Das Gen-3-Protein filamentöser Phagen als Modellsystem zur Untersuchung von Proteinstabilität, Faltungsmechanismen und Prolylisomerisierung

DISSERTATION

zur Erlangung des Grades

– Doktor der Naturwissenschaften –

der Fakultät für Biologie, Chemie und Geowissenschaften

der Universität Bayreuth

vorgelegt von

Diplom-Biochemiker

Roman Jakob

Bayreuth 2009

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 Januar 2005 bis Juni 2009 am Lehrstuhl für Biochemie der Universität Bayreuth unter der Anleitung von Prof. Dr. Franz X. Schmid angefertigt.

Promotionsgesuch eingereicht am:	1. Juli 2009
Tag des wissenschaftlichen Kolloquiums:	11. August 2009

Prüfungsausschuss:

Prof. Dr. Franz Xaver Schmid	(Erster Gutachter)
Prof. Dr. Matthias Ullmann	(Zweiter Gutachter)
Prof. Dr. Stephan Clemens	
Prof. Dr. Carlo Unverzagt	(Vorsitzender)

Für Sandra

Inhaltsverzeichnis

Kurzfassung der Arbeit

Proteine sind die funktionstragenden Moleküle in lebenden Zellen und ihre pharmazeutische und biotechnologische Anwendung gewinnt zunehmend an Bedeutung. Um ihre Funktion zu erlangen, müssen Proteine falten und die gefaltete Struktur möglichst lange beibehalten. Die Faltungsmechanismen und die Grundlagen der konformationellen Stabilität von Proteinen zu verstehen, ist daher eines der zentralen Ziele der biophysikalischen und biochemischen Forschung. Im Mittelpunkt dieser Arbeit stand das Gen-3-Protein des Phagen fd. Sein Faltungsmechanismus ist aufs Engste mit seiner biologischen Funktion verknüpft, und es ist ein Modellsystem für die Untersuchung von Proteinfaltung, -stabilität und Prolylisomerisierung.

Durch gerichtete Evolution mittels der *Proside*-Selektionsmethode konnten im Rahmen vorangegangener Arbeiten stark stabilisierte Varianten des Gen-3-Proteins und der TEM-1 β-Lactamase erhalten werden. Um die molekularen Ursachen dieser starken Stabilisierungen zu verstehen, wurden hier ihre dreidimensionalen Strukturen mittels Röntgenkristallographie bestimmt. Die Analyse zeigte für das Gen-3-Protein eine optimierte hydrophobe Packung und verbesserte Domäneninteraktion als Stabilisierungsprinzip. Bei der β-Lactamase führte die Optimierung von Helices, speziell an deren Aminotermini und die verbesserte Packung von Helices zur Stabilisierung.

Im weiteren Verlauf stand die isolierte N2-Domäne des Gen-3-Proteins im Mittelpunkt der Arbeit. Dabei zeigte sich, dass in der nativen N2-Domäne ein *cis*/*trans*-Gleichgewicht an Pro161 existiert, das entscheidend für die Stabilität und Faltung des Proteins ist. Solche Heterogenitäten an Prolinresten werden als potentiell wichtige regulatorische Schalter in biologischen Systemen diskutiert, konnten aber bislang kaum untersucht werden. Während der Faltung von N2 wird der *cis*-Anteil an Pro161 von 7 % im denaturierten Protein auf 89 % im nativen Protein erhöht, was einem Energieaufwand von 10 kJ mol^{-1} entspricht. Obwohl sich Pro161 exponiert an der Spitze einer Schleife befindet, wird die Hälfte der insgesamt verfügbaren Konformationsenergie des nativen Proteins verwendet, um das *cis*/*trans* Gleichgewicht an Pro161 zum *cis* Isomer hin zu verschieben.

Um diese Kopplung auf molekularer Ebene zu verstehen, wurde die lokale Sequenzumgebung von Pro161 systematisch variiert, sowohl im Schleifenbereich um Pro161 selbst als auch in den beiden Faltblattsträngen, die in diese Schleife einmünden. In der *cis*-Form zeigte die Schleife um Pro161 eine definierte native Konformation. Die energetische Kopplung zwischen der *cis*-Form an Pro161 und den β-Faltblattsträngen ist dabei schon im Übergangszustand der Faltung voll ausgebildet. In der *trans*-Form hingegen ist die Schleife entfaltet, und es besteht keine energetische Kopplung mit dem β-Faltblatt. Lokale Ent- und Rückfaltung wird hier also verwendet, um die energetische Kommunikation zwischen einer Prolylisomerisierung und Sekundärstrukturelementen zu brechen und wieder herzustellen. Die Beiträge der einzelnen Bereiche der N2-Domäne zur Struktur und zur Energetik des Übergangszustands der Faltung wurde mit Hilfe einer Φ-Wert-Analyse analysiert. Es sind zwar fast alle Bereiche für die Stabilität des nativen Proteins wichtig, den Faltungsprozeß hingegen bestimmt nur ein eng begrenzter Faltungsnukleus, der interessanterweise die Schleife um Pro161 und die beiden Faltblattstränge links und rechts davon beinhaltet.

In lebenden Zellen steht eine Vielzahl von Faltungshelfern bereit, um Proteine auf ihrem „Lebensweg“ zu begleiten. Eine wichtige Klasse sind dabei die Prolylisomerasen, die die *cis*-*trans*-Isomerisierung an Xaa-Pro Bindungen katalysieren. Die *trans*→*cis*-Isomerisierung an Pro161 in der N2-Domäne lässt sich sehr gut durch Prolylisomerasen katalysieren. Um die Substratspezifität von Prolylisomerasen der FKBP-Familie untersuchen zu können, wurde eine Bibliothek von N2-Varianten mit allen natürlichen Aminosäuren vor Pro161 hergestellt. Prolylisomerasen, die nur aus einer katalytischen Domäne bestehen, zeigen für kurze Peptide und für faltende Proteine die gleiche Substratspezifität, mit einer sehr starken Präferenz für hydrophobe Reste vor Prolin. Besitzen die Prolylisomerasen zusätzlich eine Chaperondomäne, wie SlyD oder Triggerfaktor, dann ist ihre Aktivität praktisch unabhängig von der Aminosäure vor Prolin. Der Grund dafür ist, dass die enzymkinetischen Parameter der katalysierten Faltung, die Affinität und die Umsatzzahl durch die Chaperondomäne bestimmt werden. Diese Bibliothek an N2-Varianten mit allen Aminosäuren vor Pro161 hat sich als ein sehr gutes Substrat erwiesen und wird in Zukunft auch verwendet werden, um Mitglieder anderer Prolylisomerasefamilien zu charakterisieren.

Die N2-Domäne besitzt nur eine marginale Stabilität. Deshalb wurde mit diesem Protein untersucht, inwieweit Proteine durch rationales *consensus design* von β-Turns stabilisiert werden können. Für N2 konnte durch Sequenzoptimierung von drei β-Turns die Stabilität tatsächlich praktisch verdoppelt werden, was die generelle Anwendbarkeit dieser Methode verdeutlicht. Interessanterweise führt diese Turnoptimierung auch zu einer sehr starken Beschleunigung der Faltung von N2, offensichtlich weil diese β-Turns bereits sehr früh im Faltungsprozess von Proteinen wichtig sind, da sie Bildung von nativen Interaktionen in den benachbarten Bereichen ermöglichen.

Im Gen-3-Protein des Phagen fd tritt während der *in-vitro*-Faltung ein langlebiges Intermediat mit einem *trans*-Prolin213 auf. Mittels NMR-Spektroskopie wurde herausgefunden, das in diesem Intermediat die N1- und die N2-Domäne bereits ihre native chemische Verschiebung zeigen, aber die Gelenkdomäne zwischen den Domänen nur teilweise strukturiert ist. In der anschließenden Domänenassemblierungsreaktion werden als Folge der *trans*→*cis* Isomerisierung an Pro213 besonders Wasserstoffbrückenbindungen stark stabilisiert, die eine Kette zwischen der Grenzfläche von N1 und der Gelenkregion bilden. Die starke Stabilisierung dieses Netzes an Wasserstoffbrücken liefert vermutlich die strukturelle Basis und die energetische Triebkraft für die feste Domänenassoziation des Gen-3-Proteins. Die Ergebnisse zur Kopplung zwischen *cis*/*trans* Isomeren an Pro213 entsprechen konzeptionell dem Mechanismus an Pro161 in der N2-Domäne. Durch diese beiden Arbeiten beginnen wir langsam zu verstehen, wie Veränderungen im Isomerenzustand an Prolinresten in einer spezifischen und gerichteten Art und Weise zu anderen Regionen im Protein weitergeleitet werden können.

