0	3.3	14.4	7.5	6.7	7.7
Sel	***Arg***	***Val***	***Leu***	***Arg***	***Val***	***Gly***	82.2	7.9	19.0	12.0	11.2	7.8
HS	Leu	***Val***	***Leu***	***Arg***	***Val***	Ala	85.6	10.3	21.4	14.1	13.3	8.1

Mutated amino acid residues are shown in bold italic type. For all proteins, the melting temperature (t_M), the Gibbs free energy of denaturation at 70 °C (ΔG_D) in the absence and in the presence of 2 M NaCl are given, as well as the changes in ΔG_D relative to the wild-type protein ($\Delta\Delta G_D$). The coulombic contribution to $\Delta\Delta G_D$ in the last column is equal to the difference between the $\Delta\Delta G_D$ values at 0 M and at 2 M NaCl. The thermodynamic parameters were derived from thermal unfolding transitions measured in 0.1 M sodium cacodylate/HCl (pH 7.0) by CD at 222.6 nm as described.[14]

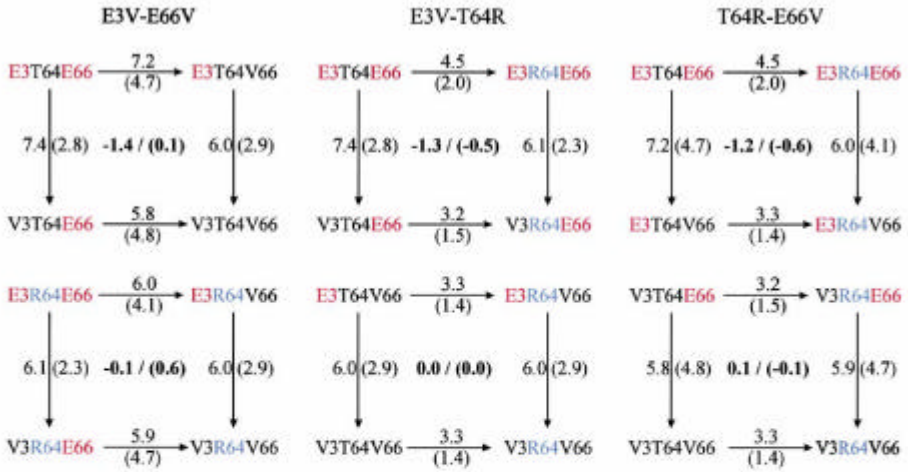


Figure 3. Double mutant cycles with all possible combinations of the amino acid residues as they occur at positions 3, 64 and 66 in wild-type *Bs*-CspB and in the most stable selected variant. Negatively charged residues are colored red, positively charged residues are colored blue. For every single mutation, the $\Delta\Delta G_D$ values at 70 °C in the absence and in the presence of 2 M NaCl (values in parentheses) are shown. In the center of every cycle, the coupling energy between the respective two amino acid residues is given, again in the absence and in the presence of 2 M NaCl. The experimental conditions are given in Table 1. The left pair of cycles specifies the interaction between Glu3 and Glu66, in the absence (upper cycle) and in the presence (lower cycle) of Arg64. The middle pair gives the interaction energies between Glu3 and Arg64, in the absence and in the presence of Glu66, and the right pair gives the energies of interaction between Arg64 and Glu66, in the absence and in the presence of Glu3.

shown that the replacements by the respective residues from the thermophilic *Bc*-Csp (Arg3 and Leu66) stabilize *Bs*-CspB by 14.2 kJ/mol.[12,13] In fact, after the final round of *Proside*, no negatively charged residue was found, and most often positions 3 or 66 were occupied by aliphatic residues such as leucine, isoleucine, or valine. The best variant contained Val3 and Val66.[10]

Both valine residues are strongly stabilizing. The Glu3Val mutation alone increased t_M by 10.0 deg. C and ΔG_D by 7.4 kJ/mol (variant 2, Table 1). The corresponding increases caused by the Glu66Val mutation are 10.4 deg. C and 7.2 kJ/mol (variant 3, Table 1). In combination, the two mutations increase t_M and ΔG_D by 19.2 deg. C and 13.2 kJ/mol, respectively (variant 6, Table 1). Thus, the interaction energy between the two Glu is only −1.4 kJ/mol, pointing to a mainly global effect of the two Glu → Val replacements.

Position 64

At position 64, only positively charged residues were found after the *Proside* selection, and the most stable variant contains an Arg64.[10] The Thr64Arg single mutation stabilized *Bs*-CspB by 4.5 kJ/mol (variant 4, Table 1), and this stabilization depends slightly on the charge distribution in the vicinity of Arg64. When Thr64 is mutated to Arg in *Bs*-CspB variants that already contain the Glu3Val single mutation (variant 7, Table 1), the Glu3Val/Glu66Val double mutation (variant 9, Table 1) or the Glu3Val/Glu66Val/Leu2Arg triple mutation (variant 11, Table 1) the $\Delta\Delta G_D$ values decrease from 4.5 to 3.2 kJ/mol. This decrease is surprisingly small, regarding the fact that the local charge changes from −2 in the wild-type background to +1 in the triple mutant.

As revealed by a triple mutant analysis (Figure 3), the interaction energies between Glu3 and Arg64 (1.3 kJ/mol) and between Glu66 and Arg64 (1.2 kJ/mol) are small, and of a magnitude similar to that of the repulsion between Glu3 and Glu66 (1.4 kJ/mol) in the absence of Arg64. When Arg64 is present, the repulsion between Glu3 and Glu66 vanishes, and conversely, when, alternatively, Glu3 or Glu66 are absent, Arg64 does not interact with the remaining Glu residue. Thus, Arg64 abolishes the repulsive interaction between Glu3 and Glu66 but, apparently, not by engaging in a pairwise attractive interaction with one of the two Glu residues.

Position 2

An Arg at position 2 was found in eight of the ten variants that were characterized after the final round of *Proside*.[10] Surprisingly, as a single mutation, Leu2Arg destabilizes wild-type *Bs*-CspB slightly. It reduces t_M by 3.3 deg. C and ΔG_D by 1.5 kJ/mol (variant 5, Table 1). This destabilization is more pronounced in variants with a lowered negative charge density near position 2, and in the Glu3Val/Glu66Val/Thr64Arg variant, the Leu2Arg mutation destabilizes *Bs*-CspB by 2.1 kJ/mol (compare variants 9 and 11 in Table 1).

We examined whether the Leu2Arg mutation might be stabilizing against proteolysis when

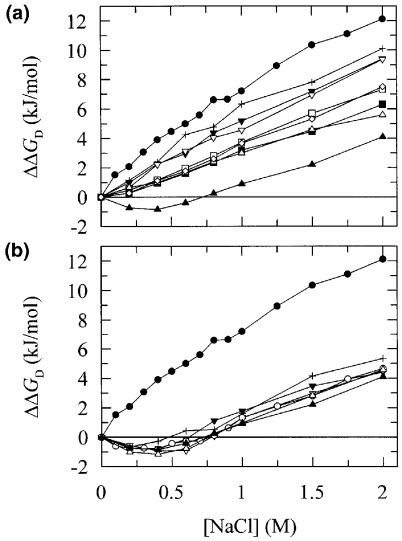


Figure 4. (a) Dependence on NaCl concentration of the change in free energy of unfolding at 70 °C, $\Delta\Delta G_D$, relative to the value measured in the absence of NaCl. (a) Wild-type *Bs*-CspB (●) in comparison to the mutants L2R (+), T64R (▼), E66V (▽), E3V (□), T64R/E66V (◇), E3V/E66V (■), E3V/T64R (△), and the most stable selected variant Sel L2R/E3V/A46L/T64R/E66V/A67G (▲). (b) Wild-type *Bs*-CspB (●),wild-type *Bc*-Csp (○), mutants L2R/E3V/E66V (+), E3V/T64R/E66V (▼), L2R/E3V/T64R/E66V (▽), Sel (▲), and the designed hyperstable variant HS E3V/A46L/T64R/E66V (△).

Bs-CspB is inserted in the gene-3 protein of the phage, as in the *Proside* selections. Comparative *in vitro* proteolysis of the corresponding phage revealed, however, that this mutation is slightly destabilizing in fusion with the gene-3 protein as well. Thus, the reason for its selection in *Proside* remains unclear. Arg2 is close to the linker between the phage protein and the *Bs*-CspB insert and might influence phage assembly or infectivity.

The nature of the contributions to stability

The combination of the selected single mutations reveals small, but significant energetic interactions between residues when the charge pattern is changed by the mutations. This suggests that long-range coulombic interactions contribute to the additional stabilization of the selected *Bs*-CspB variants.

Interactions between charges at the protein surface are screened by counterions. Thus, the stability of protein variants with differences in coulombic interactions should respond differently to changes in the ionic environment provided by the solvent. Variations of the salt concentration are very useful to characterize coulombic contributions to the stability of cold-shock proteins. The parent protein *Bs*-CspB is destabilized overall by ionic interactions, and therefore its stability increases strongly when these unfavorable interactions are screened by 0–0.5 M NaCl.[14] In contrast, the most stable variant, as selected by *Proside*,[10] as well as the thermophilic homolog *Bc*-Csp,[14] have overall stabilizing surface charge distributions, and their stabilities decrease between 0 and 0.5 M NaCl.

To identify coulombic contributions of individual residues, we measured the stabilities of all variants as a function of NaCl concentration (Figure 4). Previous experience suggested that interactions between charged residues are largely screened above 1.5 M salt, and therefore we used the differences in stability between 0 and 2 M NaCl of the various mutants as a measure for the differences in coulombic interactions.

Val3 and Val66

Residues Val3 and Val66 improve electrostatic interactions, because they abolish the negative charges of Glu3 and Glu66 that are present in the wild-type protein. The Glu3Val/Glu66Val variant is stabilized by salt much less than wild-type *Bs*-CspB (Figure 4(a)), and the difference in stability between the two proteins decreases from 13.2 kJ/mol at 0 M NaCl to 7.6 kJ/mol at 2 M NaCl (variant 6, Table 1). Under these conditions, the contributions of the two individual mutations, 2.8 kJ/mol for the Glu3Val and 4.7 kJ/mol for the Glu66Val mutation, are additive (compare the sum of the $\Delta\Delta G_D$ values at 2 M NaCl for variants 2 and 3 with the $\Delta\Delta G_D$ value for variant 6 in Table 1), which suggests that the two valine residues do not interact. The stabilization by 7.6 kJ/mol that persists at high concentrations of salt is probably caused by favorable hydrophobic interactions of Val3 and Val66 with other residues. It is also possible that unfavorable dehydration of Glu3 and Glu66 in the wild-type protein contributes to the salt-independent $\Delta\Delta G_D$.

The decrease in $\Delta\Delta G_D$ by 5.6 kJ/mol between 0 and 2 M NaCl suggests that abolishing the negative charges at positions 3 and 66 (by the Glu3Val/Glu66Val double mutation) stabilizes *Bs*-CspB electrostatically by this amount (in 0.1 M buffer, pH 7.0). Only 1.5 kJ/mol originate from a pairwise repulsion between the two Glu residues (in the absence of NaCl), as revealed by the double mutant analysis for the Glu3Val and Glu66Val replacements (Figure 3). In contrast, in the Arg3-Glu/Leu66Glu variant of the thermostable protein *Bc*-Csp, Glu3 and Glu66 repel each other by −4 kJ/mol.[13] Considering the overall coulombic contributions of the Glu3Val and the Glu66Val single mutations (4.6 and 2.5 kJ/mol, respectively), the small value for the direct pairwise interaction between the two Glu residues in wild-type *Bs*-CspB points to a more global effect of abolishing these two negative charges. The hydrophobic contributions of the two individual Glu → Val substitutions are additive, and therefore it is unlikely that the loss of a pairwise electrostatic repulsion between Glu3 and Glu66 in wild-type *Bs*-CspB is masked by a corresponding

hydrophobic repulsion between the two Val residues in the stabilized variant. The difference between the results obtained for *Bs*-CspB and the double Glu mutant of *Bc*-Csp[13] might be related to Ala67, which is present in *Bs*-CspB, but not in *Bc*-Csp. The crystal structure of *Bs*-CspB[11] shows an orientation of Ala67 that could allow a direct repulsive interaction of its carboxyl group with Glu3, which might compete with a pairwise repulsion between Glu3 and Glu66.

Arg2 and Arg64

The Thr64Arg mutation stabilizes *Bs*-CspB by 4.5 kJ/mol, which is decreased to 2.0 kJ/mol in the presence of 2.0 M NaCl (Figure 4(a); variant 4, Table 1). This suggests that Arg64 contributes about 2.5 kJ/mol of coulombic interactions and 2 kJ/mol of non-polar interactions to stability, relative to Thr64 in the wild-type protein.

Arg2 improves the coulombic interactions at the protein surface, but overall it decreases protein stability. As a consequence, the Leu2Arg mutation is more strongly destabilizing at high concentrations of salt, and $\Delta\Delta G_D$ decreases from −1.5 kJ/mol (0 M NaCl) to −3.4 kJ/mol (2 M NaCl). This suggests that the positive charge introduced by the Leu2Arg mutation stabilizes *Bs*-CspB electrostatically by about 2 kJ/mol, but this contribution is overcompensated by an unfavorable, salt-independent contribution of 3.4 kJ/mol. It might originate from favorable hydrophobic interactions of Leu2 in the wild-type protein that are not possible with an Arg at this position.

Variants with Val3/Val66 and positive charges at positions 2 and/or 64

The single Glu3Arg mutation[13] removed Glu3 and, at the same time, introduced a positive charge. Thus, it stabilized *Bs*-CspB by 11.1 kJ/mol. In the best of the selected variants, corresponding changes occurred, but they were brought about not by one, but by two replacements. Glu3Val removed the Glu residue and thus stabilized by 7.4 kJ/mol, Thr64Arg introduced the positive charge and stabilized by 4.5 kJ/mol. When introduced together the contributions of these two substitutions (10.6 kJ/mol) matched the effect of the single Glu3Arg mutation almost perfectly.

Still, the changes in the coulombic interactions caused by having the Arg residue alternatively at position 3 or 64 are not identical. This is revealed by scrutinizing the dependences on NaCl concentration of the stabilities of the respective mutants. The Glu3Arg substitution alone was sufficient to change the coulombic interactions of *Bs*-CspB from destabilizing to overall stabilizing, as in *Bc*-Csp. As a consequence, the Glu3Arg mutant showed the same salt profile as *Bc*-Csp, which is characterized by a minimum for $\Delta\Delta G_D$ near 0.4 M NaCl. A further change in charge contributed by the Glu66Leu mutation had no additional effect.[13] In contrast, the Glu3Val/Thr64Arg double mutant shows an approximately linear salt-dependence, which coincides almost perfectly with that of the Glu3Val/Glu66Val double mutant (Figure 4(a)). This coincidence suggests that similar improvements in the coulombic interactions can be achieved either by introducing a positive charge at position 64 (by the Thr64Arg replacement) or by deleting a negative charge at position 66 (by the Glu66Val replacement). This is confirmed by results obtained for the Thr64Arg and Glu66Val single mutants. Both reduce the net negative charge by 1 and show identical, largely linear salt profiles (Figure 4(a)).

The Glu3Val mutation reduces the net negative charge also by 1, but it contributes screenable coulombic interactions that are 2.1 kJ/mol larger than the contributions of Glu66Val. This might be due to a better hydration of Glu3 or due to stronger (unfavorable) interactions with other charged residues. This difference is apparent in the presence of Arg64, and the coulombic stabilization brought about by the Glu3Val/Thr64Arg double mutation is 1.9 kJ/mol higher than that of the Thr64Arg/Glu66Val double mutation.

The double mutant cycles (Figure 3) gave similar interaction energies between the mutation pairs Glu3Val and Glu66Val, Glu3Val and Thr64Arg, and Glu66Val and Thr64Arg. The unfavorable interaction between Glu3 and Glu66 vanished, however, when Arg64 was present. Similarly, no favorable interaction could be detected between one of the two Glu residues and Arg64 when the other Glu residue was missing. Thus, the introduction of an Arg residue at position 64 seems to compensate for the small repulsive interaction between Glu3 and Glu66. As soon as this pairwise repulsion is removed (either by the introduction of Arg64 or by deletion of one of the Glu residues), all further substitutions improve the general electrostatic properties of the protein successively, according to the reduction of the overall negative net charge of the protein (Figure 4(a)), and the individual contributions become independent of the nature of the neighboring residues at positions 3, 64 and 66. This is the simplest explanation for the observed effects, but we cannot rule out the possibility of more complex scenarios in which changes in potential pairwise interactions are compensated by counteracting global changes.

The stabilization *via* coulombic effects by the single Glu3Arg mutation is so strong that the respective variant, Glu3Arg–*Bs*-CspB, resembles the naturally thermostable homolog *Bc*-Csp in the salt profile of its stability.[13] Such a strong change in the coulombic interactions could be achieved for our mutants only when the two negative charges at positions 3 and 66 were deleted and a positive charge was inserted in the local environment, as in the Glu3Val/Thr64Arg/Glu66Val triple mutant (Figure 4(b)). The distribution of coulombic effects over more than one residue is accompanied by a gain in hydrophobic stabilization. Compared

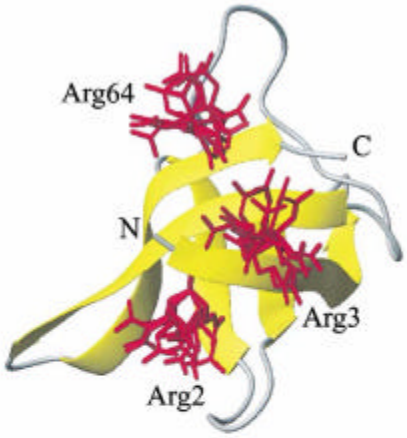


Figure 5. Side-chain rotamers of Arg2, Arg3 and Arg64 in *Bs*-CspB. Based on the structure of the wild-type protein,[11] the rotamers for the three Arg mutations were calculated using the Swiss-pdb viewer. Those with the lowest energy (13 out of 18 for Arg2, 17 for Arg3 and 16 for Arg64) are shown as ensembles in stick representation.

to the respective amino acids present in the thermophilic *Bc*-Csp (Arg3, Val64, and Leu66), the selected residues Val3, Arg64 and Val66 represent the better combination for coulombic and non-coulombic stabilization of *Bs*-CspB. The triple mutant Glu3Val/Thr64Arg/Glu66Val shows the same salt profile as the *Bs*-CspB variant containing the residues from *Bc*-Csp, Glu3Arg/Thr64Val/Glu66Leu,[13] but overall it is 1.2 kJ/mol more stable due to better hydrophobic interactions.

To reach a salt profile like that for *Bc*-Csp, the positive charge could be contributed by the residue at position 64, and by residue 2, as in the Leu2Arg/Glu3Val/Glu66Val mutant (variant 10 in Table 1, Figure 4(b)). This electrostatic stabilization is remarkable, regarding the fact that, overall, the Leu2Arg mutation is destabilizing. Arg2 and

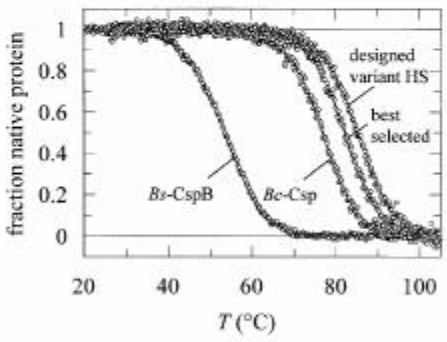


Figure 6. Thermal stabilities of *Bs*-CspB, *Bc*-Csp, the best selected variant Sel and the designed variant HS. The fractions of native protein as obtained after a two-state analysis of the data from thermal unfolding transitions are shown as a function of temperature. The experimental conditions are given in Table 1.

Arg64 differ considerably in the location and orientation of their side-chains (Figure 5), but still their influences on the coulombic interactions in *Bs*-CspB are virtually identical. This again suggests that most of the contributions of the charge mutations are non-local in nature.

Design of a hyperstable variant

We used the information about the individual contributions to stability (Table 1) to design a variant with only four amino acid differences from wild-type *Bs*-CspB. Gly67 does not contribute to stability, and Arg2 destabilizes. Therefore, in the selected variant these two residues were reverted back to Ala67 and Leu2, as in the wild-type protein. The resulting quadruple mutant with Val3, Val66, Leu46, and Arg64 (variant HS, Table 1) shows a t_M value of 85.6 °C, 31.6 deg. C higher than that of the wild-type protein (Figure 6), and its stability is increased by 21.4 kJ/mol. It is thus considerably more stable than the selected variant and even more stable than the homologous cold-shock protein from the hyperthermophilic bacterium *Thermotoga maritima*.[15]

Discussion

By using the *Proside* technique,[7] we had selected stabilized forms of *Bs*-CspB from a combinatorial library of variants in which six sequence positions had been randomized.[10] These positions were chosen on the basis of a sequence comparison with the naturally thermostable homolog *Bc*-Csp, but the most stable of the selected variants (with a t_M value 28.2 deg. C higher than that of the wild-type protein) differed from *Bc*-Csp at all six positions.

We have now analyzed the individual contributions to stability of the selected residues, and we assigned the contributions that could be abolished by adding 2 M NaCl to additional coulombic interactions. This treatment would not account for coulombic contributions that persist in 2 M NaCl. It is very difficult to determine experimentally whether such unscreenable ionic interactions exist in proteins.

The contributions to stability that are still present in 2 M NaCl were assumed to reflect primarily the improvement of hydrophobic interactions. This is probably also an oversimplification for positions that are occupied by charged or hydrogen-bonded residues in the wild-type protein or in the stabilized variant. In these cases, changes in the self-contribution to electrostatic energy (changes in the hydration of charges or in hydrogen bonding) might also contribute to salt-independent stabilization. To further dissect the electrostatic contributions to stability it would be necessary to measure the p*K* values of the respective residues in the folded and the unfolded proteins. Unfortunately, this is not possible for wild-type *Bs*-CspB,

because it aggregates below pH 4.5 and unfolds above pH 9.0.

Our analysis shows that not all of the selected amino acids contribute to the increase in stability. Residue 67 (at the carboxy terminus) is unimportant, and the Arg residue selected at position 2 is slightly destabilizing overall, although it improves the coulombic interactions. The remaining four residues, Val3, Leu46, Arg64 and Val66, all contribute strongly to the high degree of stability of the selected variant.

The coulombic contributions of the replacements that introduce or abolish charges (Leu2Arg, Glu3Val, Thr64Arg, and Glu66Val) are slightly interdependent, evidently because coulombic interactions are far-reaching. The salt-independent contributions to stability of the individual selected residues are additive, except for Leu46, which probably engages in a hydrophobic interaction with Val66 in the stabilized variant (cf. Figure 2).

The selected variant with the greatest stability differed from the naturally thermostable homolog *Bc*-Csp at all randomized positions, which, at first glance, suggested that selections in nature and in the laboratory are governed by different principles. A closer inspection of the results in Table 1 reveals that this is not the case. Both the naturally thermostable protein and the variant selected by *Proside* avoid negatively charged residues at position 3 and 66, as present in wild-type *Bs*-CspB, and prefer an aliphatic residue at position 66, which is Leu in *Bc*-Csp and Val in the most stable of the selected variants. Leu residues have been found at this position in several other stabilized variants during the *in vitro* selections.[10]

It seemed surprising that a positively charged residue was not found for position 3, although the single Glu3Arg mutation stabilized *Bs*-CspB by more than 11 kJ/mol.[12] Instead, a positively charged residue was present at position 64 in all selected variants after the final *Proside* round.[10] Positions 3 and 64 are near each other (their C^{β} atoms are separated by 10 Å, Figure 2), and we suggest that, indeed, a positive charge in this region is very important for stability, but its exact location seems to be less important. The positive charge can be provided by residues at position 3, as in *Bc*-Csp, or at positions 2 or 64, as in the selected variants. The triple mutant cycle for Arg64 revealed that this residue compensates for the local repulsion between Glu3 and Glu66 in wild-type *Bs*-CspB, but shows no direct energetic interaction with either of these Glu residues. This confirms and extends our previous suggestion, that the positively charged guanidinium group of Arg3, as present in the thermostable *Bc*-Csp, contributes to stability not by engaging in a pairwise coulombic interaction with a specific carboxyl group but rather by generally improving the charge distribution in this particular region of the protein surface.

We suggest that the high mobility of surface-exposed residues, combined with the long-range nature of coulombic interactions provide the key for understanding our results. The high-resolution structures of *Bc*-Csp and of five variants[16] demonstrate that, in fact, the guanidinium group of Arg3 can adopt many different orientations. Crystal structures are not available for the selected variants. Modeling of *Bs*-CspB variants with Arg residues at positions 2 or 64 suggests that the side-chains of these residues are highly mobile as well. The spaces accessible for the guanidinium groups of Arg2, Arg3, and Arg64 do not overlap on the surface of *Bs*-CspB (Figure 5), which might explain the small differences in the magnitudes of their contributions to stability. But, remarkably, the general effects on electrostatics caused by these Arg residues at different positions are virtually identical, due to their long-range coulombic interactions with other charged residues. Charged groups at the tip of a flexible chain, as in the case of Lys or Arg, can rearrange rapidly to adopt the conformation of lowest Gibbs free energy in the electrostatic field imposed on them by the other charges on the protein surface. Perhaps more importantly, they can respond to local fluctuations in this field by fast re-orientation. This description is compatible with the concept of ion networks, rather than with that of localized ion pairs.[17] For ion pair formation, a high mobility of the charged groups is unfavorable, because it increases the entropic cost of the interaction.

Only six positions could be probed by our *in vitro* selection technique, and only a few of the many combinations of stabilizing residues could be analyzed by directed mutagenesis. Nevertheless, our results point to a common strategy that underlies both natural and *in vitro* evolution of the cold-shock proteins. In both cases, electrostatic interactions are improved by an optimization of the surface charge distribution; in particular, by disrupting a patch of negative charge density. This local effect as well as the correlated decrease in the net charge of *Bs*-CspB contribute to stabilization, in accordance with current concepts on the contributions of charged groups to protein stability (C. Brooks & B. Dominy, personal communications).[17–23]

Hydrophobic interactions near the surface are improved in natural and in *in vitro* evolution by the selection of small aliphatic residues, such as Leu or Val. By using a computer algorithm, Mayo and co-workers showed that protein stability can be increased by introducing Leu or Ile residues at "boundary" positions that are close to the protein surface.[24]

This leaves little doubt that surface positions are good candidates for optimizing protein stability. Many different combinations of residues can be found, and the contributions to stability of the individual residues are largely additive. The steric restrictions are usually low, because surface-exposed side-chains tend to be mobile, and therefore, unlike in the protein core, deleterious effects of surface substitutions are hardly observed.