Am Beispiel des Gen-3-Proteins des Phagen IKe wurde untersucht, ob alle Gen-3-Proteine filamentöser Phagen eine ähnliche Struktur und Faltungsmechanismus besitzen. Dabei zeigte sich, dass die TolA-bindenden Domänen beider Proteine strukturell und funktionell stark konserviert sind. Die jeweiligen pilusbindenden Domänen sind dagegen architektonisch komplett verschieden, da sie spezifisch an den jeweiligen Wirt adaptiert wurden. Deshalb sind vermutlich die Faltungsmechanismen dieser Mehrdomänenproteine, und damit auch die Infektionsmechanismen grundlegend verschieden.

Abstract

Proteins are the vital molecules in living cells, and at the same time they are of increasing importance for pharmaceutical and biotechnological applications. To reach their function, these proteins have to fold and retain their native structure for a long time. Thus, the folding mechanism and the fundamentals of conformational protein stability are major objectives in biochemical and biophysical research. The focus of my work was on the gene-3-protein of the filamentous phage fd. Its folding mechanism is strongly interrelated with its biological function, and it is used as a model protein for the investigation of protein folding, protein stability and prolyl isomerization.

In previous work, strongly stabilized variants of the gene-3-protein and of TEM-1 β-lactamase were obtained by directed evolution using the *Proside* (*protein stability increased by directed evolution*) selection method. To understand the molecular basis of this strong stabilization, the crystal structures of these proteins were determined. The analysis revealed that the stabilization of the gene-3-protein originates mostly from optimized hydrophobic packing and from improved domain interaction. β-lactamase was stabilized by optimization of helices, in particular at their amino termini, and by improved helix packing.

In the further course of my work the isolated N2 domain of the gene-3-protein was investigated. Interestingly, the folded N2 domain exhibits a *cis*/*trans* equilibrium at Pro161 which is important for stability and folding of the protein. Heterogeneities at proline residues are discussed as potential important regulators in biological systems, but could hitherto not be studied in detail. During the folding of N2 the *cis*-content increases from 7 % in the unfolded protein to 89 % in the native protein, which requires the input of a folding free energy of 10 kJ mol^{-1}. Although Pro161 is located at the tip of an exposed loop, half of the available conformational energy is thus used to shift the *cis*/*trans* equilibrium at Pro161 to the *cis* isomer.

To elucidate the molecular path for this energetic coupling, the local sequence around Pro161 was systematically varied, in the loop region of Pro161 as well as in the two β-strands that lead into the loop. In the *cis* form of Pro161 the loop shows a well ordered native structure and the energetic linkage between the *cis* form of Pro161 and the β-strands is fully established in transition state of folding already. In the presence of a *trans*-Pro161 the loop is locally unfolded and Pro161 is structurally and energetically uncoupled from the β-sheet. Local unfolding and refolding are thus used to break and to reestablish the energetic communication between a prolyl isomerization and the stability of an element of secondary structure (in this case a two stranded β-sheet).

The contributions of the individual regions of the N2 domain for the structure and the energetics of the transition state of folding were analyzed using Φ-value analysis. All parts of the native protein are important for stability, but the folding process is determined by a locally restricted folding nucleus which, interestingly, includes the loop around Pro161 and the adjacent β sheets.

In living cells plenty of folding helpers accompany a protein from the cradle to the grave. An important class are the prolyl isomerases that catalyze the isomerization at Xaa-Pro

peptide bonds. The *trans→cis* isomerization at Pro161 in the N2 domain is catalyzed very well by prolyl isomerases. To characterize the substrate specificity of prolyl isomerases of the FKBP family, a library of N2 variants with all natural amino acids before Pro161 was produced. Prolyl isomerases that consist of a catalytic domain only, show high substrate specificities for both peptides and proteins, with a strong preference for hydrophobic residues before proline. When the prolyl isomerases contain an additional chaperone domain, such as in the case of SlyD or trigger factor, their activity becomes independent of the amino acid before proline. The reason for this is that the enzymatic parameters of the catalyzed folding, the affinity and the turn over number are determined by the chaperone domain. This library of N2 variants with every amino acid before proline turns out to be an excellent collection of substrates and will be used in future to characterize other prolyl isomerase families as well.

The N2 domain of the phage gene-3-protein exhibits a low stability. Therefore, this protein was used to explore to which extent rational *consensus design* of β-turns can be used for stabilization. The sequence optimization of three β-turns in N2 virtually doubled its conformational stability, which illustrates the general applicability of this method. Interestingly, the turn optimization led to a strong acceleration of refolding of N2, apparently, because β-turns are important in the early steps of protein folding, since they allow the formation of native interactions in adjacent parts of the molecule.

During the *in-vitro* folding of the gene-3-protein a persistent intermediate with a *trans*-Pro213 is populated. Using NMR spectroscopy we found out that, in this intermediate, the N1 and the N2 domains have native chemical shifts already, but the hinge region between these domains is only partially structured. In the subsequent domain assembly reaction hydrogen bonds were strongly stabilized that form a chain of interactions between N1 and the hinge region. The strong stabilization of a defined cluster of hydrogen bonds presumably provides the structural and energetic explanation for how the *trans→cis* isomerization at Pro213 controls the tight domain association in the folding of the gene-3-protein. The results on the coupling between *cis*/*trans* isomers at Pro213 are conceptually related to the mechanism of how the isomerization at Pro161 is interrelated with the conformational folding of the N2 domain. Due to both sets of results we now begin to understand how changes in the isomeric state at proline residues can be propagated in a specific and directional fashion in a folding or a folded protein.

In the last part of this work, I used the gene-3-protein of the phage IKe to examine, if all gene-3-proteins of different types of filamentous phage have similar structure and folding mechanisms. The TolA binding domains of the two proteins are in fact structurally and functionally highly conserved. In contrast, the respective pilus binding domains and the domain interactions are radically different. As a consequence the folding and infection mechanisms are not conserved in the family of filamentous bacteriophages.

1 Einleitung

1.1 Struktur und Stabilität von Proteinen

Proteine stellen die strukturell und funktionell vielseitigste Klasse biologischer Makromoleküle dar und übernehmen vielfältige Aufgaben in allen biologischen Systemen. Neben der Aminosäuresequenz ist die definierte Raumstruktur, die aus der Proteinfaltung resultiert, für die biologische Funktion eines Proteins essentiell. Die Stabilität einer Proteinstruktur ist dabei an die Erfordernisse *in vivo* angepasst. Sie muss einerseits groß genug sein, um die native Konformation eines Proteins aufrechtzuerhalten, darf aber andererseits die konformationelle Flexibilität nicht einschränken, da eine hohe Dynamik für die Funktion vieler Proteine unerlässlich ist [1]. Stabilität, Faltungsmechanismus und biologische Aktivität eines Proteins müssen demnach in einem fein abgestimmten Gleichgewicht stehen, damit es durch geringe Änderungen auf unterschiedliche Anforderungen reagieren kann.

Die Stabilität von Proteinen resultiert aus einer Vielzahl schwacher nichtkovalenter Wechselwirkungen, deren relative Beiträge immer noch diskutiert werden. Den Hauptanteil liefert dabei der hydrophobe Effekt, das Abschirmen unpolarer Seitenketten im Inneren von gefalteten Proteinen [2; 3]. Weitere wichtige Beiträge zur Stabilität des nativen Proteins liefern Disulfid- und Wasserstoffbrückenbindungen [4] zwischen verborgenen polaren Seitengruppen [5; 6; 7] und langreichweitige ionische Wechselwirkungen [3; 7; 8].

Allgemeine Prinzipien zur Stabilisierung von Proteinen sind schwer zu definieren, da die Nettostabilität schon durch geringfügige Strukturänderungen enorm verändert werden kann. Deswegen verstehen wir bis heute nicht wirklich, wie die Stabilität eines Proteins in seiner Aminosäuresequenz kodiert ist.

Die Bedeutung von Proteinen als Wirkstoffe und Katalysatoren in der Industrie nimmt stetig zu. Deshalb sind generelle, möglichst rationale Ansätze von großem Interesse um langzeitstabile und -aktive Proteine zu entwickeln. Bisher ist es gelungen, Proteine durch Optimierung von Oberflächenladungen zu verbessern [9; 10; 11; 12; 13; 14]. In den letzten Jahren hat darüber hinaus die Optimierung von lokalen Sekundärstrukturelementen, wie α-Helices und β-Turns, an Bedeutung gewonnen [15; 16; 17; 18; 19; 20], da durch statistische Konsensussequenzanalysen deren Aminosäurepräferenzen mittlerweile detailliert untersucht wurden [21; 22; 23; 24; 25; 26; 27]. Für das *protein engineering* ist dabei das Verständnis der Prinzipien von Proteinstabilität und -faltung von grundlegender Bedeutung.

1.2 Mechanismen der Proteinfaltung

Die Proteinfaltungsforschung versucht aufzuklären, wie die eindimensionale Sequenzinformation einer Polypeptidkette verwendet wird, um die dreidimensionale Struktur eines aktiven Proteins und den Weg dorthin festzulegen. Durch jahrzehntelange Anstrengungen sind inzwischen viele Details von Faltungsvorgängen bekannt, und die generellen Mechanismen, durch die eine Polypeptidkette ihre native Konformation erreicht, werden immer besser verstanden.

Die native und biologisch aktive Form eines Proteins entspricht normalerweise der thermodynamisch stabilsten Struktur unter physiologischen Bedingungen. Lange Zeit wurde die Proteinfaltung mit einer chemischen Reaktion verglichen: der entfaltete Zustand U eines Proteins muss entlang der Reaktionskoordinate einen energiereichen Übergangszustand überwinden, um zur nativen Struktur zu finden. Dieser Reaktionsverlauf steht im Einklang mit der von Levinthal begründeten Vorstellung eines definierten, sequentiellen Faltungsweges [28; 29]. In den letzten Jahren hat sich diese Sichtweise stark verändert. Der Startpunkt einer Proteinfaltungsreaktion wird nicht mehr als eine einzige Konformation des denaturierten Zustands, sondern als eine große Anzahl vieler möglicher Zustände angesehen. Folgerichtig ist die Proteinfaltung deshalb eher durch eine Energiehyperfläche zu beschreiben, bei der eine entfaltete Proteinkette entlang eines trichterartigen Energieprofils den nativen Zustand erreicht [29; 30; 31; 32]. Die Polypeptidkette findet in solchen Modellen ihr Energieminimum durch den Prozess von Versuch und Irrtum. Nativähnliche Kontakte sind stabiler als nichtnative Kontakte und daher langlebiger. Aus diesem Grund wird die Anzahl möglicher Konformationen im Verlauf der Faltung eingeschränkt [31; 33; 34]. Durch natürliche Selektion wurden Proteine so evolviert, dass sie sehr schnell und effizient falten und fehlgefaltete Zustände vermieden werden.

Die Oberfläche einer Faltungshyperfläche ist dabei einzigartig und charakteristisch für jede Polypeptidkette und wird durch deren thermodynamische und kinetische Eigenschaften bestimmt. Auch auf derartigen Energiehyperflächen wird der Faltungsweg durch die (vermutlich stark eingeschränkten) Bereiche niedriger Energie auf wenige, oder nur einen „Weg“ verengt, sodass sich hier die “alte“ Sichtweise definierter Faltungswege und die „neue“ Sichtweise des Herumirrens auf Energiehyperflächen wieder treffen. Übergänge zwischen verschiedenen Zuständen können *in vitro* detailliert durch verschiedene spektroskopische Techniken, von optischen Methoden bis hin zur NMR Spektroskopie untersucht werden [29; 35].

In den letzten 15 Jahren, resultierten die größten Fortschritte im Verständnis der Proteinfaltung vor allem aus Studien an verschiedenen kleinen Modellproteinen (< 100 Reste). Diese sind sehr einfach handhabbar und falten häufig in einer kooperativen Zwei-Zustandsreaktion ohne kinetische Intermediate direkt in den nativen Zustand [36; 37; 38].

Der Beitrag individueller Aminosäurereste am Übergangszustand der Faltung kann durch die Untersuchung von Varianten eines spezifischen Proteins auf die Entfaltungs- und Rückfaltungskinetik ermittelt werden (Φ-Wert-Analyse). Die Analysen vieler Proteine deuteten darauf hin, dass in der Regel durch frühe Interaktionen weniger Reste ein Faltungsnukleus ausgebildet wird, um den sich anschließend die übrigen Bereiche des Proteins strukturieren [39; 40; 41]. So wurde für verschiedene SH3-Domänen, das Kälteschockprotein aus *Bacillus subtilis* und Chymotrypsininhibitor 2 gefunden, dass im Übergangszustand der Faltung nur zwei Faltblattstränge interagieren[42; 43]. Diese Ergebnisse verdeutlichen den Beitrag spezifischer Seitenkettenwechselwirkungen für den Beginn eines spezifischen Faltungsprozesses. Die energetisch treibende Kraft für die Faltung liegt hingegen in den unterschiedlichen Interaktionen des Proteinrückgrats mit dem wässrigen Medium im ungefalteten Zustand bzw. mit sich selbst im nativen Zustand begründet [44]. Es ist nach wie vor unklar, wie diese generische Stabilisierung durch Wechselwirkungen des Rückgrats durch die Seitenketteninteraktionen in eine spezifische Struktur im Faltungsverlauf übersetzt wird.

1.3 Faltung von Mehrdomänenproteinen

Der Großteil der Faltungsstudien konzentrierte sich, wie schon ausgeführt, bislang auf kleine, globuläre Proteine. Die meisten Proteine (65-80 %) sind allerdings aus mehreren Domänen aufgebaut, welche strukturelle, funktionale oder separat faltende Einheiten bilden können [45; 46; 47]. Dieser modulare Aufbau ermöglicht die Kombination unterschiedlicher Funktionen [48], wobei die Wechselwirkung von Domänen oftmals von großer Bedeutung für die biologische Regulation ist. Die Faltungsmechanismen von Mehrdomänenproteinen sind in der Regel wesentlich komplexer, als die der Eindomänenproteine [49; 50]. Je nach Stabilität der Einzeldomänen und der Wechselwirkungen zwischen den Domänen sind vier Faltungsmodelle für Mehrdomänenproteine denkbar: (i) Jede der Domänen ist eine unabhängige Faltungseinheit. (ii) Die Domänen sind isoliert stabil, interagieren jedoch im nativen Zustand miteinander. (iii) Die Faltung einer Domäne katalysiert die Faltung der anderen Domänen. (iv) Die Domänen sind isoliert instabil und ihre Faltung ist direkt an die Domänenassoziation gekoppelt.

Das isolierte N1N2-Fragment des Gen-3-Proteins (G3P*, Abb. 1-1a) des Phagen fd ist ein Modellsystem, um den Faltungsmechanismus eines Mehrdomänenproteins *in vitro* zu analysieren. Die Faltung von G3P* verläuft in mehreren, zeitlich klar getrennten Schritten und umfasst einen sehr großen Zeitbereich [52; 53]. Im ersten Schritt faltet die N1- Domäne sehr schnell innerhalb weniger Millisekunden. Die N2-Domäne bildet ihre Struktur etwa tausendfach langsamer, in zwei sequentiellen Reaktionen, aus. Dabei entspricht der erste Schritt vermutlich der Faltung des globulären Teils von N2 (N2') und der zweite Schritt der lockeren Assemblierung mit der N1-Domäne. Die feste Assoziation der beiden vorgefalteten Domänen und damit die Faltung in den stabilen nativen Zustand ist nochmals zwei Größenordnungen langsamer und besitzt eine Zeitkonstante von 6200 s (bei 25 °C). Diese langsame Domänenassoziation wird durch die *trans*→*cis*-Isomerisierung der Peptidbindung Gln212-Pro213 in der Gelenkdomäne kontrolliert [53]. Pro213 weist im nativen Zustand eine *cis*-Konformation auf [51; 54] (Abb. 1-1a).