Materials and Methods

The genes for CspB variants were PCR-amplified with oligonucleotide primers coding for the mutations at the gene termini (comprising positions 2, 3, 64, 66 and 67) and single-stranded phage DNA as template, and cloned into the expression plasmid pET11a. The proteins were overproduced and purified as described for wild-type *Bc*-Csp[14] and *Bs*-CspB[25] with minor modifications. Thermal unfolding transitions of 4 μM protein were measured in 0.1 M sodium cacodylate/HCl (pH 7.0) in the presence of 0–2.0 M NaCl by circular dichroism at 222.6 nm and analyzed as described.[14] The reversibility of unfolding was examined by rapid heating to a temperature near the end of the transition, followed by rapid cooling and a second heating.

Acknowledgments

We thank T. Oas and the members of our laboratory for many discussions. This work was supported by grants from 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. *Advan. Protein Chem.* **55**, 161–225.
2. Hellinga, H. W. (1998). Computational protein engineering. *Nature Struct. Biol.* **5**, 525–527.
3. Street, A. G. & Mayo, S. L. (1999). Computational protein design. *Structure*, **7**, R105–R109.
4. Wintrode, P. L. & Arnold, P. H. (2001). Temperature adaptation of enzymes: lessons from laboratory evolution. In *Advances in Protein Chemistry*, vol. 55, pp. 161–225, Academic Press, San Diego.
5. Schmidt-Dannert, C. (2001). Directed evolution of single proteins, metabolic pathways, and viruses. *Biochemistry*, **40**, 13125–13136.
6. Plückthun, A., Schaffitzel, C., Hanes, J. & Jermutus, L. (2001). *In vitro* selection and evolution of proteins. In *Advances in Protein Chemistry*, vol. 55, pp. 367–403, Academic Press, San Diego.
7. Sieber, V., Plückthun, A. & Schmid, F. X. (1998). Selecting proteins with improved stability by a phage-based method. *Nature Biotechnol.* **16**, 955–960.
8. Kristensen, P. & Winter, G. (1998). Proteolytic selection for protein folding using filamentous bacteriophages. *Fold. Des.* **3**, 321–328.
9. Finucane, M. D., Tuna, M., Lees, J. H. & Woolfson, D. N. (1999). Core-directed protein design. I. An experimental method for selecting stable proteins from combinatorial libraries. *Biochemistry*, **38**, 11604–11612.
10. Martin, A., Sieber, V. & Schmid, F. X. (2001). *In vitro* selection of highly stabilized protein variants with optimized surface. *J. Mol. Biol.* **309**, 717–726.
11. Schindelin, H., Marahiel, M. A. & Heinemann, U. (1993). Universal nucleic acid-binding domain revealed by crystal structure of the *B. subtilis* major cold-shock protein. *Nature*, **364**, 164–168.
12. Perl, D., Mueller, U., Heinemann, U. & Schmid, F. X. (2000). Two exposed amino acid residues confer thermostability on a cold shock protein. *Nature Struct. Biol.* **7**, 380–383.
13. Perl, D. & Schmid, F. X. (2001). Electrostatic stabilization of a thermophilic cold shock protein. *J. Mol. Biol.* **313**, 343–357.
14. Mueller, U., Perl, D., Schmid, F. X. & Heinemann, U. (2000). Thermal stability and atomic-resolution crystal structure of the *Bacillus caldolyticus* cold shock protein. *J. Mol. Biol.* **297**, 975–988.
15. Wassenberg, D., Welker, C. & Jaenicke, R. (1999). Thermodynamics of the unfolding of the cold-shock protein from *Thermotoga maritima*. *J. Mol. Biol.* **289**, 187–193.
16. Delbrück, H., Mueller, U., Perl, D., Schmid, F. X. & Heinemann, U. (2001). Crystal structures of mutant forms of the *Bacillus caldolyticus* cold shock protein differing in thermal stability. *J. Mol. Biol.* **313**, 359–369.
17. Karshikoff, A. & Ladenstein, R. (2001). Ion pairs and the thermotolerance of proteins from hyperthermophiles: a traffic rule for hot roads. *Trends Biochem. Sci.* **26**, 550–556.
18. Akke, M. & Forsen, S. (1990). Protein stability and electrostatic interactions between solvent exposed charged side chains. *Proteins: Struct. Funct. Genet.* **8**, 23–29.
19. Shaw, K. L., Grimsley, G. R., Yakovlev, G. I., Makarov, A. A. & Pace, C. N. (2001). The effect of net charge on the solubility, activity, and stability of ribonuclease Sa. *Protein Sci.* **10**, 1206–1215.
20. Grimsley, G. R., Shaw, K. L., Fee, L. R., Alston, R. W., Huyghues-Despointes, B. M., Thurlkill, R. L. *et al.* (1999). Increasing protein stability by altering long-range coulombic interactions. *Protein Sci.* **8**, 1843–1849.
21. Xiao, L. & Honig, B. (1999). Electrostatic contributions to the stability of hyperthermophilic proteins. *J. Mol. Biol.* **289**, 1435–1444.
22. de Bakker, P. I., Hunenberger, P. H. & McCammon, J. A. (1999). Molecular dynamics simulations of the hyperthermophilic protein sac7d from *Sulfolobus acidocaldarius*: contribution of salt bridges to thermostability. *J. Mol. Biol.* **285**, 1811–1830.
23. Sanchez-Ruiz, J. M. & Makhatadze, G. I. (2001). To charge or not to charge? *Trends Biotechnol.* **19**, 132–135.
24. Malakauskas, S. M. & Mayo, S. L. (1998). Design, structure and stability of a hyperthermophilic protein variant. *Nature Struct. Biol.* **5**, 470–475.
25. Schindelin, H., Herrler, M., Willimsky, G., Marahiel, M. A. & Heinemann, U. (1992). Overproduction, crystallization, and preliminary X-ray diffraction studies of the major cold shock protein from *Bacillus subtilis*, CspB. *Proteins: Struct. Funct. Genet.* **14**, 120–124.

Edited by C. R. Matthews

(Received 31 January 2002; received in revised form 14 March 2002; accepted 15 March 2002)

7.4 Teilarbeit D

D

Andreas Martin and Franz X. Schmid.

Evolutionary Stabilization of the Gene-3-protein of Phage fd Reveals the Principles that Govern the Thermodynamic Stability of Two-domain Proteins.

The Journal of Molecular Biology **328**, 863-875 (2003)

doi:10.1016/S0022-2836(03)00359-0 *J. Mol. Biol.* (2003) **328**, 863–875

JMB Available online at www.sciencedirect.com SCIENCE DIRECT®

Evolutionary Stabilization of the Gene-3-protein of Phage fd Reveals the Principles that Govern the Thermodynamic Stability of Two-domain Proteins

Andreas Martin and Franz X. Schmid*

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

The gene-3-protein (G3P) of filamentous phage is essential for their propagation. It consists of three domains. The CT domain anchors G3P in the phage coat, the N2 domain binds to the F pilus of *Escherichia coli* and thus initiates infection, and the N1 domain continues by interacting with the TolA receptor. Phage are thus only infective when the three domains of G3P are tightly linked, and this requirement is exploited by *Proside*, an *in vitro* selection method for proteins with increased stability. In *Proside*, a repertoire of variants of the protein to be stabilized is inserted between the N2 and the CT domains of G3P. Stabilized variants can be selected because they resist cleavage by a protease and thus maintain the essential linkage between the domains. The method is limited by the proteolytic stability of G3P itself. We improved the stability of G3P by subjecting the phage without a guest protein to rounds of random *in vivo* mutagenesis and proteolytic *Proside* selections. Variants of G3P with one to four mutations were selected, and the temperature at which the corresponding phage became accessible for a protease increased in a stepwise manner from 40 °C to almost 60 °C. The N1–N2 fragments of wild-type gene-3-protein and of the four selected variants were purified and their stabilities towards thermal and denaturant-induced unfolding were determined. In the biphasic transitions of these proteins domain dissociation and unfolding of N2 occur in a concerted reaction in the first step, followed by the independent unfolding of domain N1 in the second step. N2 is thus less stable than N1, and it unfolds when the interactions with N1 are broken. The strongest stabilizations were caused by mutations in domain N2, in particular in its hinge subdomain, which provides many stabilizing interactions between the N1 and N2 domains. These results reveal how the individual domains and their assembly contribute to the overall stability of two-domain proteins and how mutations are optimally placed to improve the stability of such proteins.

Keywords: *Proside*; protein domains; directed evolution; protein stability; phage display

*Corresponding author

Introduction

Recently we developed a general method for stabilizing proteins, termed *Proside* (protein stability increased by directed evolution).[1] This method is based on phage display and links the increased protease resistance of stabilized protein variants with the infectivity of a filamentous phage. Infection of *Escherichia coli* by the phage is mediated by its gene-3-protein (G3P), located at one end of the virion in three to five copies. G3P consists of three distinct domains, N1 (68 aa), N2 (131 aa) and CT (150 aa), which are connected by

Abbreviations used: G3P, gene-3-protein of phage fd; N1, N2, CT, the two N-terminal and the C-terminal domains of G3P, respectively; G3P*, a fragment of G3P that consists of domains N1 and N2; GdmCl, guanidinium chloride; $[GdmCl]_M$, midpoint of a GdmCl-induced unfolding transition; CD, circular dichroism; DSC, differential scanning calorimetry; t_M, midpoint of a thermal unfolding transition; ΔG_D, Gibbs free energy of denaturation; ΔH_{vH}, van't Hoff enthalpy of denaturation at t_M; C_p^E, excess molar heat capacity; m, cooperativity value of a denaturant-induced unfolding transition.

E-mail address of the corresponding author: fx.schmid@uni-bayreuth.de

0022-2836/03/$ - see front matter

glycine-rich linkers.[2–4] The CT domain anchors G3P in the phage coat, and the N2 domain initiates infection of *E. coli* by binding to the tip of the F pilus.[5,6] Then, after pilus retraction, the N1 domain interacts with TolA of *E. coli*, and the phage DNA enters the cell.[3,4,7]

Thus, a tight linkage between the domains of G3P is essential for phage infectivity,[8] and this requirement is exploited in the *Proside* technique.[1] A repertoire of sequences coding for the protein to be stabilized is inserted into the gene for G3P between the N2 and CT domains, and the resulting library of recombinant phage is exposed to a protease. The physical linkage between the domains of G3P can be maintained only when the inserted variant of the guest protein is tightly folded and thus inaccessible for the protease. Resistant phage with stable guest proteins are enriched in repeated cycles of *in vitro* proteolysis of the phage library, infection of *E. coli*, and phage propagation.[1] This selection method is generic and does not depend on a specific function of the protein to be stabilized, such as enzymatic activity or ligand binding.

Generally, the selection is most successful when the proteolysis step is performed under conditions where the phage itself is stable, but the guest protein is partially unfolded, e.g. in the presence of a denaturant or at increased temperature. The extent of stabilization that can be reached in *Proside* selections is thus limited by the stability of the phage and, in particular, of its G3P.[1] Wild-type G3P is sensitive to subtilisin and proteinase K at room temperature, but resistant to most other proteases, such as trypsin or chymotrypsin. Above 45 °C G3P becomes sensitive to cleavage by these proteases as well.[1,9] To widen the applicability of *Proside* we have now used this strategy to improve the stability of G3P itself, in the absence of a guest protein, by cycles of *in vivo* mutagenesis and proteolytic selections.

The N1–N2 fragment of G3P forms a bilobal, horseshoe-like structure (Figure 1), in which the larger N2 domain partially wraps around the smaller N1 domain.[3,10,11] N2 is, in fact, composed of two subdomains: a globular part, which resembles N1 in size and in tertiary structure, and a hinge subdomain, which provides an extended surface for interactions between N1 and N2. These two domains are thus connected by a mobile glycine-rich covalent linker (which is invisible in the crystal structures) and by extensive non-covalent interactions that involve a contact area of about 1000 Å^2.[3]

In the course of the *Proside* selections we identified successively four variants of G3P with one to four mutations in the domains N1 and N2. These mutations shifted the onset of the proteolytic susceptibility of the phage in a stepwise fashion by about 20 °C. The increases in proteolytic stability of the intact phage were caused by corresponding increases in the thermodynamic stability of the mutated gene-3-proteins. In particular, N2, the least stable domain of G3P, was strongly stabilized, because three of the four selected mutations occurred in this domain. The results of this directed evolution experiment reflect the principles that govern the stability of two-domain proteins.

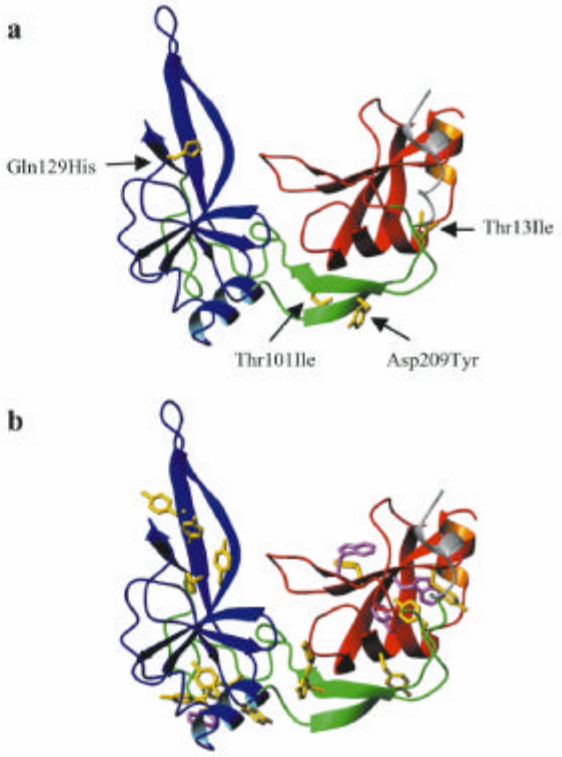


Figure 1. Tertiary structure of G3P* (coordinates from Holliger *et al.*[11]). Domain N1 is shown in red, the globular part of domain N2 in blue, and the hinge subdomain of N2 in green. (a) The side-chains of the four mutations selected with *Proside* are shown in stick representation and labeled. They were modelled by using the Swiss-pdb viewer[32] based on the structure of the wild-type protein. (b) The 14 Tyr residues (yellow) as well as the four Trp residues (pink) of G3P* are shown in stick representation. The Figure was prepared by using MolMol.[33]

Results

Selection of stabilized variants of G3P

To find stabilized variants of G3P we first propagated wild-type phage in the mutator strain *E. coli* XL1Red to introduce random mutations into the phage genome. The resulting library of mutated phage was then exposed to several rounds of selection, each consisting of a proteolysis step (exposure to 0.25 μM chymotrypsin for 15 minutes at different temperatures), followed by infection of *E. coli* XL1Blue, phage propagation and isolation. Thereafter, individual clones of infected *E. coli* cells were grown, the phage isolated, and those with the highest protease resistance were used to

produce a new panel of mutants by again infecting *E. coli* XL1Red. This whole procedure was repeated four times.

After the first mutagenesis and five selection rounds at 45 °C the first stabilizing mutation, T101I in domain N2, was identified. The respective phage was further mutated by passage through *E. coli* XL1Red, and the library was used for six additional selection rounds. During these rounds the temperature was increased in a stepwise fashion to 55 °C, and then a phage with the mutations T101I and D209Y in its G3P was isolated. This phage was again propagated in the mutator strain XL1Red, and, after three selection rounds at 55 °C, a phage with the mutations T13I/T101I/D209Y was identified. The final variant, the quadruple mutant T13I/T101I/Q129H/D209Y, was found after an additional *in vivo* mutagenesis in *E. coli* XL1Red and four selection rounds at 57.5 °C.

The locations of the selected mutations are shown in Figure 1(a). The first two mutations (T101I and D209Y) map to the less stable N2 domain, the third mutation (T13I) is in domain N1, and the fourth mutation (Q129H) again maps to N2. The mutations stabilize the phage against proteolytic cleavage in a manner that reflects their successive appearance in the course of the selection (Figure 2). In combination, the four mutations in G3P shift the onset of the inactivation by chymotrypsin from 40 °C to 60 °C and thus extend the temperature range for *Proside* selections by about 20 degrees.

Thermal stabilities measured by circular dichroism

To examine whether the improved proteolytic stabilities of the selected phage variants originate from an increased conformational stability of their G3P molecules, we purified the selected G3P variants and measured their thermodynamic stabilities. Full length G3P cannot be produced and investigated *in vitro*, because the CT domain, which anchors G3P in the phage coat, renders the protein insoluble. Instead, we used a fragment of G3P that consists of the N1 and N2 domains, and we denote this fragment as G3P*. This fragment represents the part of G3P that protrudes from the phage and mediates the infection of *E. coli* cells; its crystal structure is known (cf. Figure 1).[10,11]

The far-UV circular dichroism (CD) spectra of wild-type G3P* at 25, 60, and 80 °C are shown in Figure 3. Major changes in CD occur between 25 °C and 60 °C. Below 210 nm these changes are complete already at 60 °C, but between 210 nm and 240 nm the CD continues to change when the temperature is increased to 80 °C. These CD changes cannot solely be assigned to changes in the secondary structure, because G3P* is very rich in aromatic residues, which probably contribute to the CD, particularly near 230 nm, where the amide CD is small.

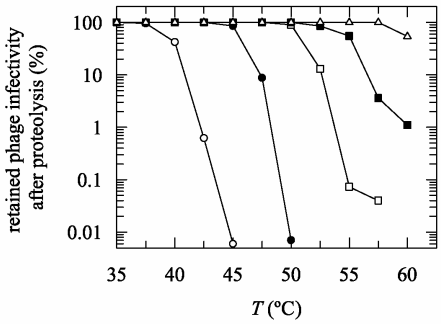


Figure 2. Retained infectivities of the wild-type phage fd (○) and the selected phage with the mutations T101I (●), T101I/D209Y (□), T13I/T101I/D209Y (■), and T13I/T101I/Q129H/D209Y (△) in their G3Ps after *in vitro* proteolysis at different temperatures. The proteolyses were performed with about 10^9 phage in 100 mM potassium phosphate (pH 8.0), with 0.25 μM chymotrypsin for 15 minutes after two minutes pre-incubation. The infectivity after a control incubation at the same temperature without protease is set as 100%.

Thermal transitions measured by CD at 230 nm (Figure 4(a)) and at 210 nm (Figure 4(b)) confirm the qualitative conclusion drawn from the spectra in Figure 3. The transitions of wild-type G3P* and the four selected variants at 230 nm (Figure 4(a)) are indeed biphasic with a highly cooperative transition at low temperature followed by a less cooperative transition at high temperature. All the variants are more stable than the wild-type protein, and the two transitions are affected in a complex fashion by the mutations that were identified in the selections. At 210 nm (Figure 4(b)) the CD changes in an apparent two-state manner and monitors only the first transition. As expected from Figure 3, no further changes are observed at 210 nm above 60 °C (for wild-type G3P*).

To elucidate whether the two steps in the thermal transitions originate from the consecutive unfolding of the two domains of G3P*, we expressed the stabilized N1 domain with the T13I mutation separately. Its unfolding transition

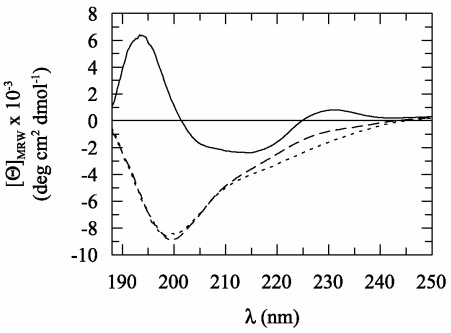


Figure 3. Far-UV CD spectra of wild-type G3P* at 25 °C (continuous line), 60 °C (broken line), and 80 °C (dotted line). The spectra were measured with 10 μM protein in 10 mM potassium phosphate (pH 7.0), at a pathlength of 1 mm and a band width of 1 nm.

monitored by CD at 230 nm (Figure 4(c)) is monophasic and coincides with the second transition of those G3P* variants that also contain the T13I substitution (cf. Figure 4(a)). Attempts to express and purify the globular part of domain N2 without the hinge subdomain have been unsuccessful.

The comparison of the transitions in Figure 4 suggests that the CD at 230 nm indeed reflects the stepwise unfolding of G3P*, in which N2 unfolds first, followed by N1, and that the CD at 210 nm (Figure 4(b)) monitors selectively the unfolding of the less stable N2 domain. The stability data for N2, as derived from a two-state analysis of these transitions, are listed in Table 1. The biphasic transitions at 230 nm (Figure 4(a)) were analyzed according to a sequential three-state model,[12] and the resulting stability data are also given in Table 1. After the stabilization by the T101I and the D209Y mutations N2 approaches N1 in stability. Thus the transitions at 230 nm are almost monophasic and the three-state analyses ambiguous. Therefore, to improve reliability, we used the stability parameters for N2, as obtained separately from the transitions at 210 nm (Figure 4(b)), and held those constant in the three-state analyses for all mutants. The results obtained in this way for the domain N1 are also shown in Table 1. The thermodynamic parameters for the unfolding of N1 and N2 derived from the different analyses agree fairly well, which supports our assignment of the changes in CD to the unfolding reactions of the two domains. Thermal unfolding transitions measured at 2 μM and 15 μM protein were coincident, indicating that G3P* and the isolated domain N1 are monomeric in solution. The unfolding of N2 was reversible when heating was stopped at 55 °C where its own transition had occurred already. It became irreversible, however, after further heating to 90 °C (to unfold both N2 and N1). Domain N1 unfolded reversibly, both when linked with N2 in G3P* and when present as an isolated domain.

The first mutation to be identified in the *Proside* selection, T101I, strongly stabilized the phage against proteolytic inactivation (cf. Figure 2). The stability data obtained with purified G3P* (Table 1) explain this result. The T101I mutation maps to N2, which is the least stable and thus the most protease-sensitive domain of G3P*. Therefore, the stabilization of N2 by 16.5 kJ mol^{-1}, caused by this mutation, improved the proteolytic stability of the entire phage, as observed.

Surprisingly, the subsequent mutation that was selected in N2 (D209Y) left the thermal unfolding transition of G3P* virtually unchanged (Figure 4(a) and (b)). We have no simple explanation for this observation. It might be related to the fact that after the strong stabilization by the first mutation (T101I) domain N2 approaches N1 in stability. This leads to a coupling of the two transitions and an apparent increase in cooperativity, as reflected in the calculated van't Hoff enthalpy of the unfolding of N2 (Table 1). Due to this coupling a potential stabilization of domain N2 might be invisible.

After the first two mutations N2 approached N1 in stability, and, as a consequence, the third stabilizing mutation (T13I) that was selected in the *Proside* experiment occurred in the N1 domain. It increased the t_M value of this domain by about 6°C and thus re-established the three-state character of the unfolding transition of G3P* (Figure 4). The fourth selected mutation (Q129H) again mapped to the N2 domain and brought its t_M to a final

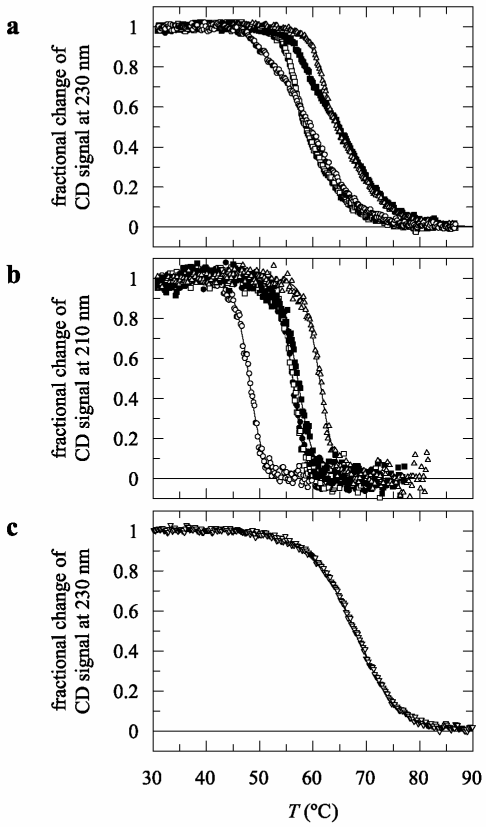


Figure 4. Thermal stabilities of wild-type G3P* and the selected variants. (a) Thermal transitions measured by CD at 230 nm for the wild-type protein (○), and the variants T101I (●), T101I/D209Y (□), T13I/T101I/D209Y (■), and T13I/T101I/Q129H/D209Y (△). The fractional changes of the CD signal as obtained after three-state analyses of the data are shown as a function of temperature. (b) Thermal transitions of the proteins measured by CD at 210 nm. The fractional changes as a function of temperature were obtained from two-state analyses of the data. (c) Thermal transition of the isolated N1 domain with the T13I mutation (▽) measured by CD at 230 nm. The fractional change as obtained after a two-state analysis of the data is shown as a function of temperature. The continuous lines in (a)–(c) show the results of the three-state and the two-state analyses, respectively; the stability data are given in Table 1. All transitions were measured with 4 μM protein in 100 mM potassium phosphate (pH 7.0), at a pathlength of 10 mm.