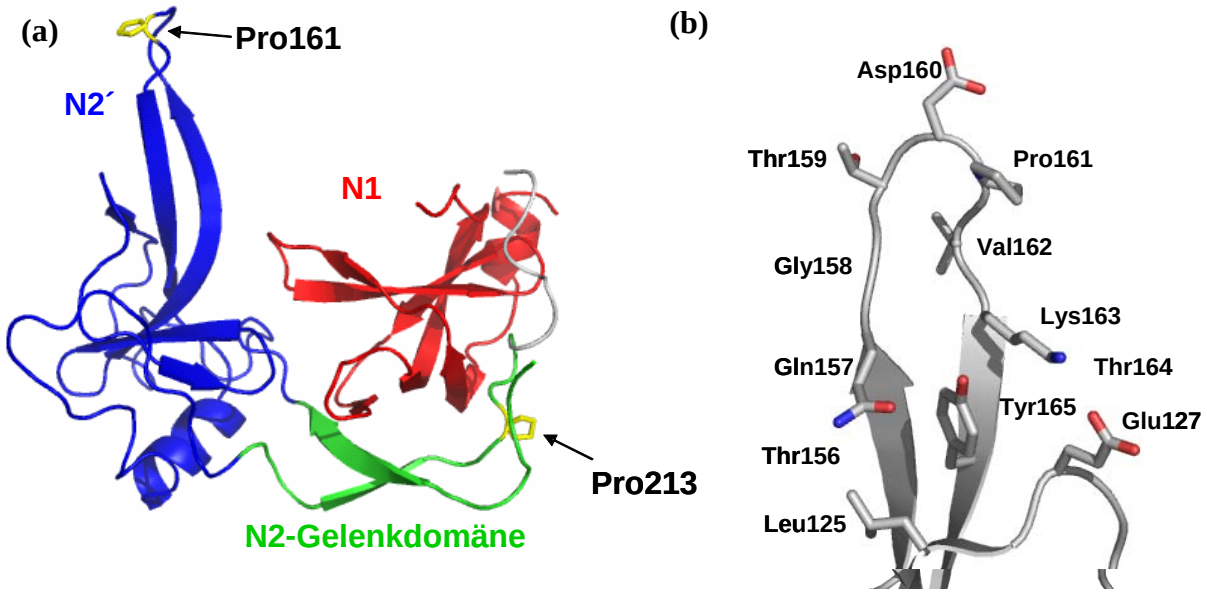


Abbildung 1-1: **(a)** Tertiärstruktur des N1N2-Fragments von G3P (G3P*). des Phagen fd Domäne N1 ist in rot dargestellt, der globuläre Teil der N2-Domäne (N2') in blau, die Gelenkdomäne von N2 in grün und der Histidintag in grau. Pro161 und Pro213 sind als Stabmodell gelb dargestellt. **(b)** vergrößerter Ausschnitt der Pro161-Schleifenregion von Thr156 bis Tyr165. Die Abbildung basiert auf der Kristallstruktur von Holliger *et al.* (1999)[51] (PDBEintrag: 2G3P) und wurde mit dem Programm Pymol erstellt.

1.4 Prolylisomerisierung als langsamer Schritt in der Proteinfaltung

Die Peptidbindung besitzt einen partiellen Doppelbindungscharakter und zeigt daher eine planare Anordnung. Die flankierenden C_α-Atome sind deshalb entweder *trans* oder *cis* zueinander angeordnet. Für alle Peptidbindungen, außer denjenigen vor Prolin, ist der *cis*

Zustand aus sterischen Gründen sehr viel ungünstiger als der *trans* Zustand. Dementsprechend sind Nichtprolyl-*cis*-Peptidbindungen in gefalteten Proteinen äußerst selten [55; 56; 57; 58; 59]. Im Gegensatz dazu kommen Peptidbindungen vor Prolin häufig in einer *cis*-Konformation vor, da der energetische Unterschied zwischen den beiden Isomeren nur gering ist. Aufgrund des Pyrrolidinrings des Prolins steht das C^{α}-Atom der Aminosäure vor Prolin immer *cis*-ständig zu einem C-Atom (C_{α} oder C_{δ}) des Prolins (Abb. 1-2). Das *trans*-Isomer ist gegenüber dem *cis*-Isomer daher nur leicht energetisch favorisiert. In kurzen Peptiden treten beide Formen auf, wobei der *cis*-Anteil bei etwa 10 bis 30 % liegt, abhängig von der Aminosäure vor dem Prolin [61; 62; 63]. Da die Aktivierungsenergie E_A für die Drehung um die Peptidbindung aufgrund des partiellen Doppelbindungscharakters sehr hoch ist (etwa 80 kJ mol^{-1}), sind Prolyl *cis*/*trans*-Isomerisierungen ausgesprochen langsame Reaktionen [64; 65].

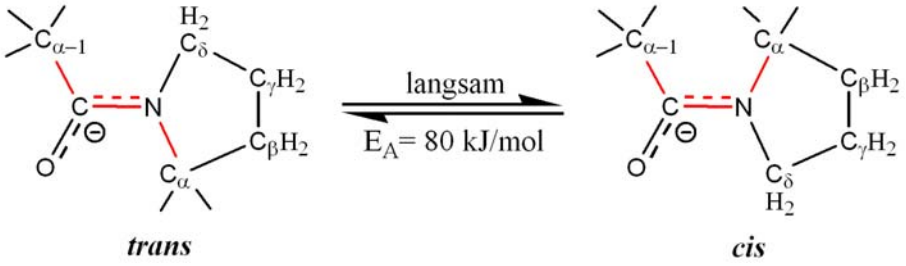


Abbildung 1-2: Isomerisierung zwischen einer *cis*- und einer *trans*-Form einer Peptidbindung (nach Eckert & Schmid)[60].

In gefalteten Proteinen wird der Konformationszustand der Prolylpeptidbindungen normalerweise durch die native Struktur festgelegt. Er ist also in allen Molekülen gleich, entweder *cis* oder *trans*. Tatsächlich sind etwa 5-7 % aller Peptidylprolylprolinbindungen in gefalteten Proteinen in *cis* [58]. Von dieser Regel gibt es nur wenige Ausnahmen. Mit hochauflösender NMR-Spektroskopie konnte in einigen Proteinen im nativen Zustand ein Gleichgewicht zwischen *cis*- und *trans*-Isomeren beobachtet werden, z.B. in Staphylokokkennuklease [66], Calbindin [67] oder humanem Interleukin-3 [68]. Eine mögliche Ursache hierfür ist, dass die entsprechenden Proline in flexiblen oder lokal entfalteten Strukturbereichen liegen. Der isomere Zustand ist hierbei nicht eindeutig fixiert, da die Kopplung mit der konformationellen Faltung zu schwach ist. Eine solche Heterogenität könnte aber auch auf eine funktionell wichtige Isomerisierung hindeuten, die z.B. als langsamer molekularer Schalter fungiert und möglicherweise durch Prolylisomerasen in der Schalterfrequenz gesteuert wird. Allerdings gibt es bislang keine auf molekularer Ebene charakterisierten Prolinschalter.

Ein derartiger Schalter wird für die SH2-Domäne der Tyrosinkinase Itk diskutiert. Diese Domäne liegt sowohl in der *cis*- als auch der *trans*-Form an Pro287 vor, die sich beide in ihrer Affinität für phosphotyrosinhaltige Peptide unterscheiden [69; 70; 71]. Kürzlich wurde eine ähnliche Heterogenität für das CrkII-Protein vorgeschlagen [72; 73]. Hier fungiert die *cis/trans*-Isomerisierung der Gly237-Pro238-Peptidbindung möglicherweise als lokaler Schalter, der die intramolekulare Domänenwechselwirkung beeinflusst.

1.5. Prolylisomerasen

Weil sie intrinsisch langsam sind, bestimmen Prolylisomerisierungen die Geschwindigkeit vieler Proteinfaltungsreaktionen [73; 74; 75; 76]. Für alle langsamen Prozesse im Metabolismus der Zellen existieren Enzyme, die diese Reaktionen katalysieren, so wie in diesem Fall die Prolylisomerasen (PPIasen). Diese Enzyme sind für die Zellen wichtig, da Faltungsintermediate mit falschen Prolinisomeren häufig anfällig für Aggregation, oder Amyloidbildung sind [74; 77]. PPIasen werden in drei unabhängige Familien unterteilt, die keine Sequenzhomologie zueinander zeigen: Cyclophiline, Parvuline und FK506-Bindungsproteine (FKBP) [78; 79; 80; 81].

In Bakterien und Archaeen scheinen Prolylisomerasen hauptsächlich in die Reifung von Proteinen involviert zu sein, da sie zumeist als isolierte Proteine, oder in Verbindung mit Chaperondomänen vorkommen. Chaperondomänen binden reversibel an exponierte hydrophobe Oberflächen entfalteter, oder partiell gefalteter Proteine und unterdrücken so die Aggregation [82]. Dabei erkennen sie ihre „Substrate“ an Oberflächenelementen, die charakteristisch für nichtnative Proteine sind [83]. Chaperone sind für das Krisenmanagement der Zelle wichtig, und werden vor allem in Stresssituationen, wie beispielsweise Hitze, oxidativem oder chemischem Stress verstärkt exprimiert. Deshalb wurden viele Chaperone zunächst als Hitzeschockproteine (Hsp) entdeckt.

In Eukaryonten sind Prolylisomerasen ebenfalls weit verbreitet. Sie kommen als einzelne Proteine und als Domänen größerer Proteine vor, sind jedoch seltener an Chaperondomänen fusioniert. Stattdessen besitzen sie häufig eine TPR-Domäne, wodurch sie spezifisch an Hsp90 und Hsp110 binden und dadurch nicht mehr selbstständig, sondern als „assoziierte Faktoren“ in großen Proteinreifungskomplexen tätig sind [84; 85; 86]. Darüber hinaus sind PPIasen in Eukaryonten verstärkt in die Regulation und Signaltransduktion involviert [87; 88].

Pin1 beispielsweise ist entscheidend für den korrekten Ablauf des Zellzyklus[89; 90]. In diesem Protein ist eine PPIase-Domäne vom Parvulin-Typ an eine WW-Domäne fusioniert,

wodurch spezifisch phosphorylierte Substrate erkannt und deren Isomerisierung katalysiert wird. Des Weiteren besitzen viele PPIasen auch Calbindin-, RNA-, oder Spliceosom-Bindedomänen, und sind so in eine Vielzahl verschiedener Prozessen involviert [88; 91; 92; 93; 94].

Die Substratspezifität von PPIasen wurde bisher nur mit Hilfe von Tetrapeptiden untersucht. Dabei zeigte sich, dass Cyclophiline eine sehr hohe Aktivität und eine breite Substratspezifität besitzen. Im Gegensatz dazu zeigten Mitglieder der FKBP-Familie für die Position vor Prolin eine ausgeprägte Präferenz für hydrophobe Aminosäuren, wie Leucin oder Phenylalanin [64; 65; 95], wohingegen die Katalyse in Gegenwart negativ geladener Reste vor Prolin 1000-fach schlechter ist [95]. Parvuline bevorzugen ebenfalls hydrophobe Aminosäuren vor Prolin und ähneln damit hinsichtlich der Substratspezifität den FKBPs [96; 97].

In *E. coli* sind die Prolylisomerasen Triggerfaktor und SlyD im Cytosol bzw. SurA und FkpA im Periplasma besonders wichtig für die Reifung von Proteinen. Alle vier Proteine besitzen zusätzlich zu ihrer PPIase-Domäne noch eine Chaperondomäne [98; 99; 100; 101]. Chaperone haben, wie bereits erwähnt, eine hohe Affinität für denaturierte oder teilgefaltete Proteine, wodurch zusätzlich Substrate für die Prolylisomerisierung gebunden werden können [102; 103]. Das Zusammenspiel von Chaperon- und PPIase-Domäne bei diesen faltungsaktiven Prolylisomerasen ist aber noch weitgehend unverstanden [104].

1.6 Filamentöse Phagen

Das Gen-3-Protein des filamentösen Phagen fd war im Rahmen meiner Dissertation von zentraler Bedeutung. Filamentöse Bakteriophagen sind sehr weit verbreitet und infizieren sowohl Gram-negative als auch Gram-positive Bakterienzellen [105]. Die Ff-Phagen (M13, f1, fd) sind dabei am besten untersucht. Sie besitzen eine Länge von 1 bis 2 µm und einen Durchmesser von ca. 7 nm [106] und infizieren spezifisch F-Pilus tragende (F^+) *E. coli*-Stämme. Filamentöse Phagen bestehen aus einer zirkulären einzelsträngigen DNA (ssDNA) von ca. 6500 Nukleotiden, die für elf verschiedene Proteine kodiert, und einer Proteinhülle, die aus etwa 2700 Kopien des Gen-8-Proteins (G8P) gebildet wird. Filamentöse Phagen sind resistent gegenüber den meisten Proteasen sowie stabil bei hoher Salzkonzentration und hohen Temperaturen [106].

Für die Infektion von *E. coli* ist das Gen-3-Protein (G3P) des Phagen verantwortlich, das in drei bis fünf Kopien an einem Ende des Phagenpartikels lokalisiert ist. G3P besteht aus drei Domänen, N1 (68 Aminosäuren), N2 (131 Aminosäuren) und CT (150 Aminosäuren), die durch glycinreiche Linker von 18 bzw. 39 Aminosäuren verknüpft sind [107, 108, 109]. Die CT-Domäne verankert das Protein in der Phagenhülle [110, 111]. Die N-terminalen Domänen N1 und

N2 stehen von der Phagenoberfläche ab, und werden für die sequentielle Interaktion mit den beiden Rezeptoren an der *E. coli*-Zelle verwendet [112]. Im Ruhezustand des Phagen liegt G3P in einer geschlossenen Form vor, in der alle drei Domänen des G3P miteinander interagieren. In diesem Zustand ist G3P robust und stabil, der Phage jedoch nicht infektiös. Dieser intramolekulare Komplex muss daher im Verlauf der Infektion gelöst werden (Abb. 1-3) [113]. Im ersten Schritt der Infektion bindet die N2-Domäne an die Spitze des bakteriellen F-Pilus, den primären Rezeptor [114; 115]. Durch eine anschließende Konformationsänderung wird die Bindungsstelle für den sekundären TolA zugänglich, und G3P aktiviert. Die TolA-Bindungsstelle ist auf der N1-Domäne im Bereich der Interaktionsfläche zwischen N1 und N2 lokalisiert, so dass sie in der geschlossenen Ruheform des Phagen mit fest assoziierten Domänen unzugänglich ist [51; 54]. Während des Infektionsprozesses muss also die Interaktion zwischen N1 und N2 aufgebrochen werden und die bindungsfähige Form von G3P muss erhalten bleiben, bis die N1-Domäne TolA erreicht [112].

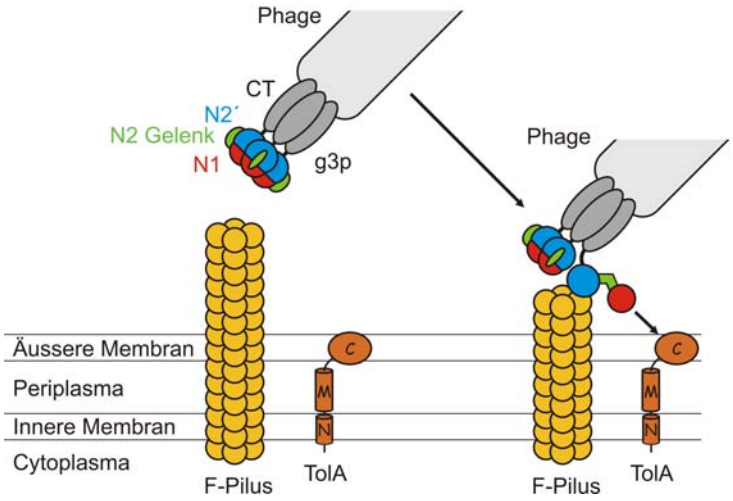


Abbildung 1-3: Die Rolle von G3P bei der Infektion von *E. coli* durch den Phagen fd. Gezeigt sind die ersten Schritte der Infektion: die Bindung der N2-Domäne an die Spitze des F-Pilus löst die Interaktion zwischen N1 und N2. Die N1-Domäne kann dann an die C-terminale Domäne des Rezeptors TolA binden (entnommen aus Eckert & Schmid) [60].