Table 1. Stability data for wild-type G3P*, the four selected variants, and the isolated domain N1 T13I as derived from thermal unfolding transitions measured by CD and by differential scanning calorimetry (DSC).

	unfolding of domain N2					unfolding of domain N1				
	CD1			DSC		CD2			DSC	
variant of G3P*	t_M (°C)	ΔH_{vH} (kJ·mol^{-1})	$\Delta G_D^{55°C}$ (kJ·mol^{-1})	t_M (°C)	ΔH_{vH} (kJ·mol^{-1})	t_M (°C)	ΔH_{vH} (kJ·mol^{-1})	$\Delta G_D^{65°C}$ (kJ·mol^{-1})	t_M (°C)	ΔH_{vH} (kJ·mol^{-1})
wild-type	48.0 ± 0.1	621 ± 22	-14.2 ± 0.3	48.9 ± 0.1	652 ± 5	59.8 ± 0.1	213 ± 3	-3.3 ± 0.1	61.8 ± 0.1	235 ± 5
T101I	56.2 ± 0.1	626 ± 24	2.3 ± 0.1	57.0 ± 0.1	711 ± 9	59.9 ± 0.7	222 ± 14	-3.4 ± 0.4	59.4 ± 0.1	210 ± 9
T101I/D209Y	56.2 ± 0.1	764 ± 32	2.7 ± 0.1	56.9 ± 0.1	716 ± 12	60.2 ± 0.5	243 ± 15	-3.6 ± 0.4	59.3 ± 0.3	216 ± 8
T13I/T101I/D209Y	57.1 ± 0.1	576 ± 23	3.6 ± 0.1	58.1 ± 0.1	557 ± 7	66.4 ± 0.1	228 ± 5	0.9 ± 0.1	67.0 ± 0.1	269 ± 7
T13I/T101I/Q129H/D209Y	61.2 ± 0.1	640 ± 24	11.3 ± 0.5	62.2 ± 0.1	665 ± 5	66.0 ± 0.3	228 ± 9	0.7 ± 0.2	68.1 ± 0.1	294 ± 9
isolated domain N1 T13I	-	-	-	-	-	67.5 ± 0.0	235 ± 2	1.7 ± 0.0	67.7 ± 0.1	237 ± 2

	CD3			CD3		
wild-type	48.5 ± 0.3	591 ± 72	-12.6 ± 1.3	60.0 ± 0.1	223 ± 6	-3.4 ± 0.1
T101I	56.7 ± 0.1	600 ± 16	3.1 ± 0.2	62.7 ± 0.8	262 ± 23	-1.8 ± 0.7
T101I/D209Y	56.7 ± 0.1	603 ± 17	3.0 ± 0.2	62.0 ± 0.9	253 ± 22	-2.3 ± 0.8
T13I/T101I/D209Y	58.2 ± 0.2	471 ± 21	4.4 ± 0.2	67.6 ± 0.3	292 ± 13	2.2 ± 0.3
T13I/T101I/Q129H/D209Y	61.8 ± 0.1	648 ± 18	12.4 ± 0.4	68.6 ± 0.4	336 ± 23	3.5 ± 0.4

For all proteins the melting temperature (t_M), the van´t Hoff enthalpy of denaturation at t_M (ΔH_{vH}), and the Gibbs free energy of denaturation (ΔG_D) at 55 °C and 65 °C, respectively, are given. The parameters for domain N2 were obtained from a two-state analysis of the changes in CD at 210 nm (**CD1**) and from the first transition at 230 nm after a three-state analysis (**CD3**). The parameters for domain N1 were obtained from the second transition at 230 nm after a three-state analysis with the data for the first transition (domain N2) set fixed as obtained from transitions at 210 nm (**CD2**), and from the second transition at 230 nm after a free three-state analysis (**CD3**). The CD data were analyzed by nonlinear regression as described[12,31], with a fixed heat capacity change ΔC_p of 10000 J·mol^{-1}·K^{-1} for the transition at 210 nm and the first transition at 230 nm, respectively (domain N2), and a ΔC_p of 1000 J·mol^{-1}·K^{-1} for the second transition at 230 nm (domain N1). The transitions monitored by CD were measured with 4 µM protein in 100 mM K phosphate, pH 7.0, at a pathlength of 10 mm.

In the DSC measurements the excess molar heat capacity C_p^E (T) upon unfolding was analyzed by the Levenberg-Marquardt non-linear least-squares method according to a non-two-state model with zero ΔC_p. For all proteins t_M and ΔH_{vH} are given, the calorimetric enthalpy is not listed due to an unreliable baseline for the unfolded protein. The transitions were measured between 20 and 110 °C with a heating rate of 1.5 K·min^{-1} at a protein concentration of 30 µM in 100 mM K phosphate, pH 7.0.

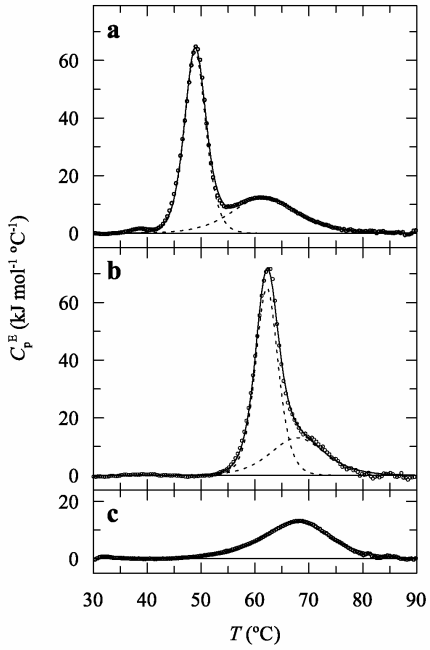


Figure 5. Excess molar heat capacities C_p^E (T) for the unfolding of (a) wild-type G3P*, (b) the most stable variant G3P* T13I/T101I/Q129H/D209Y, and (c) the isolated domain N1 T13I. The open circles represent the experimental data, the continuous lines show the results of the Levenberg–Marquardt non-linear least-squares analyses of the data according to a non-two-state model, and the broken lines in (a) and (b) show the deconvolution into two independent transitions. The derived thermodynamic parameters are given in Table 1. The experiments were performed with 30 μM protein in 100 mM potassium phosphate (pH 7.0), at a scan rate of 1.5 K min^{-1}.

value of about 62 °C, which is 14°C higher than the t_M of N2 in wild-type G3P* (Table 1).

Calorimetric measurements

The thermal unfolding of wild-type G3P and the selected variants was also followed by differential scanning calorimetry (DSC). Two peaks were observed in all DSC experiments. The profiles for wild-type G3P* and for the most stable variant are shown in Figure 5(a) and (b), respectively, together with the transition of the isolated N1 domain with the T13I mutation (Figure 5(c)). The thermal transitions of domain N1 in G3P* and of the isolated domain N1 were reversible. The transition of N2 was reversible to about 80% when heating was stopped after its unfolding in the first transition (at 55 °C), but became irreversible when the protein samples were heated to 90 °C. The calorimetric t_M values and the van't Hoff enthalpies (ΔH_{vH}) for all variants are given in Table 1. The t_M values were independent of protein concentration between 5 μM and 50 μM. Values for the calorimetric enthalpies and for the changes in heat capacity upon unfolding are not given, because the baselines at high temperature are unreliable due to aggregation. The t_M and ΔH_{vH} values from the calorimetric experiments agree very well with the data from the transitions measured by CD (Table 1). They confirm that the unfolding of G3P* occurs in two steps. Domain N2 unfolds first, coupled with domain dissociation, followed by the unfolding of the N1 domain. The data also validate our assumption that the CD at 210 nm (Figure 4(b)) monitors selectively the unfolding of the N2 domain.

GdmCl-induced unfolding transitions

Unfolding of G3P* by GdmCl also occurs in two phases (Figure 6). The CD at 230 nm first increases in a transition that is centered around 1.2 M GdmCl, and then it decreases strongly in a second transition with a midpoint at 2.6 M GdmCl. As in the thermal transitions, we tentatively assign the first transition to the unfolding of the N2 domain and the second transition to the unfolding of N1.

Unlike in thermal unfolding (cf. Figure 4(a)) the CD signal at 230 nm increases in the first transition of GdmCl-induced unfolding. We have no simple explanation for this difference. Control experiments with NaCl show that it is not caused by a salt effect of the ionic denaturant GdmCl on the native G3P*.

G3P* is rich in Tyr and Trp residues, which are unevenly distributed among the two domains (Figure 1(b)). They provide selective probes for measuring the unfolding of the individual domains by fluorescence. N1 contains three Trp and three Tyr residues, N2 contains a single, exposed Trp and 11 Tyr residues, seven of which are in the globular part and the other four in the hinge subdomain. Figure 7 shows the emission spectra of G3P* after selective excitation of the Trp residues at 295 nm (Figure 7(a)) and after excitation of both

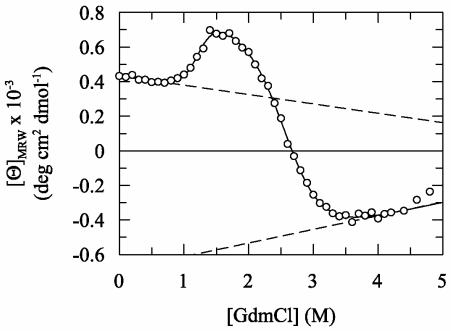


Figure 6. GdmCl-induced equilibrium unfolding transition of wild-type G3P* at 25 °C. The transition was measured by CD at 230 nm with a protein concentration of 6 μM in 100 mM potassium phosphate (pH 7.0). The continuous line represents the analysis of the data based on a three-state unfolding mechanism. The broken lines represent the baselines for the native and the unfolded protein. The transition midpoints, cooperativity values (m), and the ΔG_D values are given in Table 2.

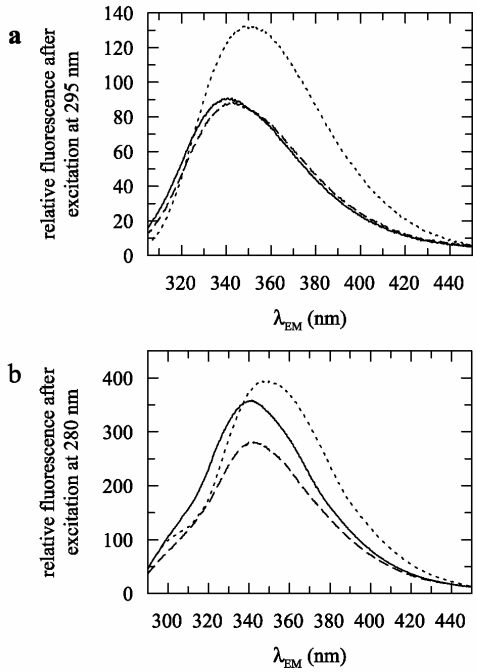


Figure 7. Fluorescence spectra of wild-type G3P* at 0 M (continuous lines), 2.0 M (broken lines) and 5.0 M GdmCl (dotted lines) (a) after excitation of Trp residues at 295 nm and (b) after excitation of both Tyr and Trp residues at 280 nm. The spectra were measured with 2 μM protein in 100 mM potassium phosphate (pH 7.0), at 25 °C.

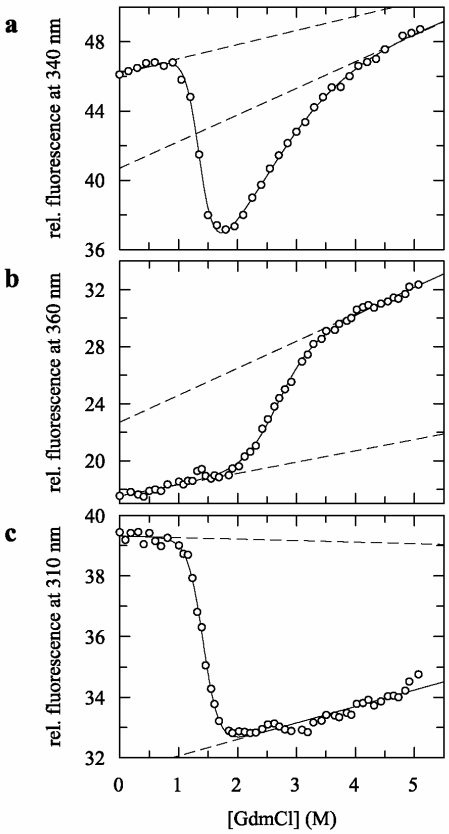


Figure 8. (a) GdmCl-induced equilibrium unfolding transitions of wild-type G3P* at 25 °C measured by protein fluorescence at 340 nm after excitation at 280 nm, (b) by fluorescence at 360 nm after excitation at 295 nm and (c) by fluorescence at 310 nm after excitation at 280 nm. The continuous lines represent the analyses of the data based on a two-state model for the fluorescence at 360 nm and 310 nm (b and c), and based on a three-state model for the fluorescence at 340 nm (a). The broken lines represent the baselines for the native and the unfolded protein. The protein concentration was 0.5 μM in 100 mM potassium phosphate (pH 7.0).

Tyr and Trp residues at 280 nm (Figure 7(b)), in the presence of 0 M, 2.0 M and 5.0 M GdmCl.

When only the Trp residues are excited, the fluorescence remains virtually constant between 0 M and 2.0 M GdmCl and then increases between 2.0 M and 5.0 M GdmCl (Figure 7(a)). When both Tyr and Trp are excited the fluorescence changes in a complex fashion (Figure 7(b)). The emission around 310 nm, which contains large contributions from Tyr residues, decreases between 0 M and 2.0 M GdmCl and then remains almost constant up to 5.0 M GdmCl. The Trp fluorescence around 350 nm first decreases (between 0 M and 2.0 M GdmCl) and then increases again.

To exploit these fluorescence changes for characterizing the unfolding reactions of G3P* we measured GdmCl-induced transitions in three different ways. First, both Tyr and Trp were excited at 280 nm, and Trp emission was measured at 340 nm (Figure 8(a)). Then, the Trp residues were excited selectively at 295 nm, and their emission was measured at 360 nm (Figure 8(b)). Finally, both Tyr and Trp were excited at 280 nm, and Tyr emission was measured at 310 nm (Figure 8(c)).

Two transitions were observed when, after the excitation of both Tyr and Trp, folding was followed by Trp fluorescence (Figure 8(a)). The first transition, between 1.0 M and 1.8 M GdmCl, led to a decrease in Trp emission, the second one, between 2.0 M and 4.0 M GdmCl, led to an increase. Only the second transition was observed when Trp fluorescence was excited selectively at 295 nm and measured at 360 nm (Figure 8(b)). In a reciprocal fashion, only the first transition was observed when Tyr fluorescence at 310 nm was used as a probe (Figure 8(c)).

Together, the data in Figure 8 suggest that the unfolding transition of the less stable N2 domain can be followed separately by Tyr fluorescence at 310 nm, after excitation at 280 nm (as in Figure 8(c)), with little perturbation from the three Tyr residues that are located in domain N1. The transition of domain N1 can be monitored separately by Trp fluorescence at 360 nm, after excitation at 295 nm (as in Figure 8(b)). The single exposed tryptophan residue in N2 (W181) seems to hardly change its emission upon unfolding (when excited at 295 nm). Its fluorescence decreases, however, during the first transition

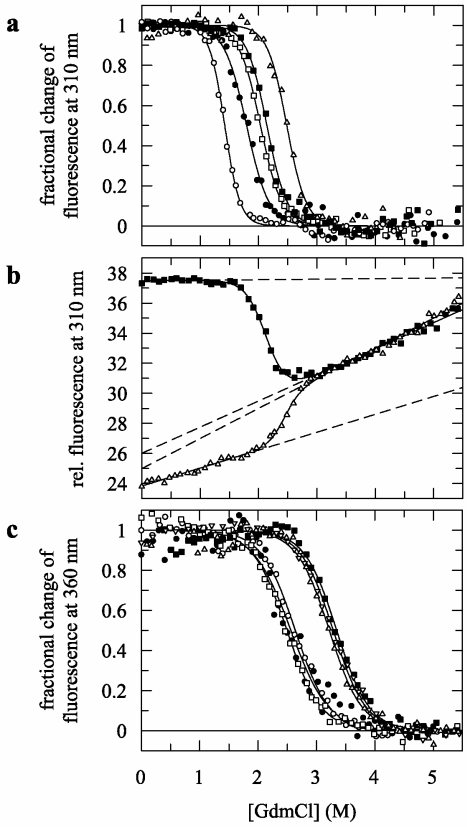


Figure 9. GdmCl-induced equilibrium unfolding transitions of wild-type G3P* (○), the variants G3P* T101I (●), T101I/D209Y (□), T13I/T101I/D209Y (■), T13I/T101I/Q129H/D209Y (△) and the isolated domain N1 T13I (▽) at 25 °C. (a) The fractional changes of Tyr fluorescence at 310 nm after excitation at 280 nm as obtained from two-state analyses of the data are shown as a function of the GdmCl concentration. (b) Transitions of G3P* T13I/T101I/D209Y (■) and G3P* T13I/T101I/Q129H/D209Y (△) measured by protein fluorescence at 310 nm after excitation at 280 nm. (c) The fractional changes of Trp fluorescence at 360 nm after excitation at 295 nm as obtained from two-state analyses are shown as a function of the GdmCl concentration. The continuous lines represent the analyses of the data based on a two-state unfolding mechanism. The broken lines represent the baselines for the native and the unfolded protein. The transition midpoints, cooperativity values (m) and the ΔG_D values at 0 M GdmCl are given in Table 2. The protein concentration was 0.5 μM in 100 mM potassium phosphate (pH 7.0).

when both Tyr and Trp are excited (at 280 nm), as in Figure 8(a). In folded G3P* W181 is very close to three Tyr residues (Y110, Y177, Y180, Figure 1(b)), which can transfer energy to W181. We think that the decrease in Trp fluorescence at 340 nm (Figure 8(a)) originates from a loss of energy transfer from these Tyr residues when the domain N2 unfolds.

For the selected variants of G3P* the GdmCl-induced unfolding transitions of the two domains are shown in Figure 9, and the respective stability parameters are given in Table 2. In the comparison of the variants we focus on the transition midpoints rather than on the ΔG_D values extrapolated to 0 M denaturant, because the m values are difficult to determine when the transitions of N2 and N1 are closely spaced. The successive stabilization of the N2 domain by the first two mutations (T101I and D209Y) led to incremental increases in the transition midpoints from 1.4 M to 2.0 M GdmCl (Figure 9(a), Table 2). Unlike in thermal unfolding (cf. Figure 4(b)), there is a clear stabilizing effect observable for the second mutation (D209Y) when introduced into the T101I variant.

The third of the selected mutations (T13I), which is in N1, leaves the stability of N2 almost unchanged, but the fourth mutation (Q129H), again in N2, increased the transition midpoint of this domain to 2.5 M GdmCl. Interestingly, this mutation decreased the Tyr fluorescence of the folded N2 domain, and, unlike for all other variants, Tyr emission increased upon unfolding (Figure 9(b)). The Q129H substitution introduced a His residue into a pocket in the N2 domain that is lined by three Tyr residues (Y151, Y166, Y168, Figures 1 and 10(c)). It is likely that H129 quenches their fluorescence in the folded, but not in the unfolded state. This specific effect of the Q129H mutation strongly supports our assignment of the Tyr fluorescence changes to the unfolding of the N2 domain.

The transitions of the N1 domains of the selected G3P* variants are shown in Figure 9(c). Only one of the stabilizing mutations, T13I, maps to the N1 domain. Accordingly, two clusters of GdmCl-induced transitions are observed: one for those proteins that do not contain this mutation (the wild-type protein, the T101I and the T101I/D209Y variants) and one for those that do contain it (the T13I/T101I/D209Y and T13I/T101I/Q129H/D209Y variants). The latter transitions match the transition of the isolated N1 domain with the T13I mutation (Table 2). This confirms our assignment of the fluorescence changes at 360 nm to the N1 domain (cf. Figure 8(b)). The stabilization of the N1 domain by the T13I mutation is independent of the mutations in N2, as anticipated, because N2 is largely unfolded already at those denaturant concentrations that are necessary for unfolding of N1.

A comparison of the ΔH_{vH} (Table 1) and m values (Table 2) shows that the unfolding transitions of the N1 domain are much less cooperative than the transitions of N2. This is expected, because N2 is about twice as large as N1. Also, it suggests that the globular part of N2 and the hinge subdomain unfold cooperatively. A more detailed interpretation of the calculated cooperativity parameters is not advisable, because the analysis based on a two-step sequential unfolding mechanism provides a good approximation only when the transitions of the two domains are well separated.[13,14]

Table 2. Stability data for wild-type G3P*, the four selected variants, and the isolated domain N1 T13I as derived from GdmCl-induced equilibrium unfolding transitions

Variant of G3P*	Unfolding of N2 (F1)			Unfolding of N1 (F2)		
	$[GdmCl]_M$ (M)	m (kJ mol^{-1} M^{-1})	$\Delta G_D^{25\,°C}$ (kJ mol^{-1})	$[GdmCl]_M$ (M)	m (kJ mol^{-1} M^{-1})	$\Delta G_D^{25\,°C}$ (kJ mol^{-1})
Wild-type	1.4	18.8	26.6	2.6	8.6	22.6
Wild-type (CD)	1.2	20.1	24.5	2.6	8.8	22.8
T101I	1.8	12.9	23.2	2.6	7.7	19.8
T101I/D209Y	2.0	13.5	27.5	2.5	8.5	21.1
T13I/T101I/D209Y	2.1	14.2	30.5	3.3	9.0	30.0
T13I/T101I/Q129H/D209Y	2.5	14.9	37.2	3.2	8.9	28.6
Isolated domain N1 T13I	–	–	–	3.3	8.9	29.2

The transition midpoints ($[GdmCl]_M$), the cooperativity values (m), and the Gibbs free energies of denaturation at 25 °C ($\Delta G_D^{25\,°C}$) are given for domains N2 and N1. Due to the relatively small changes in fluorescence the accuracy of the thermodynamic parameters is estimated to be about ±10%. F1, The transitions of domain N2 were measured by fluorescence at 310 nm after excitation at 280 nm. F2, The transitions of domain N1 were measured by fluorescence at 360 nm after excitation at 295 nm. The experimental data were analyzed according to two-state models for both domains. The protein concentration was 0.5 μM in 100 mM potassium phosphate (pH 7.0), in 10-mm cells. CD, The transition of the wild-type protein was also monitored by CD at 230 nm at a protein concentration of 6 μM (cf. Figure 6).

Discussion

Stabilized G3P for phage display

The limits of *Proside*, our phage-based selection technique for stabilized proteins, are dictated by the protease resistance of the phage G3P. We applied *Proside* to G3P itself, in the absence of a guest protein, and after successive rounds of mutation and proteolytic selection we identified four variants of G3P that were more resistant towards cleavage by chymotrypsin than the wild-type protein. For the best variant the proteolytic stability increased by about 20 °C, and thus the upper temperature limit of *Proside* could be extended from 40 °C to almost 60 °C. Indeed, the use of phage with the stabilized G3P enabled us to perform a *Proside* selection of the cold shock protein *Bs*-CspB from *Bacillus subtilis* at 57.5 °C. In this selection many stabilized variants of *Bs*-CspB were identified with t_M-values up to 82.2 °C, which is more than 28 degrees above the t_M of the wild-type protein.[15]

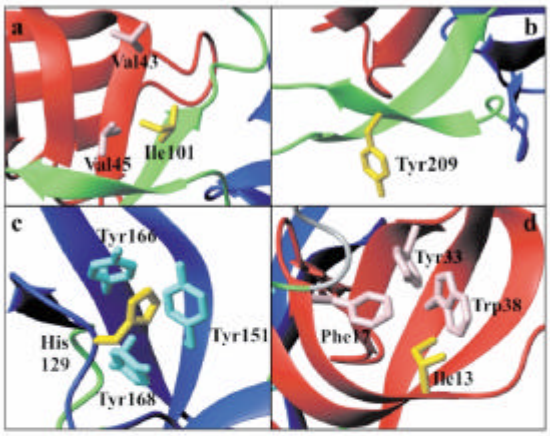


Figure 10. Structural environment of the four selected stabilizing mutations in G3P*. Domain N1 is shown in red, the globular part of domain N2 in blue, and the hinge subdomain of N2 in green. (a) I101 in close proximity to the side-chains of V43 and V45. (b) The surface-exposed Y209 in the hinge subdomain of N2. (c) H129 pointing into a pocket lined by Y151, Y166, and Y168. (d) I13 pointing into a pocket formed by the aromatic residues F17, Y33, and W38. The four mutated amino acid residues (yellow, stick representation) were modelled using the Swiss-pdb viewer,[32] based on the tertiary structure of the wild-type protein (coordinates from Holliger *et al.*[11]). The Figure was prepared by using MolMol.[33]

The increase in the protease resistance of the mutated phage correlates well with the increase in the thermodynamic stability of the purified N1–N2 fragments (G3P*) of the corresponding G3P variants. This is reassuring and confirms our basic assumption that in *Proside* the phage is stabilized against proteolytic inactivation when the guest protein in G3P shows an increased thermodynamic stability. There is, of course, no rigorous correlation between proteolytic and thermodynamic stability, but exceptions are rare, both *in vivo*[16] and in our selections.[15]

The two structural domains of G3P*, N1 and N2 (Figure 1), differ in stability. N1 is small and stable in isolation, N2 is larger, but less stable than N1. N2 consists of two elements: a compact globular subdomain and a hinge that wraps around N1 in the folded protein (cf. Figure 1). The two-step equilibrium unfolding of G3P* reflects this structural organization. In the first transition, the two domains become unlocked and N2 unfolds. The coupling of these two processes leads to a strong increase in enthalpy. The stability of N2 is thus largely determined by the domain–domain interactions, as suggested by the extended interaction area between N1 and N2 in the crystal structure.[10,11] It explains why we were unable to produce a folded fragment of G3P that comprises

only the globular part of N2. In the second transition, N1 unfolds, and this transition is independent of the presence of N2. This is expected, because in G3P* domain N1 unfolds under conditions where N2 is largely unfolded already and the domain–domain interactions are abolished.

The low stability of the N2 domain and the linkage between its unfolding and the dissociation from N1 explains why three out of the four stabilizing mutations map to N2 and why two of them are, in fact, in the hinge that mediates the domain interactions (Figure 1).

Structural basis of the observed stabilizations

The sites of the mutations in the selected variants of G3P are well separated from each other (Figure 1(b)) and therefore, in a first approximation, we assume that their contributions to the Gibbs free energy of denaturation (ΔG_D) are additive. In the following we interpret these contributions on the basis of the crystal structure of G3P*. We are of course aware that changes in ΔG_D upon mutation can be caused by effects on both the native and the unfolded protein,[17] and that a discussion based only on the structure of the folded protein might give an incomplete picture.

T101I was the first substitution to be identified in the selection, and, indeed, it provides the strongest stabilization of both the intact phage (against proteolysis) and the N2 domain in purified G3P* (against unfolding). T101 is in the hinge region of the N2 domain and in wild-type G3P[10] the side-chain of T101 is only 4 Å apart from the valine residues 43 and 45 in domain N1. In the selected variant I101 probably engages in hydrophobic interactions with these two valine residues (Figure 10(a)) and thus improves the interdomain contacts between N2 and N1.

The next mutation that was identified in the selection, D209Y, is also in the hinge region of N2 (Figure 1). The improved proteolytic stability of the phage with this mutation was reflected in an increase in the $[GdmCl]_M$ value of the corresponding G3P* variant, but not in its t_M value. We have no simple explanation for this difference, but it should be noted that small differences in the respective cooperativity parameters m and ΔH_{vH} might account for it. A close inspection of the structure of G3P* (Figure 10(b)) gives no clues on the role of the D209Y mutation.

The Q129H mutation is in the globular part of N2, remote from the domain interface. In the wild-type protein the side-chain of Q129 points into a pocket that is lined by three Tyr residues. A His side-chain is probably accommodated very well in this pocket (Figure 10(c)), and we suggest that the observed stabilization by the Q129H substitution originates from favorable interactions between H129 and these three tyrosine residues. This suggestion is supported by the strong static quenching of the Tyr fluorescence of N2 caused by the Q129H mutation.

In the N1 domain only one stabilizing mutation was selected. It maps to position 13, where a Thr residue was replaced by Ile. As in the case of Q129, the side-chain of T13 points into a pocket that is lined by three aromatic residues, F17, Y33, and W38, and, apparently, this pocket provides more favorable interactions for an Ile than for a Thr residue (Figure 10(d)). The T13I mutation in N1 leads to a slight stabilization of N2 as well. This is expected, because the interaction with the folded N1 domain contributes strongly to the thermodynamic stability of N2.

In the *in vivo* mutagenesis with *E. coli* XL1Red single nucleotide mutations are introduced and thus only a limited number of amino acid residues were available. For T13 the mutations to Arg, Lys, Ile, Pro, Ser or Ala were possible, for T101 the mutations to Ser, Asn, Ile, Pro or Ala, for Q129 to Arg, Pro, Leu, His, Glu or Lys, and for D209 to Ala, Val, Gly, Glu, Asn, His or Tyr.

Both residues 13 and 101 in G3P are near the protein surface, but not exposed, and thus can be classified as boundary.[18] In the selected stabilized variants of G3P at both positions the polar Thr was replaced by the hydrophobic Ile, which is consistent with the stabilizing Ile mutations found for the boundary positions in the β1 domain of the streptococcal protein G.[19] Ile at these positions seems to stabilize through an increased burial of hydrophobic surface area.

The course of the selection

Interestingly, the first mutation in N1 (T13I) was selected only after N2 had been strongly stabilized already by the mutations T101I and D209Y. This is well explained by the principles of our selection. In the wild-type protein N2 is the least stable and most protease-sensitive part of G3P. As a consequence, in the initial rounds of the selection only those protein variants could escape proteolysis that had acquired stabilizing mutations in N2, as observed. With the first two mutations (T101I and D209Y) N2 gained about 17 kJ mol^{-1} in stability. Thus it approached N1 in its stability, and the thermal unfolding transition of the corresponding T101I/D209Y variant of G3P* became highly cooperative. Now, after this initial stabilization of N2, G3P could also be improved by changes in N1, such as the T13I mutation, which occurred at the third stage of the *Proside* selection. In the resulting triple mutant, N1 re-established its energetic advantage over N2 and, as a consequence, the fourth mutation that was selected (Q129I) again mapped to the N2 domain.

Three of the four mutations in G3P (T101I, D209Y, and T13I) introduced cleavage sites for chymotrypsin, the protease that was used in the *Proside* selections. Apparently, the indirect protection against proteolysis by the conformational stabilization dominated over the decrease in protease resistance caused by the introduction of the additional cleavage sites. This suggests that, as

originally anticipated,[1] conformational stability is indeed the most important criterion in *Proside*[1,15] and related proteolytic selection techniques.[9,20]

G3P* as a two-domain protein

Our results on the N1–N2 fragment of G3P illustrate the principles of stabilization of two-domain proteins.[13,14,21–25] Three components contribute to their stability: the stabilities of the two individual domains and the interdomain interaction. In the course of unfolding, first the two domains become unlocked and then they unfold following the order of their intrinsic stabilities. In most cases the free energy of domain interaction is substantial, and the unfolding of the less stable domain becomes thermodynamically coupled with unlocking. Therefore, only two transitions are usually observed experimentally.[13,14,21,22,26] In G3P, N2 is the less stable domain, and its unfolding thus accompanies domain unlocking.

These considerations point to strategies for stabilizing two-domain proteins. The less stable domain can be stabilized by ligand binding[14,22] or by mutagenesis.[24] When the stability of the less stable domain is limited by unlocking then promising candidate residues for stabilization might be located in the domain contact region. Our selection results confirm this. The first and at the same time most stabilizing mutation (T101I) in N2 maps to the domain interface.

Domain unlocking and unfolding of the less stable domain are thermodynamically coupled when the unlocked folded domain is unstable relative to its unfolded form and thus not populated at equilibrium. It should be noted that in this case the overall change in Gibbs free energy is given by the sum of the free energies of unlocking and domain unfolding. Therefore, mutations that improve interactions within the less stable domain will also shift the onset of thermal unfolding to higher temperatures and thus stabilize the entire protein. Such an effect was observed for the Q129H mutation in the N2 domain of G3P, which is remote from the domain contact area. G3P* appears to be well suited for investigating the principles that govern the stability of two-domain proteins, in particular since specific probes are available to follow the folding of the two domains independently.

Materials and Methods

Proside-selection of stabilized G3P

For random mutagenesis and selection of stabilized G3P variants we used a phage derived from fCKCBS,[27] containing wild-type G3P without a guest protein but with a modified linker between the N2 and CT domains.[1] To introduce random mutations, the phage were propagated in the *E. coli* mutator strain XL1Red (Stratagene, La Jolla, USA). Phage were isolated from *E. coli* culture medium by polyethylene glycol precipitation[28] and stored in phosphate buffered saline containing 0.01% (w/v) sodium azide. The *Proside* selections were performed with about 10^9 phage in 50 μl 100 mM potassium phosphate (pH 7.0) with 0.25 μM chymotrypsin for 15 minutes at temperatures between 40 °C and 57.5 °C (increased stepwise in the course of the selection). The reaction mixtures were used for the infection of 5 ml cultures of *E. coli* XL1Blue (Stratagene, La Jolla, USA) with $A_{600} \approx 1.0$, and phage were propagated at 37 °C for eight hours. To determine the retained infectivities of phage for every selection cycle and to measure the proteolytic stabilities of selected phage variants two samples of 25 μl with about 10^9 phage each in 100 mM potassium phosphate (pH 7.0), one in the absence and one in the presence of 0.25 μM chymotrypsin, were incubated for 15 minutes at the respective temperature. Aliquots (10 μl) of these samples were added to 490 μl *E. coli* XL1-Blue ($A_{600} \approx 1.0$) and incubated at 37 °C for 25 minutes for infection. The number of infected cells and thus the number of infectious phage after proteolysis was determined by plating serial dilutions on dYT^{cam} agar.[28]

Expression and purification of G3P* and of the isolated domain N1

For the expression of G3P* (residues 1–217 of mature G3P, plus ProSerGly$(His)_6$) the gene fragments coding for the two N-terminal domains of the wild-type G3P and of the four selected variants were PCR-amplified from the corresponding single-stranded phage DNA with the primers 5′-GGAATTCCATATGGCTGAAACTG TTGAAAGTTG-3′ and 5′-CGGGATCCTTAATGATGAT GATGATGATGCCCGGACGGAGCATTGACAGGAGG-3′. The fragments were cloned into the expression plasmid pET11a (Novagen, Madison, Wisconsin, USA) *via* its *Nde* I and *Bam* HI restriction sites, and the proteins were overproduced in *E. coli* BL21(DE3)pLysS (Stratagene, La Jolla, USA). After cell lysis in lysis buffer (50 mM Tris–HCl, 50 mM NaCl, 10 mM EDTA, 20 mM DTT, pH 8.0) with a microfluidizer and centrifugation the protein was found as inclusion bodies in the pellet. The pellet was washed once with 1 mM EDTA, 1 mg/ml deoxycholate, 20 mM DTT, 0.2 mg/ml hen lysozyme (pH 8.0), three times with the same buffer without lysozyme, three times with lysis buffer, and once with 100 mM Tris–HCl, 1 mM EDTA, 50 mM reduced glutathione (pH 8.5). The inclusion bodies were solubilized in 100 mM Tris–HCl, 6.0 M GdmCl, 1 mM EDTA, 50 mM reduced glutathione (pH 8.0), and stirred for two hours at room temperature to reduce all disulfide bonds. For the oxidative refolding the protein solution was diluted 1:10 in 100 mM Tris–HCl, 1 mM EDTA, 0.5 mM oxidized glutathione (pH 8.5), and stirred for four hours at room temperature.[29] After the separation of aggregates by centrifugation, the refolded G3P* was purified by immobilized metal-affinity chromatography on a Ni-NTA column (elution with 250 mM imidazole) and gel filtration on a Superdex 75 prep grade column (both from Amersham Biosciences, Uppsala, Sweden) with 50 mM Tris–HCl, 100 mM KCl (pH 7.8). Yields were about 60 mg protein per liter.

For the expression of the isolated N1 domain (residues 1–67 of mature G3P, plus Ala$(His)_6$) the G3P-N1 fragment with signal sequence was amplified from the phage DNA with the primers 5′-GGAATTCCATATGAA AAAATTATTATTCGCAATTCC-3′ and 5′-CGGGATCCT TAATGATGATGATGATGATGAGCATTTTCAGGGATA GCAAGCCC-3′, cloned into pET11a, and the protein was

overproduced in *E. coli* BL21(DE3)pLysS. It could be isolated in its soluble oxidized form, and was purified as described for the refolded G3P*. The yield was about 15 mg/l.

Measurements of CD and fluorescence spectra

CD was measured with a JASCO J600A CD spectrometer equipped with a PTC 348 WI peltier element. The spectra of G3P* in the far-UV region (180–250 nm, 1 nm band width) were measured at a protein concentration of 10 μM in 10 mM potassium phosphate (pH 7.0), at a pathlength of 1 mm. Fluorescence spectra were measured with 2 μM protein in 100 mM potassium phosphate (pH 7.0), in 10-mm cells, with an excitation wavelength of 280 or 295 nm, and the emission detected between 290 nm and 450 nm or 304 and 450 nm, respectively. The band widths were generally 5 and 10 nm for excitation and emission, respectively. Hitachi F4010 and F4500 fluorescence spectrometers were used.

GdmCl-induced unfolding transitions

Samples of G3P* (0.5 μM) were incubated for two hours at 25 °C in the presence of 100 mM potassium phosphate (pH 7.0) and varying concentrations of GdmCl. The fluorescence of the samples was measured in 10-mm cells at 310 nm and 340 nm after excitation at 280 nm, or at 360 nm after excitation at 295 nm. The experimental data were analyzed according to two-state models by assuming a linear dependence of the fluorescence emission on the GdmCl concentration. A non-linear least-squares fit with proportional weighting of the experimental data was used to obtain the Gibbs free energy of denaturation ΔG_D as a function of the GdmCl concentration.[30]

For the CD measurements of the GdmCl-induced unfolding transition the protein concentration was 6 μM in 100 mM potassium phosphate (pH 7.0). The CD signal was measured at 230 nm with 1 nm band width and 5 mm pathlength.

Thermal unfolding transitions

The heat-induced unfolding transitions were measured by CD at a protein concentration of 4 μM in 100 mM potassium phosphate (pH 7.0). Samples were heated at a rate of 60 K h^{-1}. The transitions were monitored by the decrease of the CD signals at 210 nm and 230 nm with 1 nm band width and 10 mm pathlength. The experimental data were analyzed on the basis of the two-state approximation for the transition at 210 nm and of the three-state approximation for those at 230 nm by non-linear regression as described,[12,31] with a fixed heat capacity change ΔC_p of 10,000 J mol^{-1} K^{-1} for the transition at 210 nm and the first transition at 230 nm (domain N2), and a ΔC_p of 1000 J mol^{-1} K^{-1} for the second transition at 230 nm (domain N1).

DSC measurements

The DSC measurements were performed with a VP-DSC instrument (MicroCal, Northampton, MA, USA) at protein concentrations between 5 μM and 50 μM in 100 mM potassium phosphate (pH 7.0), with a scan rate of 1.5 K min^{-1} (cell volume 0.523 ml). The measured excess molar heat capacity C_p^E (T) was analyzed by the Levenberg–Marquardt non-linear least-squares method according to a non-two-state model after correction for a progressive baseline (assumption of zero ΔC_p). For the analysis the Origin software provided by MicroCal was used.

Acknowledgements

We thank Volker Sieber and the members of our laboratory for many discussions. This work was supported by grants from the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie.

References

1. Sieber, V., Plückthun, A. & Schmid, F. X. (1998). Selecting proteins with improved stability by a phage-based method. *Nature Biotechnol.* **16**, 955–960.
2. Marvin, D. A. (1998). Filamentous phage structure, infection and assembly. *Curr. Opin. Struct. Biol.* **8**, 150–158.
3. Lubkowski, J., Hennecke, F., Plückthun, A. & Wlodawer, A. (1999). Filamentous phage infection: crystal structure of g3p in complex with its coreceptor, the C-terminal domain of TolA. *Structure*, **7**, 711–722.
4. Riechmann, L. & Holliger, P. (1997). The C-terminal domain of TolA is the coreceptor for filamentous phage infection of *E. coli*. *Cell*, **90**, 351–360.
5. Stengele, I., Bross, P., Garces, X., Giray, J. & Rasched, I. (1990). Dissection of functional domains in phage fd adsorption protein. Discrimination between attachment and penetration sites. *J. Mol. Biol.* **212**, 143–149.
6. Deng, L. W. & Perham, R. N. (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**, 603–614.
7. Click, E. M. & Webster, R. E. (1997). Filamentous phage infection: required interactions with the TolA protein. *J. Bacteriol.* **179**, 6464–6471.
8. Spada, S. & Plückthun, A. (1997). Selectively infective phage (SIP) technology: a novel method for *in vivo* selection of interacting protein–ligand pairs. *Nature Med.* **3**, 694–696.
9. Kristensen, P. & Winter, G. (1998). Proteolytic selection for protein folding using filamentous bacteriophages. *Fold. Des.* **3**, 321–328.
10. Lubkowski, J., Hennecke, F., Plückthun, A. & Wlodawer, A. (1998). The structural basis of phage display elucidated by the crystal structure of the N-terminal domains of G3p. *Nature Struct. Biol.* **5**, 140–147.
11. Holliger, P., Riechmann, L. & Williams, R. L. (1999). Crystal structure of the two N-terminal domains of g3p from filamentous phage fd at 1.9 Angström: evidence for conformational lability. *J. Mol. Biol.* **288**, 649–657.
12. Beasty, A. M., Hurle, M. R., Manz, J. T., Stackhouse, T., Onuffer, J. J. & Matthews, C. R. (1986). Effects of the phenylalanine-22-leucine, glutamic acid-49-methionine, glycine-234-aspartic acid, and glycine-234-lysine mutations on the folding and stability of the alpha subunit of tryptophan synthase from *Escherichia coli*. *Biochemistry*, **25**, 2965–2974.

13. Tsalkova, T. N. & Privalov, P. L. (1985). Thermodynamic study of domain organization in troponin C and calmodulin. *J. Mol. Biol.* **181**, 533–544.
14. Brandts, J. F., Hu, C. Q. & Lin, L.-N. (1989). A simple model for proteins with interacting domains. Application to scanning calorimetry data. *Biochemistry*, **28**, 8588–8596.
15. Martin, A., Sieber, V. & Schmid, F. X. (2001). *In vitro* selection of highly stabilized protein variants with optimized surface. *J. Mol. Biol.* **309**, 717–726.
16. Parsell, D. & Sauer, R. (1989). The structural stability of a protein is an important determinant of its proteolytic susceptibiliy in *Escherichia coli*. *J. Biol. Chem.* **264**, 7590–7595.
17. 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.
18. Dahiyat, B. I., Sarisky, C. A. & Mayo, S. L. (1997). *De novo* protein design: towards fully automated sequence selection. *J. Mol. Biol.* **273**, 789–796.
19. Malakauskas, S. M. & Mayo, S. L. (1998). Design, structure and stability of a hyperthermophilic protein variant. *Nature Struct. Biol.* **5**, 470–475.
20. Finucane, M. D., Tuna, M., Lees, J. H. & Woolfson, D. N. (1999). Core-directed protein design. I. An experimental method for selecting stable proteins from combinatorial libraries. *Biochemistry*, **38**, 11604–11612.
21. Privalov, P. L. (1982). Stability of proteins. Proteins which do not present a single cooperative system. *Advan. Protein Chem.* **35**, 1–104.
22. Wenk, M. & Jaenicke, R. (1999). Calorimetric analysis of the Ca^{2+}-binding beta gamma-crystallin homolog protein S from *Myxococcus xanthus*: intrinsic stability and mutual stabilization of domains. *J. Mol. Biol.* **293**, 117–124.
23. Gast, K., Damaschun, G., Desmadril, M., Minard, P., Muller-Frohne, M., Pfeil, W. & Zirwer, D. (1995). Cold denaturation of yeast phosphoglycerate kinase: which domain is more stable? *FEBS Letters*, **358**, 247–250.
24. Bailey, J. M., Lin, L. N., Brandts, J. F. & Mas, M. T. (1990). Substitution of a proline for alanine 183 in the hinge region of phosphoglycerate kinase: effects on catalysis, activation by sulfate, and thermal stability. *J. Protein Chem.* **9**, 59–67.
25. Jaenicke, R. (1999). Stability and folding of domain proteins. *Prog. Biophys. Mol. Biol.* **71**, 155–241.
26. Rudolph, R., Siebendritt, R., Neslauer, G., Sharma, A. K. & Jaenicke, R. (1990). Folding of an all-beta protein: independent domain folding in gamma-II-crystallin from calf eye lens. *Proc. Natl Acad. Sci. USA*, **87**, 4625–4629.
27. Krebber, C., Spada, S., Desplancq, D., Krebber, A., Ge, L. & Plückthun, A. (1997). Selectively-infective phage (SIP): a mechanistic dissection of a novel *in vivo* selection for protein–ligand interactions. *J. Mol. Biol.* **268**, 607–618.
28. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning—A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA.
29. Rudolph, R., Böhm, G., Lilie, H. & Jaenicke, R. (1997). Folding proteins. In *Protein Function: A Practical Approach* (Creighton, T. E., ed.), pp. 57–100, Oxford University Press, Oxford.
30. Santoro, M. M. & Bolen, D. W. (1988). Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl α-chymotrypsin using different denaturants. *Biochemistry*, **27**, 8063–8068.
31. Mayr, L. M., Landt, O., Hahn, U. & Schmid, F. X. (1993). Stability and folding kinetics of ribonuclease T1 are strongly altered by the replacement of *cis*-proline 39 with alanine. *J. Mol. Biol.* **231**, 897–912.
32. Guex, N. & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, **18**, 2714–2723.
33. Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55. See also pp. 29–32.

Edited by C. R. Matthews

(Received 15 November 2002; received in revised form 5 March 2003; accepted 6 March 2003)

7.5 Teilarbeit E

E

Andreas Martin and Franz X. Schmid.
The Folding Mechanism of a Two-domain Protein: Folding Kinetics and Domain Docking of the Gene-3-protein of Phage fd.
The Journal of Molecular Biology **329**, 599-610 (2003)

doi:10.1016/S0022-2836(03)00433-9 *J. Mol. Biol.* (2003) **329**, 599–610

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

The Folding Mechanism of a Two-domain Protein: Folding Kinetics and Domain Docking of the Gene-3 Protein of Phage fd

Andreas Martin and Franz X. Schmid*

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

*Corresponding author

The gene-3 protein (G3P) of filamentous phages is essential for the infection of *Escherichia coli*. The carboxy-terminal domain anchors this protein in the phage coat, whereas the two amino-terminal domains N1 and N2 protrude from the phage surface. We analyzed the folding mechanism of the two-domain fragment N1-N2 of G3P (G3P*) and the interplay between folding and domain assembly. For this analysis, a variant of G3P* was used that contained four stabilizing mutations (IIHY-G3P*). The observed refolding kinetics extend from 10 ms to several hours. Domain N1 refolds very rapidly (with a time constant of 9.4 ms at 0.5 M guanidinium chloride, 25 °C) both as a part of IIHY-G3P* and as an isolated protein fragment. The refolding of domain N2 is slower and involves two reactions with time constants of seven seconds and 42 seconds. These folding reactions of the individual domains are followed by a very slow, spectroscopically silent docking process, which shows a time constant of 6200 seconds. This reaction was detected by a kinetic unfolding assay for native molecules. Before docking, N1 and N2 unfold fast and independently, after docking they unfold slowly in a correlated fashion. A high energy barrier is thus created by domain docking, which protects G3P kinetically against unfolding. The slow domain docking is possibly important for the infection of *E. coli* by the phage. Upon binding to the F pilus, the N2 domain separates from N1 and the binding site for TolA on domain N1 is exposed. Since domain reassembly is so slow, this binding site remains accessible until pilus retraction has brought N1 close to TolA on the bacterial surface.

Keywords: protein folding; folding kinetics; gene-3 protein; domain docking; pilus

Introduction

Infection of *Escherichia coli* by Ff filamentous phages (such as fd or M13) is initiated by two successive binding steps. The phage attaches first to the F pilus as the primary receptor, and then to the carboxy-terminal domain of the periplasmic protein TolA, the coreceptor for phage infection.[1–3] Both these early events in infection are mediated by the phage gene-3 protein (G3P), which exists in three to five copies at the tip of the phage particle. G3P consists of three domains, the two amino-terminal domains N1 (68 amino acid residues) and N2 (131 residues), and the carboxy-terminal domain CT (150 residues). The domains are connected by glycine-rich linkers.[3–6]

The CT domain of G3P is partly embedded in the phage coat,[4,7] but the amino-terminal domains

Abbreviations used: G3P, gene-3 protein of phage fd; N1, N2 and CT, the two N-terminal and the C-terminal domains of G3P, respectively; G3P*, a fragment of G3P that consists of domains N1 and N2; IIHY-G3P*, G3P* containing the four stabilizing mutations T13I, T101I, Q129H, and D209Y; GdmCl, guanidinium chloride; $[GdmCl]_M$, midpoint of a GdmCl-induced unfolding transition; t_M, midpoint of a thermal equilibrium unfolding transition; ΔG_D, Gibbs free energy of denaturation; m, cooperativity value of a denaturant-induced equilibrium unfolding transition; m_{NU}, and m_{UN}, kinetic m-value of unfolding and refolding, respectively; τ, and λ, apparent time constant and rate constant, respectively, of a folding reaction.

E-mail address of the corresponding author: fx.schmid@uni-bayreuth.de

0022-2836/03/$ - see front matter

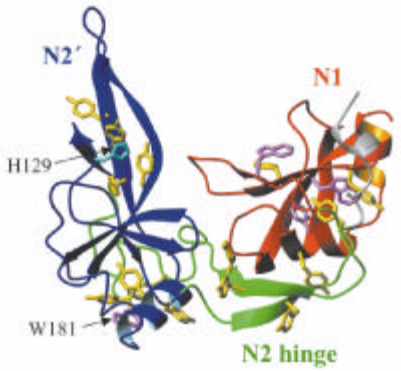


Figure 1. Tertiary structure of G3P* (coordinates from Holliger *et al.*[9]). Domain N1 is shown in red, the globular subdomain of N2 in blue, and the hinge subdomain of N2 in green. The 15 Tyr residues (yellow) as well as the four Trp residues (pink) are shown in stick representation. His129 and the single tryptophan residue Trp181 in N2 are labelled. The Figure was prepared by using MOLMOL.[34]

N1 and N2 protrude from the phage surface. Together, these two domains form a bilobal, horseshoe-like structure (Figure 1), in which the larger N2 domain wraps around the smaller N1 domain.[6,8,9] We denote the N1–N2 fragment as G3P*.

N2 is, in fact, composed of two subdomains: a globular part, which resembles N1 in size and in tertiary structure, and a hinge subdomain, which comprises two β-strands and provides most of the contacts with N1. This hinge subdomain was defined on the basis of two different crystal structures of G3P*, which suggest a rigid-body rotation of N2 relative to N1 about a "hinge" centered around residues G99 and S208.[9] The two domains are thus connected by the glycine-rich covalent linker (which is mobile and not resolved in the crystal structures), and by extensive non-covalent interactions that involve a contact area of about 1000 Å^2.[6]

Most proteins are, in fact, composed of domains, and these domains are regarded as structural, functional, or folding units.[10–12] The folding of proteins with several domains is more complex than the folding of single-domain proteins, because, in addition to the folding of the individual domains, it includes their assembly. Domain folding and assembly can be linked kinetically, in particular when individual domains are unstable in isolation. The crystal structure of G3P* reveals both a clear separation into two structural domains and an extensive domain interface. Therefore, we use G3P* as a model for investigating the folding mechanism of two-domain proteins and the interrelationship between domain folding and domain docking. In these experiments the three intradomain disulfide bonds of G3P* are left intact.

The domain architecture of G3P* provides an interesting opportunity for studying the mechanisms of protein folding, and it is essential for the infection of *E. coli* by the phage. Binding of the N2 domain to the tip of the F pilus initiates infection,[4,13] then the pilus retracts by an unknown mechanism[14,15] and brings the bound phage into close proximity of the cell surface. After pilus binding, the non-covalent interactions between N1 and N2 are broken, the two domains separate and expose the TolA binding site, which is located on domain N1. This binding site for TolA involves a surface region of N1 that is used for its interaction with N2 and thus it is inaccessible in the closed form of G3P*.[6] The domains must therefore separate before the infection can continue by the binding of N1 to the coreceptor TolA,[2] and this might depend on a local unfolding of the hinge subdomain of N2.

Here, we have analyzed how the domains of G3P* fold, how the interdomain interactions are formed and broken, and how domain folding and docking are coupled kinetically. The results reveal first insights into the domain folding and assembly reactions of G3P*, and suggest how these reactions might contribute to the mechanism of infection by filamentous phages.

Results

Guanidinium chloride (GdmCl)-induced unfolding transitions

For measuring the folding kinetics of the N1–N2 fragment of G3P, we employed a variant with four stabilizing mutations (T13I, T101I, Q129H and D209Y) that originated from an *in vitro*-selection for G3P mutants with increased protease resistance.[16] We denote it as IIHY-G3P*. The phage with these four mutations in its G3P is fully functional during the infection of *E. coli*. IIHY-G3P* was used to investigate the folding mechanism, because the unfolding transition of this stabilized variant is shifted to higher concentrations of denaturant, which improves the solubility of partially folded intermediates.

IIHY-G3P* contains four Trp and 15 Tyr residues, which are distributed unevenly among its two domains. Twelve Tyr residues are located within domain N2, and the unfolding of this domain can be followed by Tyr fluorescence. In the native state, H129 binds into a pocket that is lined by three Tyr residues and thus quenches their fluorescence.[16] Upon unfolding of N2, the quenching is relieved, and the corresponding increase in emission at 310 nm after excitation at 280 nm probes selectively the unfolding of this domain (Figure 2(a)).