Interessanterweise entspricht diese biologisch aktive Form von G3P mit den lose assoziierten Domänen dem Intermediat, das in *in-vitro* Faltungsstudien identifiziert wurde (s. 1.3). Die feste Assoziationsreaktion, d.h. der letzte Faltungsschritt, muss revertiert werden, um den Phagen für die Infektionsprozess zu aktivieren. Die entsprechende *cis*→*trans*-Isomerisierung an Pro213 ist deshalb wichtig, um die Bindungsstelle für TolA zu exponieren. Gleichzeitig dient *trans*-Pro213 als ein Zeitgeber, der gewährleistet, dass G3P so lange geöffnet und partiell entfaltet bleibt, bis die N1-Domäne ihren Interaktionspartner TolA gefunden hat [116].

Die kovalente Verbindung zwischen den Domänen von G3P ist essentiell für die Phageninfektiosität [117], und diese Vorraussetzung wird in der *Proside* (*protein stability*

increased by directed evolution) Technik für evolutionäre Proteinstabilisierung genutzt [11; 118; 119]. In *Proside* wird ein Repertoire an Sequenzen, die für das zu stabilisierende Protein kodieren, zwischen die N2 und die C-terminale Domäne in das Gen-3-Protein eingefügt. Die daraus resultierende Bibliothek an rekombinanten Phagen wird einem proteolytischen Verdau unterzogen. Die physikalische Verbindung zwischen den beiden Domänen von G3P bleibt nur erhalten, wenn die eingebaute Variante des Gastproteins stabil gefaltet und damit unzugänglich für Proteasen ist. Phagen mit stabilen Gastproteinen werden so in mehreren Zyklen von *in vitro* Proteolyse der Phagenbibliotheken, Infektion von *E. coli*-Zellen und Phagenpropagation angereichert [118; 120]

Neben den Ff-Phagen gibt es noch zwei weitere Typen von *E. coli* infizierenden filamentöse Phagen: IF1 [121; 122] und IKe [123; 124]. Diese infizieren spezifisch I-Pili bzw. N-Pili tragende *E. coli*-Zellen [107; 125; 126]. Die Infektion wird dabei ebenfalls über ein G3P vermittelt [127; 128], jedoch wurden diese Infektionsvorgänge bisher kaum untersucht. Allen *E. coli*-infizierenden Phagen gemeinsam ist eine konservierte TolA-III bindende Domäne. Die pilusbindenden Domänen dagegen besitzen untereinander keine Ähnlichkeit und sind vermutlich nicht miteinander verwandt [123; 129].

1.7 Problemstellung

Durch gerichtete Evolution mittels der *Proside*-Selektionsmethode (*protein stability increased by directed evolution*) konnten von Insa Kather in ihrer Doktorarbeit stark stabilisierte Varianten des Gen-3-Proteins und der TEM-1 β-Lactamase erhalten werden. Im Rahmen der hier vorgelegten Dissertation wurden die Kristallstrukturen dieser Proteine gelöst, um durch den Vergleich mit der Struktur der entsprechenden Wildtypproteine zu verstehen, wie die selektierten Mutationen stabilisierend wirken.

Das N1N2-Fragment von G3P (G3P*, s. Abb. 1-1a) des Phagen fd ist ein sehr gutes Modellprotein um die Prinzipien von Domänenassoziation und Domänenassemblierung zu untersuchen. Dafür wurden bisher ausschließlich die isolierte N1-Domäne (As1-70) und das N1N2-Konstrukt (As1-220) untersucht. Im Rahmen dieser Arbeit wurde deshalb die nur wenig stabile, isolierte globuläre N2-Domäne (As102-205) exprimiert, ihre Stabilität und ihre Faltungskinetik charakterisiert und anschließend mit den entsprechenden Eigenschaften des Gesamtproteins verglichen. Dabei zeigte sich überraschenderweise, dass in der nativen N2-Domäne ein *cis*/*trans*-Gleichgewicht an Pro161 existiert, das entscheidend für die Stabilität und Faltung des Proteins ist. Um dieses *cis*/*trans*-Gleichgewicht zu verändern, muss Energie aus dem Bereich der gefalteten N2’ zu Pro161 transferiert werden. Die N2-Domäne ist

dadurch ein hervorragendes Modell, um die energetische Kopplung von konformationeller Faltung und Prolylisomerisierung zu untersuchen. Um diese Kopplung auf molekularer Ebene zu verstehen, wurde die lokale Sequenzumgebung um Pro161 systematisch variiert, sowohl im Schleifenbereich als auch in den beiden Faltblattsträngen, die in diese Schleife einmünden. Die Ergebnisse dieser thermodynamischen und kinetischen Analyse deuteten darauf hin, dass die Pro161-Schleife entscheidend für den Übergangszustand der Faltung der N2-Domäne ist. Ein weiteres Ziel dieser Arbeit war deshalb, den Beitrag individueller Bereiche von N2 zur Struktur und zur Energetik des Übergangszustands der Faltung zu untersuchen. Hierfür wurden zusätzliche 45 N2-Varianten mit Aminosäureaustauschen verteilt über das gesamte Protein hergestellt, und deren Entfaltungs- und Rückfaltungskinetik im Rahmen einer Φ-Wert-Analyse charakterisiert.

Pro161 befindet sich in einer ungewöhnlich großen, irregulären Schleife von N2 (s. Abb. 1-1b). Die β-Faltblattstränge, links und rechts davon, optimieren ihre stabilisierenden Interaktionen wenn Pro161 in *cis* ist, und diese Interaktionen wiederum stabilisieren im Sinne einer thermodynamischen Kopplung die *cis*-Form an Pro161. Durch Verkürzen der Schleife zu klassischen β-Turns und rationalem Design basierend auf Konsensussequenzanalyse wurde versucht diesen Bereich optimal zu ersetzen, und so die stabilisierenden Interaktionen der Faltblattstränge zu erhalten bzw. zu verbessern. Parallel dazu wurde die Sequenz von zwei weiteren Turns in N2 durch statistische Analyse optimiert, und dadurch die generelle Anwendbarkeit von rationalem Turn-Design zur Stabilisierung von Proteinen untersucht.

Die Sequenzspezifität von Prolylisomerasen konnte bislang lediglich mit Hilfe von kurzen, chemisch modifizierten Tetrapeptiden untersucht werden. Die Spezifität in Bezug auf Prolylbindungen in rückfaltenden Proteinketten ist unbekannt, da bisher ein geeignetes Testsystem fehlte. Wie schon ausgeführt, lässt sich die *trans*→*cis*-Isomerisierung an Pro161 in N2 gut durch Prolylisomerasen katalysieren. Deshalb wurde eine Bibliothek von N2-Varianten mit allen natürlichen Aminosäuren vor Pro161 hergestellt, und damit die Substratspezifität von PPIasen bezüglich einer Proteinfaltungsreaktion untersucht. Dabei war insbesondere von Interesse wie eine zusätzliche Chaperondomäne, wie etwa in SlyD, die Substrataktivität und -spezifität beeinflusst.

Im G3P* des Phagen fd tritt während der *in-vitro* Faltung ein langlebiges Intermediat mit einem *trans*-Prolin an Position 213 auf, das für den Infektionsprozess relevant ist. Mittels NMR-Spektroskopie wurde die Struktur und Stabilität dieses Faltungsintermediates von G3P* in Zusammenarbeit mit Ulrich Weininger und Jochen Balbach (Universität Halle-Wittenberg) charakterisiert und der molekulare Weg der Domänenassemblierung analysiert.

Der filamentöse Phage IKe ist nahe verwandt zu den Ff-Phagen. Das G3P des Phagen IKe ist für die Infektion von N-Pili tragenden *E. coli*-Zellen verantwortlich. Am Beispiel dieses Zweidomänenproteins sollte untersucht werden, ob alle Gen-3-Proteine filamentöser Phagen einen ähnlichen Faltungs- und damit Infektionsmechanismus besitzen. Dafür wurden die beiden isolierten N-terminalen Domänen und das N1N2-Konstrukt von IKe-G3P* strukturell, thermodynamisch und kinetisch charakterisiert.