Three Trp residues are in the N1 domain, and only one, W181, is in N2 at a solvent-exposed position (Figure 1). A previous analysis indicated that W181 shows the same fluorescence in native and unfolded IIHY-G3P* when excited at 295 nm (to abolish energy transfer from Tyr residues).[16] Thus, Trp emission at 360 nm after excitation at

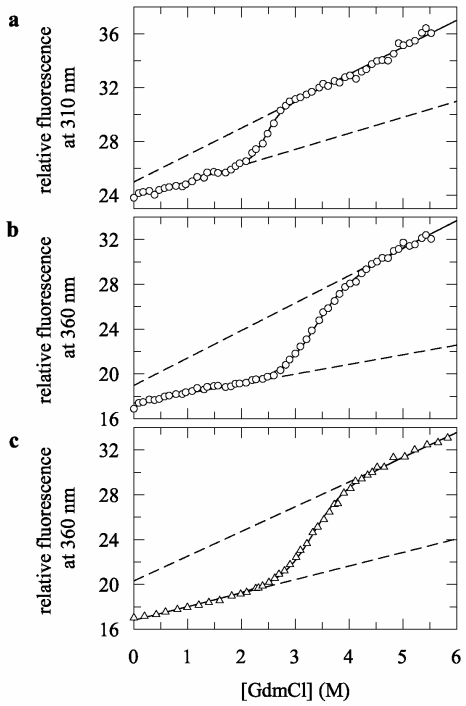


Figure 2. GdmCl-induced equilibrium unfolding transitions of IIHY-G3P* at 25 °C as measured by (a) the fluorescence at 310 nm after excitation at 280 nm to follow unfolding of N2, and (b) the fluorescence at 360 nm after excitation at 295 nm to follow unfolding of N1. (c) The GdmCl-induced transition of the isolated domain N1 with T13I mutation measured by fluorescence at 360 nm after excitation at 295 nm. The continuous lines represent the least-squares fit analyses of the data based on a two-state model for both the change in fluorescence at 310 nm and at 360 nm, the broken lines represent the baselines for the native and the denatured state, respectively. The transition midpoints and cooperativity values obtained from the two-state analyses are: (a) $[GdmCl]_M = 2.5$ M, $m = 14.9$ kJ mol^{-1} M^{-1} for domain N2; (b) $[GdmCl]_M = 3.2$ M, $m = 8.9$ kJ mol^{-1} M^{-1} for domain N1; and (c) $[GdmCl]_M = 3.3$ M, $m = 8.9$ kJ mol^{-1} M^{-1} for the isolated domain N1 T13I. The protein concentration was 0.5 μM in 100 mM potassium phosphate, pH 7.0.

295 nm monitors the folding processes in the N1 domain (Figure 2(b)).

In folded domain N2, the single exposed W181 is very close to three Tyr residues (Figure 1). They transfer energy to this Trp, and the efficiency of this transfer is reduced upon unfolding. Therefore, when excited at 280 nm, the fluorescence of G3P* at 340 nm changes in two stages in the course of unfolding. It decreases when, upon unfolding of N2, the energy transfer to W181 becomes weaker, and it increases when the three Trp residues within N1 become exposed. This change in two stages is well resolved when the domains N2 and N1 differ strongly in stability, as in wild-type G3P*.[16] In IIHY-G3P* the difference in stability between N1 and N2 is smaller, and this two-stage behavior is less obvious in the equilibrium transition. Still, the Trp fluorescence at 340 nm (after excitation at 280 nm) provides a valuable probe for measuring the kinetics of folding of domain N2 under conditions where N1 is folded.

The analyses of the unfolding transitions of IIHY-G3P* measured by fluorescence at 310 nm (Figure 2(a)) and at 360 nm (Figure 2(b)) according to two-state models yield transition midpoints of 2.5 M GdmCl for domain N2 and 3.2 M GdmCl for domain N1. The cooperativity value (m) for domain N2 (14.9 kJ mol^{-1} M^{-1}) is significantly higher than that for domain N1 (8.9 kJ mol^{-1} M^{-1}), as expected, since domain N2 is about twice as large as N1. A fragment comprising only the N1 domain (residues 1–67) with the stabilizing T13I mutation was expressed separately, and in the GdmCl-induced equilibrium unfolding transition it shows the same midpoint and cooperativity value as the N1 domain of intact IIHY-G3P* (Figure 2(c), $[GdmCl]_M = 3.3$ M, $m = 8.9$ kJ mol^{-1} M^{-1}).

Folding kinetics of IIHY-G3P*

The equilibrium transitions in Figure 2 indicate that the changes in Trp and Tyr fluorescence of IIHY-G3P* provide rather selective probes for following the folding of domains N1 and N2, respectively. Accordingly, we measured the kinetics of unfolding and refolding by Tyr fluorescence (emission at 310 nm after excitation at 280 nm) to follow the reactions of N2 and by Trp fluorescence (emission at 360 nm after excitation at 295 nm) to follow the reactions of N1. In addition we used the energy transfer to W181 (after excitation at 280 nm) to probe the unfolding of N2 in the stopped-flow measurements under conditions where N1 remains folded (up to 3.0 M GdmCl, cf. Figure 2(b)).

The two domains refold with strongly different rates. Refolding of N2 in 0.5 M GdmCl is a slow reaction, which consists of two phases with time constants of 7.0 seconds and 42 seconds (Figure 3(a)). Refolding of N1 is 500-fold faster and could be detected only after stopped-flow mixing. It is well approximated by a single-exponential function, which gives a time constant of 9.4 ms (Figure 3(b)).

Surprisingly, the unfolding reactions of the two domains of IIHY-G3P* are both slow and follow identical kinetics with a time constant of 71 seconds (at 4.5 M GdmCl, Figure 3(c)). This suggests that they are governed by a common rate-limiting step. To find out, whether this rate-limiting event of unfolding occurs in the N1 or the N2 domain, we performed a stopped-flow double-mixing experiment. N1 refolds 500-fold faster than N2. Therefore, by a short refolding pulse, it is possible to generate transiently a form of IIHY-G3P* in which N1 is already folded, but N2 is still unfolded. For this transient form, the unfolding of N1 can be measured independently of N2. In the first step of the double-mixing experiment, domain N1 was folded by a one second refolding pulse at

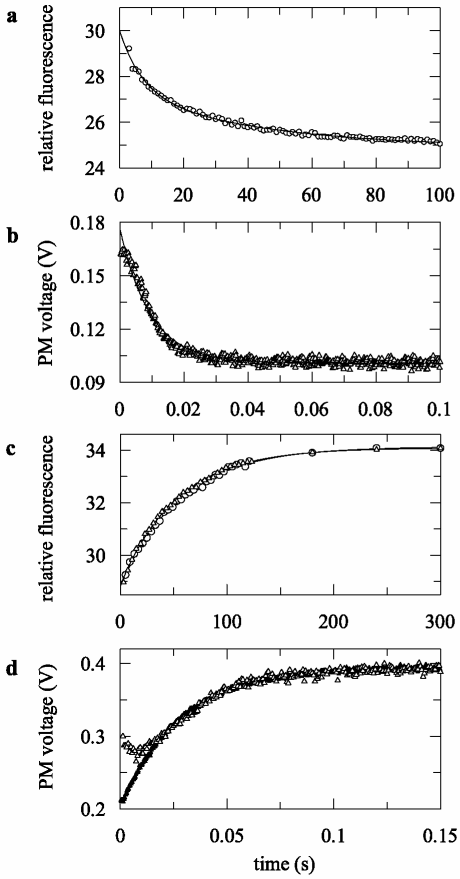


Figure 3. Refolding and unfolding kinetics of domains N2 and N1. (a) Biexponential refolding kinetics (τ_1 = seven seconds, τ_2 = 42 seconds) of domain N2 within IIHY-G3P* at 0.5 M GdmCl, starting from the denatured protein (5.0 M GdmCl) and measured by Tyr fluorescence at 310 nm after excitation at 280 nm. (b) Monoexponential refolding kinetics (τ = 9.4 ms) of domain N1 within IIHY-G3P* at 0.5 M GdmCl, starting from the denatured protein (5.0 M GdmCl) and measured after stopped-flow mixing by the change in fluorescence above 320 nm after excitation at 295 nm. (c) Identical unfolding kinetics (τ = 71 seconds) for domain N2 (○) and domain N1 (△) within IIHY-G3P* at 4.5 M GdmCl, starting from the native protein. The reactions were followed by the change in Tyr fluorescence (excitation at 280 nm, emission at 310 nm) for domain N2 and by the change in Trp fluorescence (excitation at 295 nm, emission at 360 nm) for domain N1. The two kinetic curves are normalized such that the initial and final values coincide. (d) (△) The unfolding kinetics of domain N1 within IIHY-G3P* (τ = 28 ms) at 4.5 M GdmCl, after a one second refolding pulse at 0.5 M GdmCl. (▲) The unfolding kinetics of the isolated domain N1 T13I at 4.5 M GdmCl (τ = 28 ms). The reactions were followed by the change in fluorescence above 320 nm (excitation at 295 nm) after stopped-flow double-mixing for domain N1 within IIHY-G3P* and after stopped-flow single-mixing for the isolated domain N1 T13I. The deviations in the first 10 ms (△) are probably due to an artefact caused by stopped-flow double-mixing. The continuous lines in (a)–(d) show the curve fits for single and double-exponential kinetics, the derived time constants (τ) are given above in parentheses. The protein concentration for the measurements was 0.5 μM in 100 mM potassium phosphate, pH 7.0.

0.5 M GdmCl. Then, in the second step, the concentration of GdmCl was increased to 4.5 M to initiate unfolding of N1, which was monitored by Trp fluorescence. The result is striking (Figure 3(d)). When linked with unfolded N2, N1 unfolds very fast with a time constant of 28 ms, 2500-fold faster than in the unfolding of native IIHY-G3P*, where N1 is linked with folded N2 (Figure 3(c)). Moreover, the unfolding kinetics coincides with the kinetics observed for the isolated domain N1 T13I (Figure 3(d)). This suggests that N1 cannot unfold as long as N2 is folded, and that the rate-limiting step of unfolding either occurs in the N2 domain or is governed by domain disassembly.

Denaturant-dependence of the folding kinetics

The unfolding and refolding kinetics of IIHY-G3P* were measured between 0 M and 6 M GdmCl. The observed rate constants are shown in Figure 4(a) as a function of the concentration of GdmCl. Refolding of N1, as measured after stopped-flow mixing by Trp fluorescence, is a very fast monoexponential reaction, and its rate constant is only weakly dependent on the concentration of GdmCl ($m_{UN} = -0.98\ M^{-1}$; Figure 4(a)). In contrast, the refolding of N2 (as measured by Tyr fluorescence) is slow and biphasic. The two observed rate constants differ about tenfold and decrease approximately in parallel with increasing concentration of GdmCl. Folding is slowest near 2.5 M GdmCl, which is close to the midpoint of the equilibrium transition of the N2 domain (cf. Figure 2(a)).

The relative amplitude of the faster phase of N2 refolding is about 70% between 0 M and 1 M GdmCl (Figure 4(b)) and drops to about 30% between 1 M and 2 M GdmCl. Concomitantly, the slower phase increases in amplitude and becomes the dominant phase above 2 M GdmCl. In parallel with the loss in amplitude, the apparent rate constant of the faster phase proceeds through a minimum near 1.5 M GdmCl (Figure 4(a)). This correspondence between the minimum in the measured rate and the loss in amplitude indicates that the faster phase in the folding of N2 reflects the formation of a kinetic intermediate that is destabilized between 1 M and 2 M GdmCl.

Unfolding of IIHY-G3P* appears to be a monoexponential reaction for both domains, and its time constant decreases from 6700 seconds at 2.5 M GdmCl to 6.5 seconds at 6.0 M GdmCl ($m_{NU} = 2.59\ M^{-1}$; Figure 4(a)). An additional minor unfolding reaction was detected under some conditions. It showed a small amplitude and a GdmCl-independent rate. We are not sure whether this phase reflects an unfolding event; it could be caused by protein aggregation or by dilution effects.

To measure the unfolding of domain N1 separately, stopped-flow double-mixing experiments (as in Figure 3(d)) were performed to keep the N2 domain unfolded. The observed rate constant of

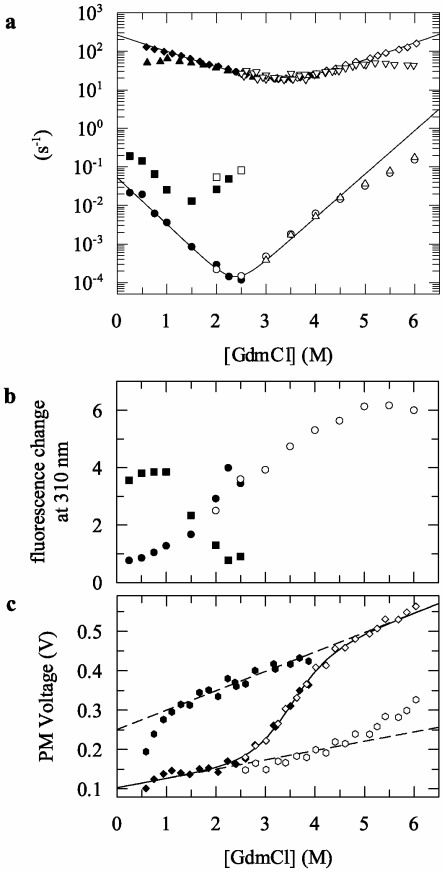


Figure 4. (a) Refolding kinetics (filled symbols) and unfolding kinetics (open symbols) of domain N2 (○, □, ●, ■) and domain N1 (△, ▲, ▽) within IIHY-G3P*, and of the isolated domain N1 T13I (◇, ◆) at 25 °C. The apparent rate constants λ are shown as a function of the concentration of GdmCl. The unfolding kinetics of domain N1 (△) and domain N2 (○, □), starting from the native IIHY-G3P*, and the refolding kinetics of domain N2 (●, ■) were measured by the changes in Trp and Tyr fluorescence, respectively, after manual tenfold dilution of the protein. The fast refolding kinetics of domain N1 within IIHY-G3P* (▲) and the unfolding and refolding kinetics of the isolated domain N1 T13I (◇, ◆) were determined by the change in fluorescence above 320 nm after stopped-flow single-mixing. To measure the fast unfolding of domain N1 in short-time refolded IIHY-G3P* (▽) the unfolded protein was first refolded for one second at 0.5 M GdmCl and then unfolded in a stopped-flow double-mixing experiment. Fits of the fast unfolding and refolding rates of domain N1 and of the slowest unfolding and refolding rates of domain N2 on the basis of linear two-state models are shown by continuous lines. The obtained parameters for domain N1 are $k_{NU}(H_2O) = 0.33\ s^{-1}$, $k_{UN}(H_2O) = 262\ s^{-1}$, $m_{NU} = 1.04\ M^{-1}$, $m_{UN} = -0.98\ M^{-1}$, and $[GdmCl]_M = 3.3$ M. The parameters for N2 are $k_{NU}(H_2O) = 1.5 \times 10^{-7}\ s^{-1}$, $k_{UN}(H_2O) = 0.052\ s^{-1}$, $m_{NU} = 2.59\ M^{-1}$, $m_{UN} = -2.77\ M^{-1}$, and $[GdmCl]_M = 2.4$ M. The final concentration of protein for all measurements was 0.5 μM in 100 mM potassium phosphate, pH 7.0. (b) Fluorescence amplitudes of the two refolding phases (●, ■) and the slow unfolding phase (○) of domain N2 as a function of the concentration of GdmCl. (c) Initial (○, ●) and final (◇, ◆) values of the unfolding (open symbols) and refolding (filled symbols) kinetics of the isolated domain N1 T13I as a function of the concentration of GdmCl. The two-state analysis of the data (assuming that the initial values define the baselines for the native and the unfolded protein, and the final values trace the equilibrium transition) gave values of $m = 6.3$ kJ mol^{-1} M^{-1} and $[GdmCl]_M = 3.5$ M, respectively.

unfolding of N1 depends only weakly on the concentration of GdmCl ($m_{NU} = 1.04\ M^{-1}$), and, in the transition region, it connects smoothly with the refolding kinetics, as measured directly in single-mixing experiments (Figure 4(a)). Together, these kinetic data result in a shallow chevron for N1 that flattens out at low and high concentrations of GdmCl. This chevron coincides well with the chevron of the isolated N1 domain (Figure 4(a)). The good agreement validates the results of the double-mixing experiments, and it supports the assumption that in intact IIHY-G3P* Trp fluorescence probes selectively the folding kinetics of the N1 domain. It confirms our conclusion that in native IIHY-G3P* unfolding of N1 is strongly decelerated, because it is linked kinetically with the slow unfolding of N2. The limbs of the chevron for the isolated N1 domain remain straight and do not flatten below 1 M or above 5 M GdmCl, as observed for N1 when it is a part of intact IIHY-G3P* (Figure 4(a)). At present we have no explanation for this difference.

The initial and the final values of the folding kinetics of the isolated N1 domain are shown in Figure 4(c). The final values trace a reversible transition that coincides with the equilibrium transition in Figure 2(c). The transition midpoints agree with the minima of the chevron plots for domain N1 as a part of IIHY-G3P* and as an isolated fragment (Figure 4(a)). In combination, the equilibrium data in Figure 2 and the kinetic data in Figure 4 show that N1 is the most stable domain of IIHY-G3P*, and that its folding transition is very dynamic, with time constants of equilibration between the native and the unfolded form that are lower than 20 ms. N1 is an independent folding unit, and its stability and folding is not affected by the presence of unfolded N2 and seems to follow a simple two-state mechanism. It should be noted, however, that an additional criterion for a two-state transition is not met: the sum of the kinetic m values:

$$RT(m_{NU} - m_{UN}) = 5.0\ \text{kJ mol}^{-1}\ \text{M}^{-1}$$

is smaller than the m values derived from the end-point analysis (6.3 kJ mol^{-1} M^{-1}; Figure 4(c)) and from the equilibrium unfolding transition of N1 (8.9 kJ mol^{-1} M^{-1}, Figure 2(c)). Together with the loss in refolding amplitude below 1 M GdmCl (Figure 4(c)), this indicates that a rapidly formed intermediate might be present when N1 folds under strongly native conditions.

N2 is less stable than N1, and its refolding is a complex, slow process, which occurs in the time range between five seconds and 6000 seconds. In

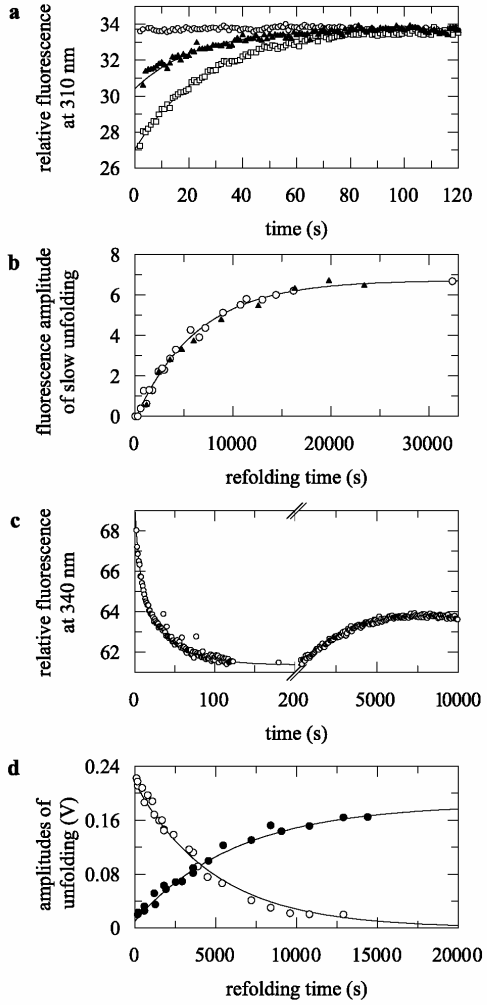


Figure 5. (a) Unfolding kinetics of domain N2 at 5.0 M GdmCl after refolding of denatured IIHY-G3P* at 0.5 M GdmCl for three minutes (○), 70 minutes (▲), and 540 minutes (□). The continuous lines represent the curve fits for monoexponential kinetics ($\tau = 25$ s). The kinetics were followed by Tyr fluorescence after manual double-mixing, at a final concentration of protein of 0.5 μM in 100 mM potassium phosphate, pH 7.0, 25 °C. (b) Amplitudes of the slow unfolding reaction ($\tau = 25$ s, cf. (a)) of domain N2 of IIHY-G3P* (○) at 5.0 M GdmCl and of domain N2 of the H129Q variant (▲) at 4.5 M GdmCl, as a function of the time of refolding at 0.5 M GdmCl. The experimental conditions are described above. The continuous line represents the monoexponential curve fit ($\tau = 6{,}200$ s) of the data for IIHY-G3P*. (c) Unfolding kinetics at 2.0 M GdmCl of the folding intermediate present after three minutes refolding of IIHY-G3P* at 0.5 M GdmCl. The kinetics were followed by Trp fluorescence at 340 nm after excitation at 280 nm, as a specific probe for domain N2. The continuous line shows a triexponential curve fit of the data ($\tau_1 = 5.5$ seconds, $\tau_2 = 43$ seconds, $\tau_3 = 8{,}300$ seconds). (d) Amplitudes of the fast (○, $\tau = 30$ ms) and the slow (●, $\tau = 25$ seconds) unfolding reactions of domain N1 within IIHY-G3P* at 5.0 M GdmCl as a function of the time of refolding at 0.5 M GdmCl. The kinetics were followed by the change in fluorescence above 320 nm after stopped-flow double-mixing. The continuous lines represent monoexponential curve fits, (○, $\tau = 5000$ seconds; ●, $\tau = 6250$ seconds). For all measuremants, the final concentration of protein was 0.5 μM in 100 mM potassium phosphate, pH 7.0, 25 °C.

their unfolding, the two domains are kinetically coupled, and N1 cannot unfold as long as it is locked in its native conformation by the folded N2 domain. Thus, although N2 is less stable than N1, it provides a kinetic protection for N1 and decelerates its unfolding up to 150,000-fold (at 2.5 M GdmCl).

Kinetics of the formation of fully folded IIHY-G3P* molecules

G3P has no catalytic activity, and its N1 and N2 domains do not exert their biological functions until the phage infects an *E. coli* cell. This occurs a long time after phage coat assembly. Therefore, functional assays cannot be used to follow the kinetics of renaturation in real time or to examine whether the product of fluorescence-detected folding is indeed the native protein. Previously, we developed a generic two-step assay for measuring the kinetics of the formation of fully folded protein molecules during a refolding reaction.[17,18] This assay is based on the fact that the protein molecules that have passed beyond the highest activation barrier in their refolding, and thus have reached the native state, are separated from the unfolded state by this energy barrier. They unfold slowly when the conditions are switched to unfolding, because they must cross the high barrier again, now in the reverse direction. Partially folded molecules, however, that have not yet passed the final transition state, unfold rapidly. This translates into a double-mixing procedure. First, unfolded protein is mixed with refolding buffer to initiate refolding. Then, after variable time intervals, samples are withdrawn, transferred to standard unfolding conditions, and the amplitude of the subsequent slow unfolding reaction is determined. It is a direct measure for the amount of native molecules that had been present at the time when refolding was interrupted.

This procedure was used to follow the formation of fully folded IIHY-G3P* molecules during refolding in 0.5 M GdmCl. The unfolding assays were performed in 5.0 M GdmCl at 25 °C, conditions under which native IIHY-G3P* unfolds in a monoexponential reaction with a time constant of 25 seconds (cf. Figure 4(a)).

In 0.5 M GdmCl, fluorescence-detected refolding of IIHY-G3P* is complete within three minutes (Figure 3(a)). An unfolding assay at 5.0 M GdmCl performed after three minutes of refolding and followed by Tyr fluorescence (unfolding of domain N2) revealed, however, that virtually all molecules had unfolded already within the dead-time of manual mixing (Figure 5(a)). This suggests that the product of the fluorescence-detected refolding

is a folding intermediate and not the native protein (which unfolds with $\tau = 25$ seconds). When the duration of refolding in the first step was extended from three minutes to 70 minutes, half of the molecules unfolded slowly in the second step, and after 540 minutes all molecules unfolded slowly (Figure 5(a)). The time-course of formation of native IIHY-G3P*, as reflected in this increase of the amplitude of slow unfolding, is shown in Figure 5(b). It is governed by a time constant of 6200 seconds, and is thus 100-fold slower than folding monitored directly by fluorescence (cf. Figure 3(a)). Apparently, this very slow reaction is not accompanied by significant changes in protein fluorescence, but it leads to a dramatic decrease in the rate of unfolding.

Folding intermediates are less stable than the native protein. Therefore, to analyze the putative folding intermediate further, we produced it by a three minute refolding pulse at 0.5 M GdmCl as in Figure 5(a) and then increased the concentration of GdmCl to only 2.0 M, which is near the onset of the equilibrium unfolding transition of domain N2 (Figure 2). The refolding intermediate present after three minutes unfolded with a time constant of 43 seconds (Figure 5(c)), and this rapid unfolding was followed by a very slow refolding reaction ($\tau = 8300$ seconds), because at 2.0 M GdmCl the folded form of IIHY-G3P* is still favored (cf. Figure 2). The equilibrium value (cf. Figure 2) was not reached completely, because a small fraction of the molecules aggregated during this very slow refolding reaction.

The transient unfolding reaction at 2.0 M GdmCl (Figure 5(c)) completes the evidence that the product of fluorescence-detected refolding is an intermediate that is less stable than the native protein. It has not yet crossed the highest energy barrier of folding and thus unfolds rapidly. At 2.0 M GdmCl, the rate of unfolding of this intermediate is very similar to the apparent rate of the faster phase of refolding of domain N2 as observed in single-mixing refolding experiments (Figure 4(a)). This suggests that the faster phase of N2 folding represents the folding of this domain before domain docking.

In the unfolding of native IIHY-G3P* the domains N1 and N2 are kinetically coupled (cf. Figure 3(c)), and the unfolding of N1 is decelerated dramatically by the interaction with N2 (Figure 4(a)). If, in refolding, the final very slow process indeed reflects a docking reaction that couples the two domains, then this process should convert domain N1 from fast unfolding to slow unfolding. Therefore, we repeated the double-mixing experiments and followed this time the unfolding of domain N1 (by Trp fluorescence) in the second step after stopped-flow mixing.

Two unfolding reactions were observed, and their amplitudes depended in a reciprocal fashion on the time of refolding in step 1 (Figure 5(d)). The fast unfolding reaction showed the same time constant (30 ms) as the unfolding of isolated N1 or of N1 coupled with unfolded N2. The slow unfolding reaction showed the same time constant (25 seconds) as the unfolding of N1 in native IIHY-G3P* (cf. Figure 4(a)). With increasing duration of refolding in step 1, the amplitude of the fast phase of unfolding in step 2 decreased with $\tau = 5000$ seconds, and the amplitude of the slow phase increased with $\tau = 6250$ seconds (Figure 5(d)). These values are virtually identical with the τ value obtained from the increase in the amplitude of slow unfolding of N2 ($\tau = 6200$ seconds, Figure 5(b)). This suggests that the very slow refolding reaction reflects indeed the final docking of the two domains.

The folding kinetics of the N2 domain in the absence of domain docking

The refolding kinetics of IIHY-G3P* extend over six orders of magnitude (Figure 4(a)). Folding of the N1 domain is complete within less than 100 ms, but domain docking takes more than five hours. Thus, between these two reactions there is a wide time window for analyzing the unfolding and refolding of domain N2 in the undocked state.

The corresponding measurements again involve double-mixing. In the first set of experiments, the refolding of N2 was investigated with N1 already folded. Molecules with a folded N1 domain were produced by a 0.5 second refolding pulse at 0.5 M GdmCl, and then the protein was transferred to 0.2–1.2 M GdmCl to measure the folding of N2 as a function of the concentration of denaturant. As expected, the resulting refolding kinetics of N2 are indistinguishable from the fast phase observed for this domain in the single-mixing refolding experiments (Figure 6). The slow phase that was observed after manual single-mixing was outside the observation window of these stopped-flow double-mixing experiments.

In the second set of experiments, the unfolding of the N2 domain was measured in the molecules in which the final docking step had not yet occurred. To this end, the protein was refolded for three minutes to allow folding of N1 and N2, but not domain docking (cf. Figures 3(a) and 4(a)). Then these folded molecules were transferred to 1.2–3.0 M GdmCl to measure the unfolding kinetics of N2. N1 remains largely folded under these conditions (cf. Figure 4(a)). For the short-time refolded IIHY-G3P* the unfolding of N2 is much faster than for fully native IIHY-G3P* (Figure 6). When N2 is in the transient, undocked state, it unfolds with a time constants of 0.3 second at 3.0 M GdmCl, whereas in fully native IIHY-G3P* it unfolds with a time constant of 2080 seconds. This huge difference in the unfolding rate confirms that the product of fluorescence-detected folding is an intermediate and not the native protein. The apparent rate of unfolding of N2 in short-time refolded IIHY-G3P* measured as a function of GdmCl concentration connects around 1.6 M GdmCl with the rate of the fast refolding reaction

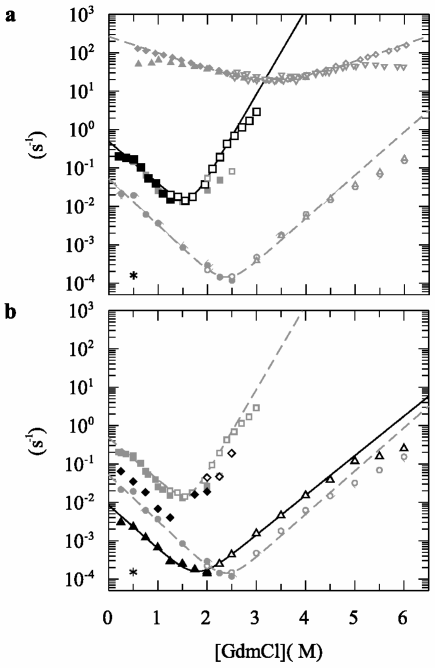


Figure 6. (a) Apparent rate constants λ of the unfolding (□) and refolding kinetics (■) of domain N2 in undocked IIHY-G3P* as a function of the concentration of GdmCl. The kinetics were measured after stopped-flow double-mixing by the change in Trp fluorescence above 320 nm after excitation at 280 nm. The denatured protein was refolded at 0.5 M GdmCl for 0.5 second (■) or for 180 seconds (□) and then further refolded or unfolded at the indicated concentrations of GdmCl. The fit of the data on the basis of a linear two-state model (the continuous line) revealed $k_{NU}(H_2O) = 2.3 \times 10^{-6}\ s^{-1}$, $k_{UN}(H_2O) = 0.48\ s^{-1}$, $m_{NU} = 5.04\ M^{-1}$, $m_{UN} = -2.52\ M^{-1}$, and $[GdmCl]_M = 1.6$ M. For comparison, the kinetic data for IIHY-G3P* folding taken from Figure 4(a), are shown in grey. (*) Rate of domain docking at 0.5 M GdmCl ($\tau = 6200$ seconds, cf. Figure 5(b)). (b) Apparent rate constants of the unfolding (△, ◇) and refolding (▲, ◆) of domain N2 of the H129Q variant as a function of the GdmCl concentration. The kinetics were measured by Tyr fluorescence at 310 nm (excitation 280 nm) after manual mixing. The fit of the slow unfolding and refolding phases on the basis of a linear two-state model (continuous line) gave $k_{NU}(H_2O) = 1.1 \times 10^{-6}\ s^{-1}$, $k_{UN}(H_2O) = 8.5 \times 10^{-3}\ s^{-1}$, $m_{NU} = 2.38\ M^{-1}$, $m_{UN} = -2.61\ M^{-1}$, and $[GdmCl]_M = 1.8$ M. The rate of domain docking at 0.5 M GdmCl (*, $\tau = 6700$ seconds) was determined by the double-mixing assay for native molecules (cf. Figure 5(b)). For comparison, the respective data for IIHY-G3P* are shown in grey. All measurements were performed with a final protein concentration of 0.5 μM in 100 mM potassium phosphate, pH 7.0, at 25 °C.

of N2, as obtained from the first set of double-mixing experiments (Figure 6(a)). Thus, it completes the chevron for domain N2 in the undocked state. Its analysis, based on a two-state model, gave a transition midpoint of $[GdmCl]_M = 1.6$ M, which is much lower than the transition midpoint of 2.5 M GdmCl as measured for N2 in native IIHY-G3P* (Figure 2(a)). This shows that N2 gains strongly in stability by the interactions that are established upon docking with N1.

The role of the hinge subdomain for domain docking

N2 consists of the globular subdomain and the hinge subdomain (Figure 1). Domain N1 and at least the globular subdomain of N2 seem to be largely folded before domain docking, which suggests that this final reaction is mediated by the hinge subdomain. In this case, the docking should be independent of mutations in the globular subdomain of N2. This part of N2 is stabilized strongly by the His at position 129, which is remote from the hinge subdomain and from the contact area with N1 (Figure 1).

The replacement of H129 by a Gln, as in wild-type G3P*, leaves the stability of N1 unchanged. N2, however, is destabilized by about 7 kJ/mol and the midpoint of its GdmCl-induced transition is decreased from 2.5 M to 2.1 M.[16] The two refolding reactions of N2 are both decelerated about fivefold and unfolding is accelerated about threefold (Figure 6(b)), as expected for a destabilizing substitution. The final docking reaction, however, as measured by the two-step unfolding assays for native molecules, is not influenced by the H129Q replacement (Figure 5(b)). Its time constant of 6700 seconds (at 0.5 M GdmCl, 25 °C) is very similar to the time constant of 6200 seconds, as obtained for IIHY-G3P*. This supports our suggestion that the docking reaction represents the association of the already folded domains N1 and N2 and that it is independent of the stabilities of the globular domains themselves.

At the minima of their chevrons (near 2.5 M and 1.8 M GdmCl, respectively) IIHY-G3P* and the H129Q variant fold with apparent rates that are virtually identical with the rate of domain docking, as measured at 0.5 M GdmCl (Figure 7). This suggests that in the transition region, under conditions where domain N2 is not stable in isolation, folding of N2 and domain docking are linked kinetically. The docking reaction is thus accompanied by changes in Tyr fluorescence, and this linkage explains why in the unfolding and subsequent refolding of short-time refolded IIHY-G3P* at 2.0 M GdmCl the final docking reaction leads to an increase in Tyr fluorescence (cf. Figure 5(c)).

Discussion

The folding mechanism of IIHY-G3P*

The refolding of IIHY-G3P* is a complex process and encompasses reactions that span the time range from 10 ms to several hours (Figure 7). N1 is the most stable domain, and it refolds very rapidly ($\tau = 9.4$ ms at 0.5 M GdmCl) in an apparently simple two-state reaction, irrespective of

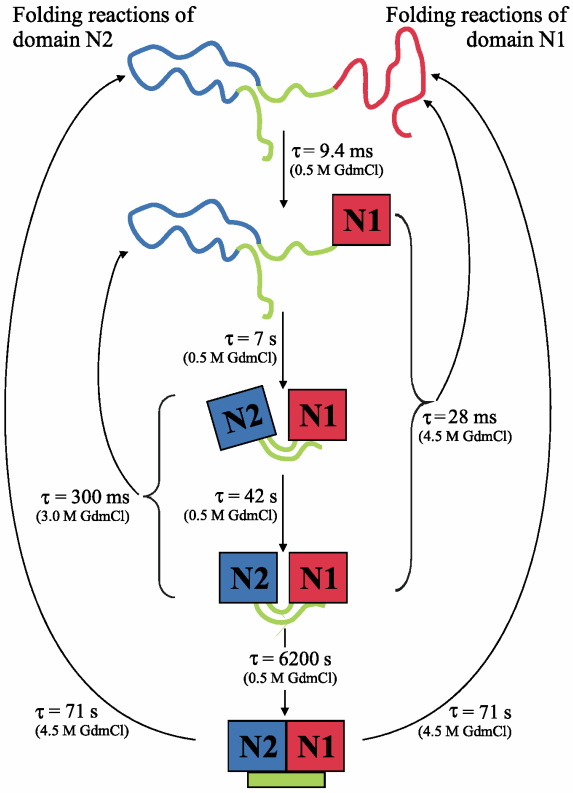


Figure 7. A possible model for the folding mechanism of IIHY-G3P*. The time constants τ are given for the refolding reactions of the domains N1 and N2 at 0.5 M GdmCl, 25 °C, and for the unfolding reactions of the two domains at 4.5 M and 3.0 M GdmCl, 25 °C. During refolding, first the domain N1 folds very rapidly (red square), followed by the folding of the globular part of N2 (blue square) and probably the formation of loose contacts between both domains. The native state with the hinge subdomain of N2 (green rectangle) in its native conformation is reached in a very slow, spectroscopically silent domain docking reaction, which couples the two domains kinetically and strongly decelerates their unfolding kinetics.

whether it is present as an isolated protein fragment or as a part of IIHY-G3P*. For unfolding, the results are different. In isolation, N1 unfolds very fast ($\tau = 28$ ms at 4.5 M GdmCl), but as a part of native IIHY-G3P* it unfolds 2500-fold more slowly ($\tau = 71$ seconds).

The N2 domain is less stable than N1 and, in fact, a major part of the Gibbs free energy of stabilization of N2 originates from its interaction with N1 in the fully folded protein. This interaction is provided primarily by the hinge subdomain of N2, which partially wraps around N1 in folded G3P*.[6,8,9] Our previous analysis of the equilibrium stability of IIHY-G3P* showed that the loss of the domain interactions and the unfolding of N2 indeed occur in a single cooperative transition.[16] N2 refolds much more slowly than N1, and its folding is a complex process, which involves two, probably sequential reactions with time constants of seven seconds and 42 seconds (at 0.5 M GdmCl, Figure 7). Both are strongly denaturant-dependent. Probably, they reflect the folding of at least the globular part of domain N2 and the formation of loose contacts with domain N1.

The folding reactions of the N1 and N2 domains are followed by a very slow, spectroscopically silent reaction with a time constant of 6200 seconds (Figure 7). This reaction would have been overlooked in conventional single-mixing experiments, because it starts from a folding intermediate that appeared native-like in its fluorescence properties. However, this intermediate unfolded more than 8000-fold faster (at 3.0 M GdmCl, 25 °C) than fully folded protein. Its final conversion to the native state could therefore be measured by a kinetic double-mixing assay for native molecules that is based on this vast difference in the unfolding rates.

We propose that the final very slow refolding reaction reflects the docking of the two domains. In fact, it links the two domains kinetically. Before docking, N1 and N2 unfold fast and independently, but after docking unfolding becomes very slow and correlated (Figure 7). This kinetic linkage explains our initial observation that, in intact G3P*, N1 unfolds much more slowly than when it is present as an isolated fragment.

Structurally, docking occurs late, in an intermediate in which Tyr and Trp fluorescence are native-like already. Domain N1 and at least the globular part of N2 have folded already at this stage. The rate of the docking reaction is apparently governed by structural changes within the hinge subdomain of N2. This assumption is supported by the finding that a destabilizing mutation in the globular subdomain of N2 decreases the rate of refolding of N2 but has no effect on the docking reaction.

G3P* as a paradigm for the folding of two-domain proteins

Single-domain proteins often fold very fast in simple two-state reactions,[19–23] but most proteins consist of more than one domain, and thus their folding mechanisms must be more complex. For two-domain proteins, a range of scenarios can be envisioned. In some proteins, the two domains represent stable structural entities that unfold and refold independently. In others, one domain might fold first and provide a scaffold for the folding of the second domain, and in still others both domains might be unstable in isolation. In this case, domain folding and docking must be tightly coupled for reaching the native state.[10,11,24–29]

An example for the first class of proteins is provided by the immunoglobulin light chain. Its two domains, C_L and V_L, are independent entities with similar stabilities, and their unfolding and refolding kinetics are the same, irrespective of whether they are linked covalently in the intact light chain or present as separate domains. Nevertheless, an additional phase could be identified in the

refolding of the intact light chain, which was ascribed to domain docking.[25] In γ_{II} crystallin of the eye lens, the two domains differ in stability, and the less stable C-terminal domain is stabilized by its interactions with the N-terminal domain. The folding kinetics of γ_{II} crystallin are well described by two interdigitated chevrons that reflect the folding of the two domains, a separate domain docking step could not be identified in the kinetic analysis.[24]

In the case of G3P*, the folding kinetics of the two domains and the kinetics of domain docking are well separated. Figure 7 shows our model for the folding mechanism of G3P*. N1 is the most stable domain and folds within a few milliseconds. It provides a framework for the folding of N2, which proceeds in the time range of a few minutes, probably in loose association with N1. Finally, the two domains become docked in a very slow reaction with a time constant of about 6000 seconds. Such a very slow domain docking reaction has not been observed before. This docking occurs after conformational folding, but nevertheless it leads to the highest energy barrier along the folding path and thus to a strong kinetic protection against unfolding for the two domains. Unfolding of N1 is decelerated up to 150,000-fold, as compared with the undocked state.

Folding and function of G3P

The very slow domain assembly and disassembly reactions are the last step in refolding and the first step in unfolding of G3P*, respectively, and they might be important for the function of this protein. G3P is located at the tip of the phage and mediates the infection of *E. coli*. The N1–N2 part protrudes from the surface of the phage virions, and the two domains must remain stably associated to protect them from inactivation in diverse environments. Such a protection is accomplished by the very high activation barrier for domain disassembly.

As soon as infection is initiated by the interaction of N2 with the tip of an F pilus, the two domains must come apart to expose the binding site for TolA on the N1 domain. The open state should persist until N1 and TolA are close enough to interact with each other. Riechmann & Holliger[3] found by ELISA measurements that isolated N1 binds to TolA about 100-fold more tightly than to domain N2, but still, within G3P binding of N1 to N2 is entropically favored because the two domains are covalently linked. The very slow domain re-assembly reaction might ensure that, once N1 and N2 became separated, domain closure is blocked long enough that the pilus can retract to bring N1 into close proximity of Tol A and allow the two to interact.

Apparently, the rates of the individual steps in the folding of G3P* seem to be determined by function rather than by the physical principles of folding. The fast folding of the N1 domain is important to rapidly present a platform for the subsequent steps in folding, and the final docking step provides a strong kinetic stabilization of G3P. Protection by a high activation barrier has its price: folding to the docked state over this barrier is very slow. Other extracellular proteins, such as secreted proteases, are protected by a high barrier of activation energy against unfolding. They avoid very slow folding by coupling barrier formation with the proteolytic removal of a pro-domain after folding. This final reaction is important for function: it generates the proteolytic activity. Such a coupling between folding and barrier formation has been well characterized for α-lytic protease.[30,31] The phage virions and thus G3P mature in the bacterial periplasm, and it remains to be seen whether they can use the host folding machinery for accelerating their own folding.

Materials and Methods

Expression and purification of G3P* and of the isolated domain N1

For the expression of IIHY-G3P* (residues 1–217 of mature G3P, plus ProSerGly$(His)_6$) and its H129Q variant the gene fragments coding for the two N-terminal domains of G3P with the stabilizing mutations T13I, T101I, Q129H, D209Y and T13I, T101I, D209Y, respectively, were PCR-amplified from the corresponding single-stranded phage DNA. The fragments were cloned into the expression plasmid pET11a (Novagen, Madison, Wisconsin, USA) *via* its *Nde* I and *Bam* HI restriction sites, the proteins were overproduced in *E. coli* BL21(DE3)-pLysS (Stratagene, La Jolla, USA) and purified as described.[16]

For the expression of the isolated domain N1 T13I (residues 1–67 of mature G3P, plus Ala$(His)_6$) the G3P-N1 fragment with signal sequence was amplified from the phage DNA, cloned into pET11a, and the protein was overproduced in *E. coli* BL21(DE3)pLysS. It was purified as described.[16]

GdmCl-induced unfolding transitions

The samples of IIHY-G3P* were prepared as described,[16] their fluorescence was measured in 10 mm cells at 310 nm after excitation at 280 nm and at 360 nm after excitation at 295 nm. The experimental data were analyzed according to two-state models by non-linear least-squares fit with proportional weighting to obtain the Gibbs free energy of denaturation ΔG_D as a function of the concentration of GdmCl.[32] The band-widths were 5 nm and 10 nm for excitation and emission, respectively. Hitachi F-4010 and F-4500 fluorescence spectrometers were used.

Kinetic single and double-mixing experiments

All GdmCl-induced unfolding and refolding experiments were performed in 100 mM potassium phosphate (pH 7.0), at 25 °C with a final protein concentration of 0.5 μM. Slow kinetics were measured after a manual ten-fold dilution of the native or denatured (in 5.0 M GdmCl) protein with GdmCl solutions of varying concentrations

to give final concentrations of GdmCl between 1.5 M and 6.0 M for unfolding and 0.25–2.5 M for refolding. The changes in Tyr fluorescence (emission at 310 nm after excitation at 280 nm) and Trp fluorescence (emission at 360 nm after excitation at 295 nm) were monitored in a 10 mm cell with a Hitachi F-4010 fluorescence spectrometer at a response of 0.5 second and band-widths of 5 nm and 10 nm for excitation and emission, respectively.

Fast refolding and unfolding reactions were measured after stopped-flow mixing in a DX.17MV spectrometer from Applied Photophysics (Leatherhead, UK). The native or unfolded (in 6.6 M GdmCl) protein was diluted 11-fold with GdmCl solutions of various concentrations. The kinetics were followed by the change in fluorescence above 320 nm after excitation at 295 nm (10 nm band-width) in an observation chamber with 2 mm path-length. A 0.5 cm cell with acetone (transparent above 320 nm) was placed between the observation chamber and the photomultiplier to absorb scattered light from the excitation beam. The kinetics were measured at least five times under identical conditions and averaged.

The fast unfolding kinetics of domain N1 linked with denatured domain N2 were measured in stopped-flow double-mixing experiments. In the first step, 27.5 μM unfolded protein in 6.0 M GdmCl was diluted 11-fold with buffer to initiate refolding at 0.55 M GdmCl, and in the second step after one second of refolding it was diluted sixfold with GdmCl solutions of varying concentrations to unfold domain N1 again. The changes in fluorescence above 320 nm after excitation at 295 nm were measured as described above.

Stopped-flow double-mixing was also used to obtain the refolding and unfolding kinetics of domain N2 in the presence of the already folded domain N1 but before the final domain docking reaction. For the measurement of refolding the unfolded protein (66 μM in 5.5 M GdmCl) was first diluted 11-fold with buffer to allow refolding of domain N1 for 0.5 second at 0.5 M GdmCl, and then diluted sixfold with GdmCl solutions of varying concentrations to obtain final concentrations of 0.2–1.2 M GdmCl and 1 μM protein. To analyze the unfolding of domain N2 before domain docking with N1 the unfolded protein (66 μM in 5.5 M GdmCl) was refolded for three minutes by an 11-fold dilution with buffer and unfolded again by a sixfold dilution with GdmCl solutions of varying concentrations to give final concentrations of GdmCl between 1.5 M and 3.0 M. The folding reactions of N2 were monitored by the changes in the fluorescence of W181 above 320 nm after excitation at 280 nm (10 nm band-width).

The dependences on GdmCl of the apparent rate constants for domains N1 and N2 were analyzed according to two-state models to obtain the microscopic rate constants and kinetic *m*-values of unfolding and refolding.[33]

Kinetics of domain docking measured by the two-step assay for native molecules

To follow the formation of native IIHY-G3P* and the H129Q variant, unfolded protein (50 μM in 5.0 M GdmCl) was first manually diluted tenfold with buffer to initiate refolding at 0.5 M GdmCl and then, after times of refolding between one minute and 540 minutes, it was unfolded again at 5.0 M GdmCl either by manual dilution tenfold with a GdmCl solution of 5.5 M or by 11-fold stopped-flow mixing with 5.45 M GdmCl. The manual mixing in the Hitachi F-4010 fluorescence spectrometer was used to follow the slow unfolding reaction ($\tau = 25$ seconds) of domain N2 by the change in Tyr fluorescence at 310 nm (10 nm band-width) after excitation at 280 nm (5 nm band-width). In the stopped-flow spectrometer, the fast ($\tau = 30$ ms) and the slow ($\tau = 25$ seconds) unfolding phases of domain N1 were detected by the fluorescence above 320 nm after excitation at 295 nm (10 nm band-width).

The unfolding of the short-time refolded intermediate after three minutes at 0.5 M GdmCl was initiated by a manual dilution tenfold to 2.0 M GdmCl and monitored by the change in the fluorescence of W181 at 340 nm after excitation at 280 nm.

Acknowledgements

We thank the members of our laboratory for many discussions. This work was supported by grants from the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie.

References

1. Levengood, S. K., Beyer, W. F., Jr & Webster, R. E. (1991). TolA: a membrane protein involved in colicin uptake contains an extended helical region. *Proc. Natl Acad. Sci. USA*, **88**, 5939–5943.
2. Click, E. M. & Webster, R. E. (1997). Filamentous phage infection: required interactions with the TolA protein. *J. Bacteriol.* **179**, 6464–6471.
3. Riechmann, L. & Holliger, P. (1997). The C-terminal domain of TolA is the coreceptor for filamentous phage infection of *E. coli*. *Cell*, **90**, 351–360.
4. Stengele, I., Bross, P., Garces, X., Giray, J. & Rasched, I. (1990). Dissection of functional domains in phage fd adsorption protein. Discrimination between attachment and penetration sites. *J. Mol. Biol.* **212**, 143–149.
5. Marvin, D. A. (1998). Filamentous phage structure, infection and assembly. *Curr. Opin. Struct. Biol.* **8**, 150–158.
6. Lubkowski, J., Hennecke, F., Plückthun, A. & Wlodawer, A. (1999). Filamentous phage infection: crystal structure of g3p in complex with its coreceptor, the C-terminal domain of TolA. *Structure*, **7**, 711–722.
7. Boeke, J. D. & Model, P. (1982). A prokaryotic membrane anchor sequence: carboxyl terminus of bacteriophage f1 gene III protein retains it in the membrane. *Proc. Natl Acad. Sci. USA*, **79**, 5200–5204.
8. Lubkowski, J., Hennecke, F., Plückthun, A. & Wlodawer, A. (1998). The structural basis of phage display elucidated by the crystal structure of the N-terminal domains of G3p. *Nature Struct. Biol.* **5**, 140–147.
9. Holliger, P., Riechmann, L. & Williams, R. L. (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**, 649–657.
10. Garel, J.-R. (1992). Folding of large proteins:multi-domain and multisubunit proteins. In *Protein Folding* (Creighton, T. E., ed.), pp. 405–454, Freeman, New York.
11. Jaenicke, R. (1999). Stability and folding of domain proteins. *Prog. Biophys. Mol. Biol.* **71**, 155–241.

12. Netzer, W. J. & Hartl, F. U. (1997). Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature*, **388**, 343–349.
13. Deng, L. W. & Perham, R. N. (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**, 603–614.
14. Jacobson, A. (1972). Role of F pili in the penetration of bacteriophage fl. *J. Virol.* **10**, 835–843.
15. Frost, L. S. (1993). Conjugative pili and pilus-specific phages. In *Bacterial Conjugation* (Clewell, D. B., ed.), pp. 189–221, Plenum Press, New York.
16. Martin, A. & Schmid, F. X. (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**, 863–875.
17. Schmid, F. X. (1983). Mechanism of folding of ribonuclease A. Slow refolding is a sequential reaction *via* structural intermediates. *Biochemistry*, **22**, 4690–4696.
18. Schmid, F. X. (1986). Fast-folding and slow-folding forms of unfolded proteins. In *Enzyme Structure Part L* (Hirs, C. H. W. & Timasheff, S. N., eds), 1st edit., vol. 131, pp. 71–82, Academic Press, New York.
19. Jackson, S. E. & Fersht, A. R. (1991). Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry*, **30**, 10428–10435.
20. Schindler, T., Herrler, M., Marahiel, M. A. & Schmid, F. X. (1995). Extremely rapid folding in the absence of intermediates: the cold-shock protein from *Bacillus subtilis*. *Nature Struct. Biol.* **2**, 663–673.
21. Huang, G. S. & Oas, T. G. (1995). Submillisecond folding of monomeric lambda repressor. *Proc. Natl Acad. Sci. USA*, **92**, 6878–6882.
22. Jackson, S. E. (1998). How do small single-domain proteins fold? *Fold. Des.* **3**, R81–R91.
23. Alm, E. & Baker, D. (1999). Matching theory and experiment in protein folding. *Curr. Opin. Struct. Biol.* **9**, 189–196.
24. Rudolph, R., Siebendritt, R., Neslauer, G., Sharma, A. K. & Jaenicke, R. (1990). Folding of an all-beta protein: independent domain folding in gamma-II-crystallin from calf eye lens. *Proc. Natl Acad. Sci. USA*, **87**, 4625–4629.
25. Tsunenaga, M., Goto, Y., Kawata, Y. & Hamaguchi, K. (1987). Unfolding and refolding of a type kappa immunoglobulin light chain and its variable and constant fragments. *Biochemistry*, **26**, 6044–6051.
26. Teschner, W., Rudolph, R. & Garel, J.-R. (1987). Intermediates on the folding pathway of octopine dehydrogenase from *Pecten jacobaeus*. *Biochemistry*, **26**, 2791–2796.
27. Dautry-Varsat, A. & Garel, J. R. (1981). Independent folding regions in aspartokinase-homoserine dehydrogenase. *Biochemistry*, **20**, 1396–1401.
28. Murry-Brelier, A. & Goldberg, M. E. (1989). Alternate succession of steps can lead to the folding of a multidomain oligomeric protein. *Proteins: Struct. Funct. Genet.* **6**, 395–404.
29. Zitzewitz, J. A. & Matthews, C. R. (1999). Molecular dissection of the folding mechanism of the alpha subunit of tryptophan synthase: an amino-terminal autonomous folding unit controls several rate-limiting steps in the folding of a single domain protein. *Biochemistry*, **38**, 10205–10214.
30. Cunningham, E. L., Mau, T., Truhlar, S. M. E. & Agard, D. A. (2002). The pro region N-terminal domain provides specific interactions required for catalysis of alpha-lytic protease folding. *Biochemistry*, **41**, 8860–8867.
31. Cunningham, E. L., Jaswal, S. S., Sohl, J. L. & Agard, D. A. (1999). Kinetic stability as a mechanism for protease longevity. *Proc. Natl Acad. Sci. USA*, **96**, 11008–11014.
32. Santoro, M. M. & Bolen, D. W. (1988). Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl α-chymotrypsin using different denaturants. *Biochemistry* **27**, 8063–8068.
33. Kiefhaber, T. (1995). Protein folding kinetics. *Methods Mol. Biol.* **40**, 313–341.
34. Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55. see also pp. 29–32.