2 Zusammenfassung und Diskussion der Ergebnisse

2.1 Strukturelle Basis der Beiträge einzelner Aminosäureaustausche zur Stabilisierung einer disulfidfreien Variante des Gen-3-Proteins

Die beiden N-terminalen Domänen N1 und N2 des G3P des Phagen fd enthalten drei Disulfidbrücken, eine in N2 (Cys188-Cys201) und zwei in N1 (Cys7-Cys36 und Cys46-Cys53). Diese sind essentiell für die konformationelle Stabilität des Proteins, es entfaltet, wenn sie durch Reduktion oder Mutation entfernt werden. Im Rahmen von *Proside*-Selektionen können diese Cysteine zu Problemen führen, da sie nichtproduktive Disulfidbrücken mit Gastproteinen eingehen können, wodurch die Stabilität und Infektiosität der Phagen deutlich reduziert wird [130]. Deshalb wurden die Disulfidbrücken mittels *in-vitro*-Selektion durch optimale nicht Cystein-Aminosäurepaare ersetzt [131]. Dadurch konnte aber nur ein geringer Teil des enormen Stabilitätsverlusts kompensiert werden, der durch den Bruch der kovalenten Verknüpfungen verursacht wurde. Die Generierung stabiler, disulfidfreier Varianten von G3P wurde durch die Selektion zusätzlicher *second-site*-Mutationen möglich. Die beste Variante ohne Disulfidbrücken die von Insa Kather erhalten wurde, zeigte eine höhere konformationelle Stabilität als das ursprüngliche dreifach disulfidverbrückte Wildtypprotein.

Für das Verständnis von stabilisierenden Mutationen sind Kristallstrukturen unabdingbar. Deshalb wurde im Rahmen von Teilarbeit A die disulfidfreie Gen-3-Protein-Variante mit allen stabilisierenden Mutationen (0SS-G3P* stab) kristallisiert und deren Struktur mit einer Auflösung von 1.9 Å gelöst. Die in den Kristallisationsbedingungen enthaltenen NH_4^+- und Ca^{2+}-Ionen waren dabei essentiell für eine hohe Streuleistung der Kristalle.

Die Interpretation der Kristallstruktur von 0SS-G3P* stab ergab folgende wichtige Kernaussagen: Die molekulare Ursache für die Stabilisierung der meisten Substitutionen kann auf eine optimierte Packung zurückgeführt werden (Abb. 10, Teilarbeit A). Die Packung von Seitenketten ist ein wichtiger Faktor für thermodynamische Stabilität [132], insbesondere in Proteinen aus thermophilen Organismen [133; 134]. Die Ergebnisse sind sehr ähnlich zu Erkenntnissen aus den Strukturen stabilisierter Varianten des Gb1-Proteins, wo durch Aminosäuresubstitutionen zumeist Packungsdefekte eliminiert wurden [135; 136]. Die Substitution R29W in der N1-Domäne von 0SS-G3P* stab hat mit ungefähr 25 kJ mol^{-1} den größten Anteil. Dieser Wert entspricht der Gesamtstabilität vieler kleiner Proteine. Diese Schlüsselmutation stabilisiert auf drei Arten: (i) Es wird eine geladenen Aminosäure ersetzt, die eine vergrabene Salzbrücke im Wildtypprotein asubildet. Verborgene Salzbrücken in

Proteinen sind ungünstig, da beide interagierenden Seitengruppen vollständig desolvatisiert werden müssen. (ii) Die lokale Packung der Reste, die die Disulfidbrücke C46-C53 ersetzen wird optimiert und (iii) die Domäneninteraktionen werden verbessert (Abb. 9, Teilarbeit A). Offensichtlich gibt es Aminosäurereste in Proteinen, die eine Schlüsselrolle für die Stabilität haben. Die Auswirkungen der weiteren stabilisierenden Substitutionen sind in Teilarbeit A diskutiert. Die Vielzahl an gefundenen stabilisierenden Mutationen zeigt, dass es verschiedene Wege gibt ein Protein zu stabilisieren.

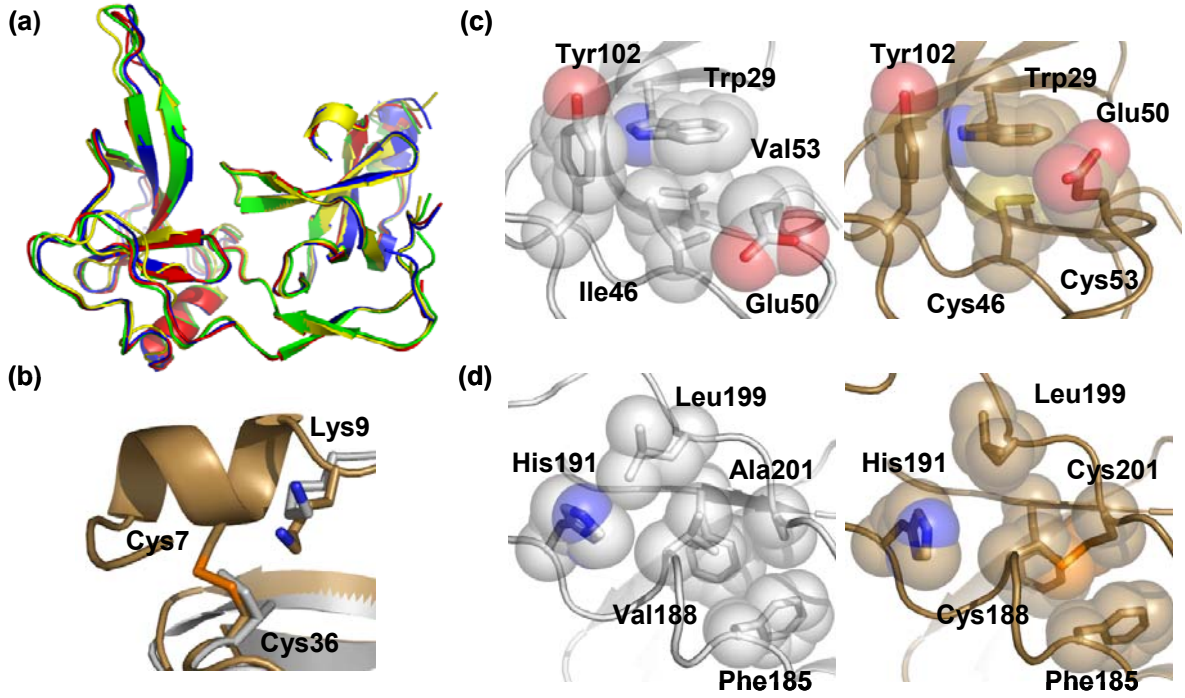


Abbildung 2-1: **(a)** Überlagerung der G3P*-Moleküle in der asymmetrischen Einheit: Molekül A (blau), Molekül B (rot), Molekül C (grün), Molekül D (gelb); Vergleich der lokalen Struktur von 0SS-G3P* stab (linke Abbildung, grau) und 3SS-G3P* stab (rechte Abbildung, braun) in der Umgebung **(b)** der Disulfidbrücke C7-C36, **(c)** der Disulfidbrücke C46-C53, **(d)** der Disulfidbrücke C188-C201.

Um aufzuklären, ob die selektierten *second-site*-Mutationen im G3P* mit intakten Disulfidbrücken ebenfalls stabilisieren, wurden in das stabilisierte disulfidfreie G3P* die ursprünglich vorhandenen drei Disulfidbrücken wieder eingeführt. Diese G3P*-Variante (3SS-G3P* stab) wurde durch denaturierungsmittelinduzierte und thermische Entfaltung von Insa Kather ausführlich charakterisiert. Das Protein zeigte eine enorme konformationelle Stabilität und konnte durch das starke Denaturierungsmittel GdmCl nicht mehr vollständig entfaltet werden. Die identifizierten Mutationen wirken also auch im Hintergrund der Disulfidbrücken sehr stark stabilisierend [131].

Diese Variante (3SS-G3P* stab) wurde ebenfalls röntgenkristallographisch untersucht (2.0 Å Auflösung), um die Grundlage ihrer extremen Stabilisierung zu verstehen. Das Protein

kristallisierte unter ähnlichen Bedingungen wie die ursprüngliche, nichtdisulfidverbrückte G3P*-Variante. Es befinden sich vier G3P*-Moleküle in der asymmetrischen Einheit, die sich nur marginal unterscheiden (Abb. 2-1a). Die einzigen Veränderungen in den beiden Kristallstrukturen betreffen die Regionen um die drei Disulfidbrücken.