Edited by C. R. Matthews

(Received 24 January 2003; received in revised form 24 March 2003; accepted 24 March 2003)

7.6 Teilarbeit F

F

Andreas Martin and Franz X. Schmid.

A Proline Switch Controls Folding and Domain Interactions in the Gene-3-protein of the Filamentous Phage fd.

The Journal of Molecular Biology **331**, 1131-1140 (2003)

doi:10.1016/S0022-2836(03)00864-7 *J. Mol. Biol.* (2003) **331**, 1131–1140

JMB Available online at www.sciencedirect.com SCIENCE DIRECT®

A Proline Switch Controls Folding and Domain Interactions in the Gene-3-protein of the Filamentous Phage fd

Andreas Martin and Franz X. Schmid*

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

The amino-terminal domains N1 and N2 of the gene-3-protein of phage fd form a bilobal structural and functional entity that protrudes from the phage tip. Domain N2 initiates the infection of *Escherichia coli* by binding to the F pilus. This binding results in the dissociation of the two domains and allows N1 to interact with the TolA receptor at the cell surface. The refolding of the N1–N2 fragment begins with the folding of domain N1, which takes a few milliseconds, followed by the folding of domain N2, which is complete within five minutes. The subsequent domain assembly is unusually slow and shows a time-constant of 6200 s at 25 °C. We found that the rate of this reaction is controlled by the *trans* to *cis* isomerization of the Gln212-Pro213 bond in the hinge subdomain of N2, a region that provides many interactions between N1 and N2 in the gene-3-protein. The substitution of Pro213 by Gly accelerated domain association 30-fold and revealed that the folding of the two individual domains and their assembly are indeed sequential steps in the refolding of the gene-3-protein. In the course of infection, the domains must separate to expose the binding site for TolA on domain N1. The kinetic block of domain reassembly caused by Pro213 isomerization could ensure that after the initial binding of N2 to the F pilus the open state persists until N1 and TolA are close enough for their mutual interaction. Pro213 isomerization might thus serve as a slow conformational switch in the function of the gene-3-protein.

Keywords: domain protein; protein folding; prolyl isomerization; pilus; TolA

*Corresponding author

Introduction

Domains are modules of protein structure and function, and they represent the basic folding units.[1–3] The folding of multi-domain proteins includes both the folding and the subsequent assembly of the domains, and thus it is more complex than the folding of single-domain proteins. In isolation, individual domains are often unstable or cannot fold to completion, and for such cases in the intact protein the folding of the domains is linked with their assembly.[2,4–6]

The gene-3-protein (G3P) of filamentous phages is essential for the infection of *Escherichia coli*. G3P consists of three domains. The carboxyterminal CT domain is partly embedded in the phage coat,[7,8] whereas the aminoterminal domains N1 and N2 protrude from the phage surface. They form a bilobal, horseshoe-like structure (Figure 1), in which the larger N2 domain (131 residues) wraps around the smaller N1 domain (68 residues).[9–11] The crystal structure of the N1–N2 part of G3P (denoted as G3P*) reveals both a clear separation

Abbreviations used: G3P, gene-3-protein of phage fd; N1, N2 and CT, the two N-terminal and the C-terminal domains of G3P, respectively; G3P*, a fragment of G3P that consists of domains N1 and N2; IIHY-G3P*, G3P* containing the four stabilizing mutations T13I, T101I, Q129H, and D209Y; GdmCl, guanidinium chloride; $[GdmCl]_M$, midpoint of a GdmCl-induced unfolding transition; λ, apparent rate constant and τ (λ^{-1}), apparent time-constant of a folding reaction; k_{NU}, microscopic rate constant and m_{NU} (= δ ln k_{NU}/δ[GdmCl]), kinetic *m*-value of unfolding; k_{UN}, m_{UN}, microscopic rate constant and kinetic *m*-value of refolding; k_{ct}, and k_{tc}, microscopic rate constants of *cis* to *trans* and *trans* to *cis* isomerizations; K_{tc}, equilibrium constant for a *trans* to *cis* isomerization.

E-mail address of the corresponding author: fx.schmid@uni-bayreuth.de

0022-2836/$ - see front matter

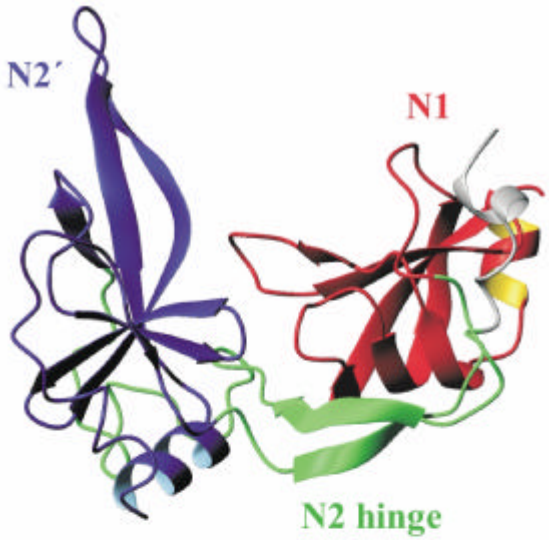


Figure 1. Tertiary structure of G3P* (coordinates from Holliger *et al.*[11]). Domain N1 is shown in red, the globular subdomain of N2 (N2′) in blue, and the hinge subdomain of N2 in green (as defined by Holliger *et al.*[11]). The Figure was prepared by using MolMol.[52]

into two structural domains and an extensive domain interface. N2 can be further decomposed into the globular subdomain, which resembles N1 in size and chain topology, and the hinge subdomain, which provides most of the contacts with N1 (Figure 1).[11]

We use G3P* as a model for investigating the folding mechanism of two-domain proteins and the interrelation between domain folding and assembly. In fact, we employ a variant with four stabilizing mutations (T13I, T101I, Q129H and D209Y) that originated from an *in vitro* selection for G3P mutants with increased thermodynamic stability.[12] This variant is denoted as IIHY-G3P*.

The equilibrium unfolding of IIHY-G3P* consists of two transitions. In the first transition, domain disassembly and unfolding of the less stable domain N2 occur in a cooperative process. In the second transition, domain N1 unfolds in a separate reaction.[12]

The folding kinetics of IIHY-G3P* are complex and extend from milliseconds to hours.[13] In the first reaction, N1 folds very fast ($\tau = 9.4$ ms at 0.5 M guanidinium chloride (GdmCl), 25 °C), both as a part of the native protein and as an isolated fragment. The subsequent folding of N2 is about 1000-fold slower and proceeds in two, presumably sequential, reactions (with time-constants of 7 s and 42 s, respectively). The final docking of the two prefolded domains is another two orders of magnitude slower and shows a time-constant of 6200 s at 25 °C. This reaction is spectroscopically silent, but it could be detected by kinetic unfolding assays for native molecules.[13] In the undocked state, N1 and N2 unfold fast and independently of each other; after docking, they unfold very slowly in a correlated fashion.[13]

Here, we investigated the molecular nature of the domain docking reaction in IIHY-G3P*. We find that its rate is governed by the unusually slow *trans* to *cis* isomerization of the Gln212-Pro213 bond in the hinge subdomain of N2. This prolyl bond is *cis* in the native protein, and it acts as a molecular switch that controls the rate of domain docking in the folding of G3P*.

Results

Slow isomerization is independent of the denaturant concentration

In our preceding work,[13] we developed a two-step assay for measuring the formation of native G3P* molecules during refolding. This assay exploits the fact that only fully folded molecules unfold slowly, whereas all partially folded intermediates unfold rapidly. Unfolded protein is first mixed with buffer to initiate refolding. Then, after variable times of refolding, samples are withdrawn and transferred to standard unfolding conditions to determine the amplitude of the subsequent slow unfolding reaction. This amplitude is a measure for the amount of native molecules that were present at the time of sample transfer. By using this assay, we found that during the refolding of G3P* in 0.5 M GdmCl the native state is reached in an extremely slow reaction ($\tau = 6200$ s), which is about 150-fold slower than refolding as followed by the changes in protein fluorescence. This final folding reaction couples domains N1 and N2 kinetically in their unfolding and reflects the docking of the two prefolded domains.[13]

Conformational folding reactions usually become slower when the conditions are less favorable for the native state, typically when the residual concentration of the denaturant is increased.[14] The very slow refolding reaction of G3P*, however, is independent of the final concentration of GdmCl. Its kinetics, as followed by the assay for native molecules at 0.5, 1.0, 1.5 and 2.0 M GdmCl, show a GdmCl-independent time-constant of about 6200 s (Figure 2). In contrast, the fluorescence-detected refolding, which reflects the folding of N2 before domain docking, is decelerated strongly when the concentration of GdmCl is increased. As a consequence, the rates of folding as measured by Tyr fluorescence and by the assay for native molecules coincide near 2.5 M GdmCl (Figure 3), which represents the midpoint of the equilibrium unfolding transition of domain N2 in the native IIHY-G3P*.[13] This indicates that near the transition midpoint the folding of N2 is driven by the assembly reaction with domain N1. The shift in the nature of the rate-limiting process from an intra-domain to an inter-domain reaction leads to a strong, apparently linear decrease of $\ln\lambda$ with increasing [GdmCl] (Figure 3). It is clear that the

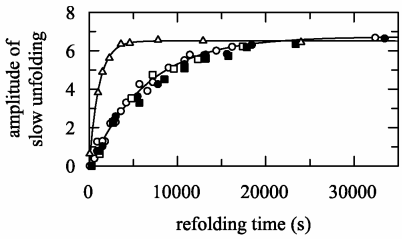


Figure 2. Time-course of formation of native IIHY-G3P* as measured by unfolding assays (see Materials and Methods). The amplitudes of the slow unfolding reaction ($\tau = 25$ s) of domain N2 in IIHY-G3P* at 5.0 M GdmCl are shown as a function of the time of refolding at 0.5 M (○), 1.0 M (●), 1.5 M (□), and 2.0 M GdmCl (■), and at 0.5 M GdmCl in the presence of 1.0 μM cyclophilin 18 (△). The continuous lines represent mono-exponential curves, fit to the data for the uncatalyzed (○, $\tau = 6200$ s) and the catalyzed (△, $\tau = 1100$ s) refolding at 0.5 M GdmCl. The unfolding kinetics were followed by Tyr fluorescence at a final protein concentration of 0.5 μM in 100 mM potassium phosphate (pH 7.0), 25 °C.

observed apparent *m*-value for refolding ($m_{UN} = \delta\, ln\, k_{UN}/\delta$[GdmCl]) cannot be interpreted as in the case of the two-state folding reactions of small, single-domain proteins.[14–16]

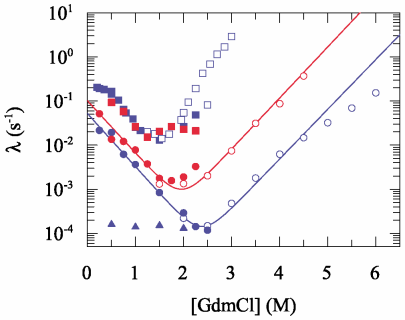


Figure 3. Refolding kinetics (●, ■) and unfolding kinetics (○, □) of domain N2 in IIHY-G3P* (blue) and in the P213G mutant (red), as measured by the changes in Tyr fluorescence. The apparent rate constants λ are shown as a function of the concentration of GdmCl. The data for the conformational folding reactions of IIHY-G3P* were taken from Martin and Schmid,[13] the rates of domain docking in IIHY-G3P* (▲) were determined by the assay for native molecules (see Figure 2). Fits of the slowest unfolding and refolding reactions of domain N2 in IIHY-G3P* and its P213G variant on the basis of linear two-state models are shown by continuous lines. The obtained parameters for IIHY-G3P* are $k_{NU}(H_2O) = 1.5 \times 10^{-7}\ s^{-1}$, $k_{UN}(H_2O) = 0.052\ s^{-1}$, $m_{NU} = 2.59\ M^{-1}$, $m_{UN} = -2.77\ M^{-1}$, and $[GdmCl]_M = 2.4$ M; for the P213G variant $k_{NU}(H_2O) = 2.8 \times 10^{-6}\ s^{-1}$, $k_{UN}(H_2O) = 0.101\ s^{-1}$, $m_{NU} = 2.66\ M^{-1}$, $m_{UN} = -2.72\ M^{-1}$, and $[GdmCl]_M = 1.95$ M were obtained. All measurements were performed at a final protein concentration of 0.5 μM in 100 mM potassium phosphate (pH 7.0), at 25 °C.

The final refolding reaction is limited in rate by prolyl isomerization

Denaturant-independent rates are a hallmark of folding reactions that are limited in rate by prolyl isomerizations. Prolyl isomerizations are, however, typically 100-fold faster than the slowest process observed in G3P* folding. Proline-limited steps in a folding reaction can be identified by using a prolyl isomerase,[17] which catalyzes the *cis* to *trans* isomerization, provided that the local sequence around the proline is recognized and the respective proline residue remains accessible during folding.

The formation of native molecules of IIHY-G3P* is in fact catalyzed by a prolyl isomerase (Figure 2), and its time-constant is reduced from 6200 s to 1100 s in the presence of 1 μM cyclophilin 18.[18,19] We conclude from this result that the final domain docking reaction in the folding of G3P is limited in rate by a prolyl isomerization. The respective prolyl bond is probably located at or near the protein surface or in an incompletely folded region, otherwise it would not be accessible for catalysis.

Prolyl isomerization occurs also in the unfolded protein

Prolyl isomerizations are slow steps in refolding and in unfolding. Isomerizations in unfolded proteins are usually silent, and therefore they cannot be measured directly, e.g. by a spectroscopic probe. Rather, a double-jump technique is used that exploits the fact that unfolded molecules with native-like prolyl isomers refold rapidly, whereas molecules with non-native isomers refold slowly.[20,21]

For G3P*, the proline-limited step is spectroscopically silent in both unfolding and refolding. In refolding, it could be followed by double-mixing experiments (refolding, followed by unfolding), because the molecules with a non-native prolyl isomer unfold fast, whereas those with a native-like prolyl isomer unfold slowly (cf. Figure 2). Therefore, to detect and measure the corresponding prolyl isomerization in unfolded G3P*, triple-mixing experiments were necessary.

In the first step of these experiments, unfolding of native G3P* was initiated by diluting it to 5 M GdmCl. Then, after various times of unfolding, samples were withdrawn and refolded for 180 s at 0.5 M GdmCl. This refolding step is long enough for the fluorescence-detected refolding reactions to go to completion (cf. Figure 3), but it is much too short for the proline-limited domain docking, which shows a time-constant of 6200 s under these conditions (cf. Figure 2). In other words, only the unfolded molecules with still native-like prolyl isomers will fold to the fully native state during the 180 s refolding step. These native G3P* molecules in the mixtures are easily identified, because they unfold slowly (cf. Figure 2). Therefore, in the third step of the assay, the samples were returned to

unfolding conditions (5.0 M GdmCl), and the amplitude of the slow unfolding reaction was determined as a function of the duration of unfolding in the first step of the assay. The Tyr fluorescence was used as a probe, which monitors the unfolding of domain N2 in G3P*.

The results are shown in Figure 4. After 60 s of unfolding in the first step, the amplitude of slow unfolding in the third step of the experiment was almost as high as the amplitude of unfolding of native G3P*. This indicates that after such a short unfolding pulse most molecules still had a native-like prolyl isomer. They were able to refold completely within the 180 s refolding pulse in the second step of the assay, and thus, in the third step, they unfolded as slowly as native G3P*. When the duration of unfolding in the first step was increased, more and more unfolded molecules with a non-native isomer were formed. These molecules cannot refold to the native state within the 180 s refolding pulse, and in the third step they unfold within the dead-time of mixing. The time-course of the decrease of the slow-unfolding molecules in Figure 4 thus reflects the kinetics of the respective *cis* to *trans* prolyl isomerization in unfolded IIHY-G3P*. It occurs with a time-constant of 350 s, well after the conformational unfolding, which shows a time-constant of 30 s under the conditions of the first step (5.0 M GdmCl).

In an analogous set of triple-jump experiments, the unfolding in the third step was followed by Trp fluorescence, which monitors the amplitude of unfolding of the N1 domain. This probe gave the same time-course as unfolding of N2 for the prolyl isomerization in the first step (Figure 4). Moreover, identical rates were observed in the third (unfolding) step, indicating that the two domains are kinetically coupled in their unfolding. This confirms that in the second step the molecules with the native prolyl isomer had indeed reached the fully folded state, in which the two domains are docked.

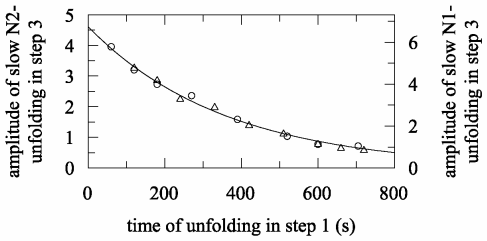


Figure 4. Kinetics of the slow isomerization in unfolded IIHY-G3P*, as measured by triple-jump assays (see Materials and Methods). The amplitude of slow unfolding ($\tau = 25$ s) in step 3 is shown as a function of the time of unfolding in step 1. Unfolding in step 3 was followed by the change in Tyr fluorescence for domain N2 (○, left ordinate) and by the change in Trp fluorescence for domain N1 (△, right ordinate). The continuous line represents the monoexponential curve, fit to the data for domains N2 and N1, the derived apparent time-constant for the isomerization is $\tau = 350$ s. The final protein concentration was 0.5 μM in 100 mM potassium phosphate (pH 7.0), at 25 °C.

A protein unfolding reaction that is followed by a prolyl isomerization in the unfolded state is described by the mechanism in equation (1), in which N stands for the native protein, and U_c and U_t denote the unfolded forms with the proline residue in the native *cis* and the non-native *trans* conformations, respectively. We assume that the critical proline residue of IIHY-G3P* is *cis* in the native protein, because most molecules must undergo isomerization during refolding to reach the native state (cf. Figure 2):

$$N \rightarrow U_c \underset{k_{tc}}{\overset{k_{ct}}{\rightleftharpoons}} U_t \qquad (1)$$

The three-step experiments in Figure 4 measure the apparent rate constant k_{app} of the $U_c \rightleftharpoons U_t$ equilibration (0.0029 s^{-1}). It is equal to the sum of the two microscopic rate constants of isomerization, k_{ct} and k_{tc} (equation (2)). The equilibrium constant for isomerization, $K_{tc} = [U_c]/[U_t]$ is given by the ratio k_{tc}/k_{ct} (equation (3)):

$$k_{app} = k_{ct} + k_{tc} \qquad (2)$$

$$K_{tc} = k_{tc}/k_{ct} \qquad (3)$$

The refolding of the U^t molecules is limited by the *trans* to *cis* isomerization (equation (1)), and the rate of this reaction (0.00016 s^{-1}, Figure 2) is equal to the microscopic rate constant k_{tc}, provided that the rate of prolyl isomerization is not affected by conformational folding. Such a coupling with conformational folding is unlikely, because the rate of the slowest step in G3P* folding is independent of the denaturant concentration (cf. Figure 3). From $k_{app} = 0.0029$ s^{-1} and $k_{tc} = 0.00016$ s^{-1}, a value for k_{ct} of 0.00274 s^{-1} is derived (equation (2)). These values for k_{ct} and k_{tc} are very low. For comparison, the k_{ct} values for the isomerization of Xaa-Pro bonds in short peptides range between 0.005 s^{-1} and 0.050 s^{-1} (at 25 °C).[22] K_{tc} for the isomerization in IIHY-G3P* is 0.06 (equation (3)), which indicates that at equilibrium in 94% of the unfolded molecules the critical proline residue is in the non-native *trans* conformation. This explains, why the very slow refolding reaction was observed for virtually all molecules (cf. Figure 2).

The very slow unfolding and refolding reactions originate from Pro213

In its native state, G3P* contains two *cis* prolyl peptide bonds; both are in N2. *Cis* Pro161 is in a solvent-exposed turn in the globular subdomain, and *cis* Pro213 is in the hinge subdomain of N2 near the interface between the two domains (Figure 7(a)).

To examine the role of Pro213 for the folding of IIHY-G3P* we exchanged it for a glycine residue. The data for the unfolding and refolding of the respective mutant are included in Figure 3. The

P213G mutation leaves the kinetics of the two refolding reactions of the N2 domain and their dependences on GdmCl concentration almost unchanged. Unfolding of IIHY-G3P* is, however, accelerated 20-fold. Together, this suggests that the P213G mutation affects a reaction that occurs very late in refolding (after the fluorescence-detected steps) but early in unfolding.

The native state of G3P* is reached much faster after the P213G mutation. Domain docking in the last step of refolding, as measured by the unfolding assay for native molecules, is accelerated about 30-fold. The respective time-constants are 6200 s for IIHY-G3P* (Figure 2) and 240 s for the P213G variant (Figure 5(a)). The time-course of native state formation shows a pronounced lag, which indicates that the native molecules are formed in a sequential reaction. A tentative kinetic analysis of the data at 0.5 M GdmCl (Figure 5(a)), on the basis of a sequential three-state mechanism, gave rate constants of 0.011 s^{-1} (equal to $\tau = 77$ s) for the first step and 0.0041 s^{-1} (equal to $\tau = 240$ s) for the second step. The rate constant obtained for the first step coincides well with the rate constant of the slowest fluorescence detected refolding reaction of domain N2 ($\lambda = 0.013\ s^{-1}$ at 0.5 M GdmCl, Figure 3).

At 1.0 M GdmCl, the lag in the formation of native molecules is even more pronounced (Figure 5(b)). In this case the analysis revealed rate constants of 0.007 s^{-1} for the first step and of 0.0041 s^{-1} for the second step (as at 0.5 M GdmCl). Again, the value of 0.007 s^{-1} agrees well with the respective rate of the slowest phase in the fluorescence-detected folding of domain N2 ($\lambda = 0.007\ s^{-1}$ at 1.0 M GdmCl, Figure 3).

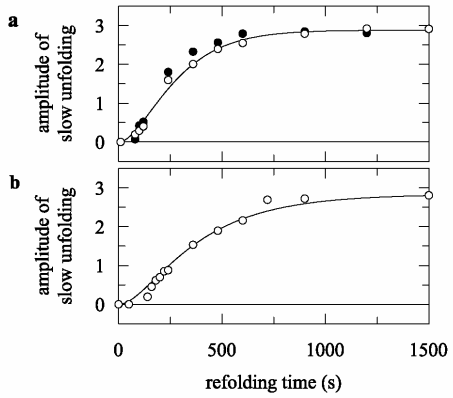


Figure 5. Time-course of formation of native IIHY-G3P* P213G at 25 °C measured by unfolding assays, as described in Materials and Methods. The amplitudes of the slow unfolding reaction ($\tau = 32$ s) of domain N2 at 3.5 M GdmCl are shown as a function of the time of refolding (a) at 0.5 M GdmCl in the absence (○) and in the presence (●) of 1.0 μM cyclophilin 18, and (b) at 1.0 M GdmCl. The continuous lines represent curves fit to the data according to a sequential two-step refolding mechanism, assuming that only the slower reaction contributes to the increase in the amplitude of slow unfolding. The obtained time-constants for the two steps are $\tau_1 = 77$ s and $\tau_2 = 240$ s at 0.5 M GdmCl (a), and $\tau_1 = 143$ s and $\tau_2 = 240$ s at 1.0 M GdmCl (b). The unfolding kinetics were measured by Tyr fluorescence at a final protein concentration of 0.5 μM in 100 mM potassium phosphate (pH 7.0), 25 °C.

In Figure 4 we had used a triple-jump procedure to detect the slow prolyl isomerization in unfolded IIHY-G3P* and to measure its kinetics. By using the same procedure we found that an isomerization occurs in the unfolded form of the P213G variant as well. Its rate is, however, accelerated 70-fold and it is complete within 20 s (Figure 6).

The strong rate enhancements in unfolding and refolding caused by the P213G substitution suggest that the very slow domain docking reaction in the parent protein IIHY-G3P* originates indeed from the *trans* to *cis* isomerization at Pro213 in the hinge subdomain. The 30-fold acceleration of this reaction in the P213G mutant led to a lag in the formation of the native protein, because it approached the conformational folding of the N2 domain in rate. The sigmoidal kinetics demonstrate that domain folding and domain docking are indeed consecutive steps on a sequential pathway. The sequential nature of folding and domain docking could not be demonstrated for IIHY-G3P* itself, because here the two steps differ too strongly in rate.

The accelerated isomerization in the P213G variant could reflect the *cis* to *trans* isomerization of the Q212-G213 bond or, alternatively, of the second *cis* prolyl bond in G3P*, Asp160-Pro161, which is located in the globular subdomain of N2, remote from the domain interface. The isomerization of G3P* P213G is not accelerated in the presence of cyclophilin 18 (Figure 5(a)), which argues against the isomerization at Pro161 as the

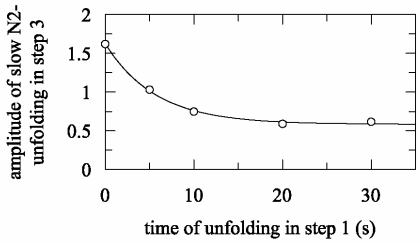


Figure 6. Kinetics of the slow isomerization in unfolded IIHY-G3P* P213G, as measured by triple-jump assays (see Materials and Methods). The amplitude of slow unfolding in step 3 ($\tau = 32$ s, at 3.5 M GdmCl) is shown as a function of the time of unfolding in step 1 (at 5.0 M GdmCl). Unfolding in step 3 was followed by Tyr fluorescence, the final protein concentration was 0.38 μM in 100 mM potassium phosphate (pH 7.0), 25 °C. The continuous line shows the single-exponential curve fit, the derived time-constant for the isomerization is $\tau = 5.6$ s. The amplitude of unfolding in step 3 does not decrease to zero after extended times of unfolding in step 1, because a fraction of the unfolded molecules with non-native isomer reaches the native state within the 180 s refolding pulse in step 2 (see Figure 5(a)).

rate-limiting reaction. Pro161 is strongly exposed, even in fully folded G3P*, and thus it should be accessible for catalysis during refolding.

We therefore suggest that the observed isomerization in the P213G variant occurs at the Gln212-Gly213 bond. The data in Figures 5 and 6 were used with equations (2) and (3) to derive the microscopic rate constants ($k_{ct} = 0.173\ s^{-1}$, $k_{tc} = 0.0042\ s^{-1}$) and the equilibrium constant ($K_{tc} = 0.024$) for this isomerization. The value for k_{ct} is outside the range of values observed for proline-containing peptides. For non-prolyl (secondary) peptide bonds, only limited reference data for the isomerization rates and the equilibrium constants are available. The published k_{ct} values for short peptides range between $0.24\ s^{-1}$ and $2.38\ s^{-1}$ and the K_{tc} values between 0.0011 and 0.0048.[23,24] For non-prolyl peptide bonds in proteins, the observed isomerization rates lie between $0.4\ s^{-1}$ and $2.3\ s^{-1}$ for k_{ct} and between $1 \times 10^{-3}\ s^{-1}$ and $4 \times 10^{-3}\ s^{-1}$ for k_{tc}.[25–27] It is likely that the domain docking reaction in the P213G variant is in fact caused by the isomerization of the Gln212-Gly213 bond, but it remains unclear why the *cis*/*trans* interconversion at this bond is rather slow and why the fraction of the *cis* isomer present at equilibrium in the unfolded protein is rather high.

Discussion

Proline-limited slow domain docking of G3P

The docking between the N1 and N2 domains is the final reaction in the folding of the gene-3-protein of phage fd. Its rate is controlled by the unusually slow *trans* to *cis* isomerization of the Gln212-Pro213 prolyl peptide bond in the hinge subdomain of N2. This prolyl bond is *cis* in native G3P*[9,11] but *trans* in 94% of the unfolded molecules. The *cis* to *trans* equilibration of the Gln212-Pro213 bond in the unfolded protein ($\tau = 360$ s) and the *trans* to *cis* isomerization in refolding ($\tau = 6200$ s) are both one to two orders of magnitude slower than typical prolyl isomerizations. In a tetrapeptide, the *cis* to *trans* isomerization of an Ala-Pro bond occurs with a time-constant of about 20 s,[28] and the *cis* to *trans* equilibrations at various Xaa-Pro bonds in unfolded proteins show time-constants in the range of 20–50 s (at 25 °C).[29–34]

Fischer and co-workers investigated prolyl *cis* to *trans* isomerizations in short peptides at 4 °C and found that the Gln-Pro bond isomerizes about threefold more slowly than Ala-Pro, and that the isomerization of Ala-Pro is decelerated threefold when it is followed by an additional proline residue (as in the peptide APPAK).[22] Pro213 in the hinge subdomain of G3P is preceded by a Gln and followed by a Pro, and thus its very low rate of isomerization might in fact be a consequence of this peculiar local sequence. In protein folding, prolyl isomerizations can be influenced by a coupling with conformational folding. In RNase A, prolyl isomerization in refolding is accelerated by such a coupling,[29,35] whereas it is decelerated in RNase T_1.[36] In the latter case, proline-limited refolding becomes faster with increasing concentration of denaturant, because the denaturant destabilizes a partially folded structure. The proline-limited domain docking reaction of G3P* is independent of the concentration of GdmCl, and therefore it is probably not coupled with conformational folding.

Domain docking is a late step in refolding

The P213G mutation strongly accelerated the final domain docking between N1 and N2 in refolding, which allowed us to identify the isomerization of the Gln212-Pro213 bond as the rate-determining event of docking. This mutation also decreased the stability of the folded protein, as reflected in a 20-fold increase in the rate of unfolding. The rates of the two refolding reactions of N2 that precede domain docking were not affected, which suggests that the region around position 213 is structured very late, after the folding of N2. This confirms our conclusion that the proline-controlled domain docking is a late step that occurs after the two domains have essentially completed their conformational folding.

Proline-dependent conformational switches

The first *cis* prolyl peptide bond in a folded protein was detected in bovine pancreatic RNase S[37] in 1970, and a few years later Brandts *et al*. postulated that prolyl *cis* to *trans* isomerization causes slow steps in the unfolding and refolding of this protein.[20] In fact, many protein folding reactions are limited in rate by prolyl isomerizations,[38] and several enzymes, prolyl isomerases, have been discovered that catalyze these reactions.[18,39,40]

Isomerizations at prolyl peptide bonds occur in the course of protein folding and in folded proteins,[38] and it was suggested that such isomerizations might be used as slow molecular switches for turning biological activities on and off.[41] The rate of switching could easily be modulated by prolyl isomerases. However, it is difficult to identify prolyl isomerizations in folded proteins and, therefore, the evidence for proline switches remained circumstantial. Interesting data pointing to a proline-dependent conformational switch were obtained for the tyrosine kinase Itk.[42,43] Unlike other proteins with *cis* and *trans* prolyl conformers in the native state,[44–48] the SH2 domain of Itk shows a pronounced change in structure upon *cis*/*trans* isomerization. This conformational switch seems to control substrate recognition and can be regulated by the prolyl isomerase cyclophilin 18. Another prolyl isomerase, PIN1, catalyzes the isomerization of phosphoserine–proline bonds, and there is compelling evidence that PIN1 integrates phosphorylation-dependent and prolyl isomerization-dependent signaling.[49]

The switching rates at prolyl bonds in folded

proteins can be determined by the coupling with conformational events, through catalysis by prolyl isomerases, or by the local environment of the proline residues. In G3P it might indeed be the local sequence around Pro213 that determines the low switching rate at this proline.

The *cis* Pro213 in its peculiar sequence context complicates the folding of G3P* because the native state is reached very slowly and thus the risk of side-reactions, such as aggregation, is increased. So, why is this *cis* proline residue retained? Previously we had suggested that the slow domain docking might be important for the function of G3P.[13] In the course of infection, N1 and N2 must come apart to expose the binding site for TolA on the N1 domain, and this open state should persist until N1 and TolA are close enough to interact with each other. N1 and N2 become separated upon binding of N2 to the tip of the F pilus. The slow switch provided by Pro213 *cis* to *trans* isomerization would ensure that domain reassembly is suppressed long enough so that N1 can interact with TolA after the pilus has been retracted.

Structural basis of proline-limited domain docking in G3P

How could the function of the hinge subdomain of N2 be controlled by Pro213 isomerization? The central part of the hinge is formed by two non-contiguous segments of the polypeptide chain that run in opposite directions (Figure 7(a)). The first segment (residues 88–123) connects the glycine-rich linker between N1 and N2 with the globular subdomain of N2. The second segment (residues 203–217) is located after the globular subdomain of N2 and leads into the other linker that connects N2 with the CT domain.

Many contacts between the hinge and the N1 domain involve the chain regions 92–105 and 208–216 (Figure 7(a)). As mentioned already, Pro213 is in a local environment that is rich in proline residues (P211–Q212–P213–P214) but, moreover, the neighboring chain segment 92–98 contains two proline residues (Pro96 and Pro98). As a consequence of the clustering of proline residues, this region is well ordered and shows low crystallographic *B*-factors.[9,11] Numerous hydrogen bonds exist between this segment of the hinge subdomain and domain N1 (Figure 7(b)). Notably, the side-chain amide of Gln212, which precedes the *cis* peptide bond, forms two H bonds with the side-chain amide group of Gln52 in N1. It is also hydrogen bonded with the CO of Pro211 and thus fixes the backbone adjacent to the *cis* Gln212-Pro213 bond. The Gln52 in N1 is flanked by two residues of the hinge: Pro214 and the aromatic ring of Tyr92. The hydroxyl group of Tyr92 is hydrogen bonded to the OH group of Thr18 in N1. The NH group of Val215 maintains a backbone hydrogen bond with the CO group of Ser16, and the side-chain OH group of Ser16 engages in a hydrogen bond with the main-chain CO group of

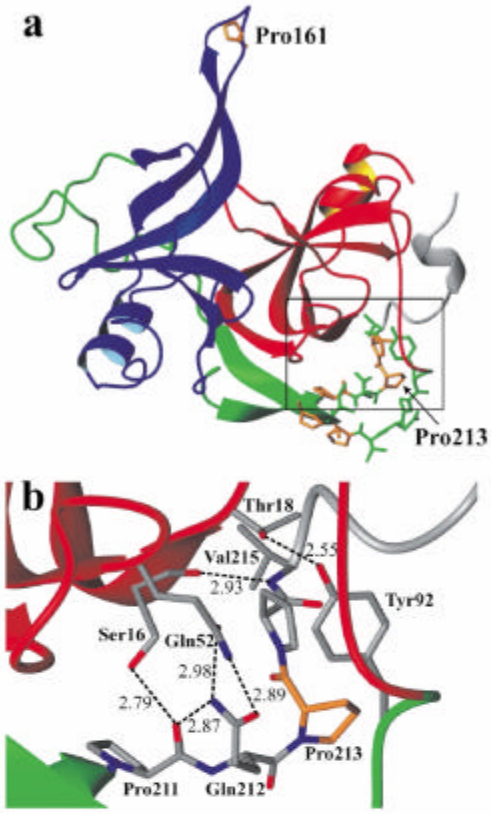


Figure 7. (a) Tertiary structure of G3P* (coordinates from Holliger *et al.*[11]). Domain N1 is shown in red, the globular subdomain of N2 (N2′) in blue, and the hinge subdomain of N2 in green. The regions Tyr92-Pro98 and Pro211-Val215 of the hinge are shown in stick representation, the five Pro residues in these stretches are colored gold. The two *cis* proline residues of G3P*, Pro161 and Pro213 are labelled. (b) A detailed view of the region framed in (a). The residues Pro211-Val215 and Tyr92 of domain N2 as well as Ser16, Thr18 and Gln52 of domain N1 are shown in stick representation. The *cis* Pro213 is colored gold. Hydrogen bonds between residues are indicated by broken lines, the distances between the respective heteroatoms are given in Å. The Figure was prepared by using MolMol.[32]

Pro211 (Figure 7(b)). We suggest that this network of interdomain interactions around the Gln212-Pro213 bond is required for the firm docking between N1 and N2. It will be lost when this prolyl bond isomerizes from the native *cis* to the incorrect *trans* conformation.

Based on the crystal structures of the N1–N2 fragments of the phages fd and M13, Holliger *et al.* suggested that domain N2 can rotate with respect to N1 about a "hinge" located at residues Gly99 and Ser208.[11] With our data, it is tempting to speculate that the isomerization at Pro213 triggers a movement of the entire N2 domain beyond this "hinge" relative to N1. Further work will be required to elucidate how the binding of the F pilus to the rim of N2 (mainly to the loop region

178–199[50]) is translated into a structural rearrangement or unfolding of the hinge subdomain to induce domain dissociation and how the *cis/trans* isomerization at Pro213 determines the kinetics of this process.

Materials and Methods

Expression and purification of G3P* and of the isolated domain N1

For the expression of IIHY-G3P* (residues 1–217 of mature G3P, plus ProSerGly(His)$_6$) the gene fragment coding for the two N-terminal domains of G3P with the stabilizing mutations T13I, T101I, Q129H and D209Y was PCR-amplified from the corresponding single-stranded phage DNA and cloned into the expression plasmid pET11a (Novagen, Madison, WI) *via* its *Nde*I and *Bam*HI restriction sites. The P213G variant was constructed by site-directed mutagenesis using the QuikChange method (Stratagene, La Jolla, USA). The proteins were overproduced in *E. coli* BL21(DE3)pLysS (Stratagene, La Jolla, USA) and purified as described.[12]

Unfolding and refolding kinetics

All GdmCl-induced unfolding and refolding experiments were performed in 100 mM potassium phosphate (pH 7.0) at 25 °C, at a final protein concentration of 0.5 μM. The kinetics were measured after a manual tenfold dilution of the native or denatured (in 5.0 M GdmCl) protein with GdmCl solutions of varying concentrations to give final concentrations of GdmCl of 1.5–5.0 M for unfolding and 0.25–2.5 M for refolding. The changes in Tyr fluorescence (emission at 310 nm after excitation at 280 nm) or Trp fluorescence (emission at 360 nm after excitation at 295 nm) were monitored in a 10 mm-cell with a Hitachi F-4010 fluorescence spectrometer at a response of 0.5 s and band widths of 5 nm and 10 nm for excitation and emission, respectively. To obtain the microscopic rate constants and kinetic *m*-values of unfolding and refolding, the apparent rate constants as a function of the concentration of GdmCl were analyzed according to a two-state model.[51]

Kinetic two-step unfolding assay for native molecules

Unfolded protein (50 μM in 5.0 M GdmCl) was first tenfold diluted by manual mixing with buffers of various concentrations of GdmCl to initiate refolding at 0.5, 1.0, 1.5 and 2.0 M GdmCl. Then, after times of refolding between one minute and 540 minutes, the protein was unfolded again at 5.0 M GdmCl by manual tenfold dilution. For measuring the catalyzed refolding, the same double-mixing experiments were performed with the refolding step at 0.5 M GdmCl in the presence of 1.0 μM cyclophilin 18. The slow unfolding reaction ($\tau = 25$ s) of domain N2 in the second step of the two-step assays was monitored by the change in Tyr fluorescence at 310 nm after excitation at 280 nm.

To measure the kinetics of domain docking for the P213G variant, the double-mixing experiments were performed in the same way as for IIHY-G3P*, but in the second step the protein was unfolded at 3.5 M GdmCl ($\tau = 32$ s). All experiments were carried out at 25 °C in 100 mM potassium phosphate (pH 7.0).

Prolyl isomerization in the denatured state measured by a triple-jump assay

The *cis/trans* interconversions in the denatured state of IIHY-G3P* and its P213G variant were measured by a triple-mixing procedure. In step 1, the native protein was unfolded at 5.0 M GdmCl for periods between 60 s and 720 s, in step 2 it was refolded by a manual tenfold dilution to 0.5 M GdmCl for 180 s, and in step 3 it was unfolded again by a further tenfold dilution to 5.0 M GdmCl (in the case of IIHY-G3P*) or to 3.5 M GdmCl (in the case of the P213G variant). The final protein concentrations were 0.5 μM for IIHY-G3P* and 0.38 μM for the P213G variant. The slow unfolding reaction in step 3 was followed by Tyr fluorescence at 310 nm after excitation at 280 nm and by Trp fluorescence at 360 nm after excitation at 295 nm.

Acknowledgements

We thank the members of our laboratory for many discussions. This work was supported by grants from the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie.

References

1. Garel, J.-R. (1992). Folding of large proteins:multidomain and multisubunit proteins. In *Protein Folding* (Creighton, T. E., ed.), pp. 405–454, Freeman, New York.
2. Jaenicke, R. (1999). Stability and folding of domain proteins. *Prog. Biophys. Mol. Biol.* **71**, 155–241.
3. Netzer, W. J. & Hartl, F. U. (1997). Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature*, **388**, 343–349.
4. Rudolph, R., Siebendritt, R., Neslauer, G., Sharma, A. K. & Jaenicke, R. (1990). Folding of an all-beta protein: independent domain folding in gamma-II-crystallin from calf eye lens. *Proc. Natl Acad. Sci. USA*, **87**, 4625–4629.
5. Privalov, P. L. (1982). Stability of proteins. Proteins which do not present a single cooperative system. *Advan. Protein Chem.* **35**, 1–104.
6. Brandts, J. F., Hu, C. Q. & Lin, L.-N. (1989). A simple model for proteins with interacting domains. Application to scanning calorimetry data. *Biochemistry*, **28**, 8588–8596.
7. Boeke, J. D. & Model, P. (1982). A prokaryotic membrane anchor sequence: carboxyl terminus of bacteriophage f1 gene III protein retains it in the membrane. *Proc. Natl Acad. Sci. USA*, **79**, 5200–5204.
8. Stengele, I., Bross, P., Garces, X., Giray, J. & Rasched, I. (1990). Dissection of functional domains in phage fd adsorption protein. Discrimination between attachment and penetration sites. *J. Mol. Biol.* **212**, 143–149.
9. Lubkowski, J., Hennecke, F., Plückthun, A. & Wlodawer, A. (1998). The structural basis of phage display elucidated by the crystal structure of the

N-terminal domains of G3p. *Nature Struct. Biol.* **5**, 140–147.
10. Lubkowski, J., Hennecke, F., Plückthun, A. & Wlodawer, A. (1999). Filamentous phage infection: crystal structure of g3p in complex with its coreceptor, the C-terminal domain of TolA. *Structure*, **7**, 711–722.
11. Holliger, P., Riechmann, L. & Williams, R. L. (1999). Crystal structure of the two N-terminal domains of g3p from filamentous phage fd at 1.9 Angström: evidence for conformational lability. *J. Mol. Biol.* **288**, 649–657.
12. Martin, A. & Schmid, F. X. (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**, 863–875.
13. Martin, A. & Schmid, F. X. (2003). Folding and domain docking of the phage fd gene-3-protein. *J. Mol. Biol.* **329**, 599–610.
14. Tanford, C. (1968). Protein denaturation. Part B. The transition state from native to denatured state. *Advan. Protein Chem.* **23**, 218–282.
15. Tanford, C. (1970). Protein denaturation. Part C. *Advan. Protein Chem.* **24**, 1–95.
16. Myers, J. K., Pace, C. N. & Scholtz, J. M. (1995). Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4**, 2138–2148.
17. Schmid, F. X. (1993). Prolyl isomerase—enzymatic catalysis of slow protein-folding reactions. *Annu. Rev. Biophys. Biomol. Struct.* **22**, 123–143.
18. Fischer, G., Bang, H. & Mech, C. (1984). Nachweis einer enzymkatalyse für die *cis–trans*-isomerisierung der peptidbindung in prolinhaltigen peptiden. *Biomed. Biochim. Acta*, **43**, 1101–1111.
19. Fischer, G., Wittmann-Liebold, B., Lang, K., Kiefhaber, T. & Schmid, F. X. (1989). Cyclophilin and peptidyl-prolyl-*cis/trans*-isomerase are probably identical proteins. *Nature*, **337**, 476–478.
20. Brandts, J. F., Halvorson, H. R. & Brennan, M. (1975). Consideration of the possibility that the slow step in protein denaturation reactions is due to *cis–trans* isomerism of proline residues. *Biochemistry*, **14**, 4953–4963.
21. Schmid, F. X. (1986). Fast-folding and slow-folding forms of unfolded proteins. In *Enzyme Structure Part L* (Hirs, C. H. W. & Timasheff, S. N., eds), vol. 131, pp. 71–82, Academic Press, New York.
22. Reimer, U., Scherer, G., Drewello, M., Kruber, S., Schutkowski, M. & Fischer, G. (1998). Side-chain effects on peptidyl-prolyl *cis/trans* isomerisation. *J. Mol. Biol.* **279**, 449–460.
23. Scherer, G., Kramer, M. L., Schutkowski, M. U. R. & Fischer, G. (1998). Barriers to rotation of secondary amide peptide bonds. *J. Am. Chem. Soc.* **120**, 5568–5574.
24. Schiene-Fischer, C. & Fischer, G. (2001). Direct measurement indicates a slow *cis/trans* isomerization at the secondary amide peptide bond of glycylglycine. *J. Am. Chem. Soc.* **123**, 6227–6231.
25. Odefey, C., Mayr, L. M. & Schmid, F. X. (1995). Non-prolyl *cis–trans* peptide bond isomerization as a rate-determining step in protein unfolding and refolding. *J. Mol. Biol.* **245**, 69–78.
26. Vanhove, M., Raquet, X., Palzkill, T., Pain, R. H. & Frere, J. M. (1996). The rate-limiting step in the folding of the *cis*-Pro167Thr mutant of TEM-1 beta-lactamase is the *trans* to *cis* isomerization of a non-proline peptide bond. *Proteins: Struct. Funct. Genet.* **25**, 104–111.
27. Wheeler, K. A., Hawkins, A. R., Pain, R. & Virden, R. (1998). The slow step of folding of Staphylococcus aureus PC1 beta-lactamase involves the collapse of a surface loop rate limited by the *trans* to *cis* isomerization of a non-proline peptide bond. *Proteins: Struct. Funct. Genet.* **33**, 550–557.
28. Stein, R. L. (1993). Mechanism of enzymatic and nonenzymatic prolyl *cis–trans* isomerization. *Advan. Protein Chem.* **44**, 1–24.
29. Cook, K. H., Schmid, F. X. & Baldwin, R. L. (1979). Role of proline isomerization in folding of ribonuclease A at low temperatures. *Proc. Natl Acad. Sci. USA*, **76**, 6157–6161.
30. Mayr, L. M., Odefey, C., Schutkowski, M. & Schmid, F. X. (1996). Kinetic analysis of the unfolding and refolding of ribonuclease T_1 by a stopped-flow double-mixing technique. *Biochemistry*, **35**, 5550–5561.
31. Schreiber, G. & Fersht, A. R. (1993). The refolding of *cis*- and *trans*-peptidylprolyl isomers of barstar. *Biochemistry*, **32**, 11195–11203.
32. Kamen, D. E. & Woody, R. W. (2002). Folding kinetics of the protein pectate lyase C reveal fast-forming intermediates and slow proline isomerization. *Biochemistry*, **41**, 4713–4723.
33. Kelley, R. F. & Richards, F. M. (1987). Replacement of proline-76 with alanine eliminates the slowest kinetic phase in thioredoxin folding. *Biochemistry*, **26**, 6765–6774.
34. Osterhout, J. J., Jr & Nall, B. T. (1985). Slow refolding kinetics in yeast iso-2 cytochrome c. *Biochemistry*, **24**, 7999–8005.
35. Schmid, F. X. & Blaschek, H. (1981). A native-like intermediate on the ribonuclease A folding pathway. 2. Comparison of its properties to native ribonuclease A. *Eur. J. Biochem.* **114**, 111–117.
36. Kiefhaber, T., Quaas, R., Hahn, U. & Schmid, F. X. (1990). Folding of ribonuclease T_1. 2. Kinetic models for the folding and unfolding reactions. *Biochemistry*, **29**, 3061–3070.
37. Wyckoff, H. W., Tsernoglou, D., Hanson, A. W., Knox, J. R., Lee, B. & Richards, F. M. (1970). The three-dimensional structure of ribonuclease-S. Interpretation of an electron density map at a nominal resolution of 2 Å. *J. Biol. Chem.* **245**, 305–328.
38. Balbach, J. & Schmid, F. X. (2000). Prolyl isomerization and its catalysis in protein folding. In *Mechanisms of Protein Folding* (Pain, R. H., ed.), pp. 212–237, Oxford University Press, Oxford.
39. Fischer, G. (1994). Peptidyl-prolyl *cis/trans* isomerases and their effectors. *Angew. Chem. Int. Ed.* **33**, 1415–1436.
40. Schmid, F. X. (2002). Prolyl isomerases. *Advan. Protein Chem.* **59**, 243–282.
41. Schmid, F. X., Lang, K., Kiefhaber, T., Mayer, S. & Schönbrunner, R. (1991). Prolyl isomerase. Its role in protein folding and speculations on its function in the cell. In *Conformations and Forces in Protein Folding* (Nall, B. T. & Dill, K. A., eds), American Association for the Advancement of Science, Washington, DC.
42. Mallis, R. J., Brazin, K. N., Fulton, D. B. & Andreotti, A. H. (2002). Structural characterization of a proline-driven conformational switch within the Itk SH2 domain. *Nature Struct. Biol.* **9**, 900–905.
43. Brazin, K. N., Mallis, R. J., Fulton, D. B. & Andreotti, A. H. (2002). Regulation of the tyrosine kinase Itk by

the peptidyl-prolyl isomerase cyclophilin A. *Proc. Natl Acad. Sci. USA*, **99**, 1899–1904.
44. Evans, P. A., Dobson, C. M., Kautz, R. A., Hatfull, G. & Fox, R. O. (1987). Proline isomerism in staphylococcal nuclease characterized by NMR and site-directed mutagenesis. *Nature*, **329**, 266–268.
45. Kördel, J., Forsen, S., Drakenberg, T. & Chazin, W. J. (1990). The rate and structural consequences of proline *cis–trans* isomerization in calbindin D9k: NMR studies of the minor (*cis* Pro43)isoform and the Pro43Gly mutant. *Biochemistry*, **29**, 4400–4409.
46. Yuan, X., Werner, J. M., Knott, V., Handford, P. A., Campbell, I. D. & Downing, K. (1998). Effects of proline *cis–trans* isomerization on TB domain secondary structure. *Protein Sci.* **7**, 2127–2135.
47. Gitti, R. K., Lee, B. M., Walker, J., Summers, M. F., Yoo, S. & Sundquist, W. I. (1996). Structure of the amino-terminal core domain of the HIV-1 capsid protein. *Science*, **273**, 231–235.
48. Ramelot, T. A. & Nicholson, L. K. (2001). Phosphorylation-induced structural changes in the amyloid precursor protein cytoplasmic tail detected by NMR. *J. Mol. Biol.* **307**, 871–884.
49. Yaffe, M. B., Schutkowski, M., Shen, M. H., Zhou, X. Z., Stukenberg, P. T., Rahfeld, J. U. *et al.* (1997). Sequence-specific and phosphorylation-dependent proline isomerization—a potential mitotic regulatory mechanism. *Science*, **278**, 1957–1960.
50. Deng, L. W. & Perham, R. N. (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**, 603–614.
51. Kiefhaber, T. (1995). Protein folding kinetics. *Methods Mol. Biol.* **40**, 313–341.
52. Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55.

Edited by C. R. Matthews

(Received 9 April 2003; received in revised form 16 June 2003; accepted 1 July 2003)

Danksagung

Zum Schluss möchte ich allen ganz herzlich danken, die an der Entstehung dieser Arbeit beteiligt waren:

Die vorliegende Arbeit wurde von September 1998 bis Februar 2003 am Lehrstuhl für Biochemie der Universität Bayreuth unter der Leitung von Prof. Dr. Franz X. Schmid angefertigt. Ihm gilt mein besonderer Dank, für die ausgezeichnete wissenschaftliche Betreuung, die gewährte Freiheit und das Vertrauen bei der Entwicklung neuer Projekte, und das stete Interesse am Fortgang meiner Arbeit. Mit zahlreichen guten Ratschlägen, nicht nur wissenschaftlicher Art, und der permanenten Hilfsbereitschaft machte er der Bezeichnung „Doktorvater“ alle Ehre.

Vielen Dank an meine Laborkolleginnen und -kollegen Jochen Balbach, Barbara Eckert, Maik Jacob, Insa Kather, Christine Magg, Peter Maier, Raimund Maier, Dieter Perl, Michael Rape, Christian Scholz, Michael Seewald, Volker Sieber, Claudia Staab, Michael Wunderlich, und Markus Zeeb für die gute Zusammenarbeit und die hervorragende Laboratmosphäre.

Bei Volker Sieber bedanke ich mich für die Einführung in die Methoden der gerichteten Evolution.

Manuela Schaffrath und Prof. Dr. Carlo Unverzagt möchte ich für die Kooperation bei der Peptidsynthese, -reinigung und -analyse danken.

Ein Dankeschön auch an Prof. Terry Oas für die tolle Zeit während seines Forschungsaufenthalts in Bayreuth.

Dem Graduiertenkolleg „Biosynthese der Proteine und Regulation ihrer Aktivität“ und seinem Sprecher Prof. Dr. Mathias Sprinzl sowie dem Fonds der Chemischen Industrie danke ich für die freundliche Unterstützung während meiner Zeit als Doktorand.

Und ein Dankeschön an alle Pro- und Eukaryonten, die zum Gelingen dieser Arbeit beigetragen haben…☺

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, eine Dissertation einzureichen oder mich der Doktorprüfung zu unterziehen.

Bayreuth, den 20. Februar 2003