Der aminoterminale Bereich (Aminosäure 1-9) ist in 0SS-G3P* stab nicht aufgelöst da er lokal entfaltet ist (Abb. 2-1b). In 3SS-G3P* stab bilden die ersten Aminosäuren aufgrund der kovalenten Verknüpfung von Cys7 und Cys36 eine kurze α-Helix, die identisch zu der im Wildtypprotein ist (Abb. 2-1b). Der Bereich um die Disulfidbrücke Cys46-Cys53 zeigt zwischen beiden Proteinen nur kleine Reorientierungen der strukturell benachbarten Seitengruppen, wie etwa Trp29 und Glu50 (Abb. 2-1c). Auch in der Region um Cys188-Cys201 sind nur geringe Packungsunterschiede zu erkennen. Durch die wiedereingeführte Disulfidbrücke verändern sich die Seitenketten von Phe185, His191, Leu199 leicht, und gehen in eine Position die sehr ähnlich zu der im Wildtypprotein ist (Abb. 2-1d). Diese Strukturanalyse wird zur Publikation vorbereitet.

2.2 Strukturelle Interpretation einzelner stabilisierender Mutationen in der β-Lactamase

Das im Labor etablierte *in vitro*-Selektionsystem auf erhöhte Stabilität (*Proside*) wurde bisher nur für das Gen-3-Protein oder kleine Eindomänenproteine wie CspB [11; 137] und Gb1 [135; 138] erfolgreich angewendet. Die TEM-1 β-Lactamase aus *E. coli* war das erste größere Protein (263 Aminosäuren) das durch *Proside* stabilisiert werden konnte. β-Lactamasen sind monomere Enzyme die β-Lactam Antibiotika spalten [139; 140; 141]. Ihre Struktur, Stabilität und Faltungskinetik sind sehr gut charakterisiert, was für die Analyse und das Verständnis von stabilisierend wirkenden Substitutionen sehr wertvoll ist [142; 143; 144; 145; 146; 147; 148; 149]. Mit Hilfe von *Proside* wurden neun stabilisierende Aminosäureaustausche identifiziert, deren thermodynamische und kinetische Analyse bereits detailliert in der Dissertation von Insa Kather beschrieben sind [131]. Im Rahmen der vorliegenden Arbeit wurde die stabilste β-Lactamase-Variante kristallisiert und ihre Kristallstruktur (2.0 Å Auflösung) gelöst.

Die meisten Aminosäureaustausche erhöhen die konformationelle Stabilität durch verbesserte Seitenkettenpackung von TEM-1 β-Lactamase, wie im Falle von Ile208Met, Ala224Val (vergleiche Abb. 6 in Teilarbeit B). Drei der vier am stärksten stabilisierenden Mutationen befinden sich am oder in der Nähe des N-Terminus von Helices. Die Substitution Met182Thr stabilisiert, in dem sie das N-Cap der α-Helix 183-195 bildet. Die Mutation Leu201Pro stabilisiert, da sie optimal für die erste Position einer Helix ist und Glu147Gly

befindet sich zwischen zwei Helices und optimiert deren Packung (s. Abb. 7, Teilarbeit B). Diese drei Mutationen zeigen möglicherweise einen generellen Weg, um Protein zu stabilisieren. Durch Konsensussequenzanalyse sind die Präferenzen von bestimmten Sekundärstrukturbereichen wie α-Helices [21; 25; 26; 27] und β-Turns sehr gut bekannt [22; 23]. Positionen, in denen Proteine davon abweichen, können anhand von Kristallstrukturen leicht identifiziert werden, und die Analyse der lokalen Struktur kann Aufschluss über zusätzliche unvorteilhafte Wechselwirkungen liefern. Ist die Kristallstruktur eines Proteins bekannt, so sollte es durch Substitution dieser Aminosäuren hin zu den statistisch bevorzugten Resten möglich sein, jedes Protein zu stabilisieren (s. 2.6).

Insgesamt zeigte die Charakterisierung und Kristallstrukturanalyse der durch *in-vitro* Selektion stabilisierten Varianten von G3P* und der TEM 1 β-Lactamase, dass Selektionen generell am schwächsten Punkt eines Proteins ansetzen. Bei dem bisher untersuchten Kälteschockprotein CspB wurde durch die Eliminierung ungünstiger Ladungswechsel-wirkungen eine Stabilisierung erzielt [11; 119; 137]. Bei der β-Domäne des Streptokokken-Proteins G (Gb1) führte die Eliminierung von Packungsdefekten zur Stabilisierung [135; 138]. Die Analyse größerer Proteine in dieser Arbeit zeigt für G3P* ebenfalls eine optimierte Packung und Domäneninteraktion als Stabilisierungsprinzip. Bei der β-Lactamase führte die Optimierung von Helices, speziell an deren Aminoterminus und die verbesserte Packung von Helices zur Stabilisierung. Die Kombination aus thermodynamischer Stabilitätsanalyse und Kristallstruk-turanalyse war dabei sehr hilfreich für die molekulare Interpretation stabilisierend wirkender Mutationen, es zeigte sich aber auch deutlich, dass die Stabilitätsbeiträge einzelner Reste und ihrer Interaktionen nicht allein aus der Struktur ablesbar sind.

2.3 Stabilität und Faltungsmechanismus der isolierten N2-Domäne des Gen-3-Proteins des Phagen fd

Bislang wurde vor allem die Faltung von Eindomänenproteinen untersucht. Diese zeichnen sich häufig durch sehr schnelle Faltungsreaktionen nach einem einfachen Zwei-Zustandsmodell aus [36]. Die meisten Proteine sind allerdings aus mehreren Domänen aufgebaut, welche strukturelle, funktionale oder separat faltende Einheiten bilden [150; 151; 152]. Die Faltung von Mehrdomänenproteinen ist ein wesentlich komplexerer Vorgang als die der Eindomänenproteine und wurde bisher kaum untersucht [49; 50]. Das N1N2 Zweidomänenfragment von G3P (G3P*) des Phagen fd wird als Modellprotein verwendet, um die Faltungsmechanismen von Mehrdomänenproteinen zu verstehen [52; 53]. Dabei konnte nur

die Faltung der isolierten N1-Domäne und des Gesamtproteins untersucht werden, da die Expression der isolierten N2-Domäne bislang gescheitert war.

In dieser Arbeit wurde die Expression und Reinigung der isolierten N2-Domäne (Aminosäuren 102-205) sukzessive optimiert, sodass jetzt Ausbeuten von bis zu 20 mg/l Fermentation erzielt werden (Material und Methoden, Teilarbeit C und D,). Es wurde zunächst die Stabilität und den Faltungsweg der isolierten N2-Domäne charakterisiert, und dann ihre Faltung im Gesamtprotein untersucht (s. 2.7).

Die isolierte, globuläre N2-Domäne besitzt nur eine marginale Stabilität und neigt zur Aggregation. Die Substitution Q129H stammt aus der *Proside*-Selektion des G3P* (s. 2.1) und stabilisiert um 4 kJ mol^{-1}. Diese Q129H-Variante von N2 hat immer noch einen niedrigen thermischen Übergangsmittelpunkt T_M von 38.4 °C. Dieses Protein wird im Folgenden als N2' bezeichnet. N2' besitzt ein Tryptophan (Trp181), dessen Fluoreszenz bei Denaturierung in Harnstoff um 40 % abnimmt und die entsprechende Gleichgewichtsentfaltung ergab einen Übergangsmittelpunkt von 2.7 M Harnstoff. Die extrapolierten Enthalpien der Entfaltung ΔG_D aus thermischer und denaturierungsmittelinduzierter Entfaltung von N2' stimmen sehr gut überein, was andeutet, dass die Gleichgewichtsentfaltung adäquat durch einen Zwei-Zustandmodell beschrieben wird.

Die Faltung von N2' hingegen ist ein komplexer Vorgang, wobei in der Entfaltung zwei und in der Rückfaltung drei Phasen (schnell, intermediär, langsam) beobachtet werden (Abb. 3, Teilarbeit C). Die schnelle und die intermediäre Rückfaltungsreaktion folgen einem gemeinsamen Entfaltungsast (Abb. 2-2a). In der Rückfaltung unterscheiden sich beide Phasen in ihrer Rate ungefähr um den Faktor 30, ihre Denaturierungsmittelabhängigkeit ist jedoch ähnlich. Interessanterweise folgt keine der Reaktionen dem Gleichgewichtsübergang. Die Übergangsmittelpunkte von 3,8 und 1,8 M Harnstoff stimmen nicht mit denen aus der Gleichgewichtsentfaltung (2,7 M Harnstoff) überein.

Der Grund hierfür ist die zusätzliche langsame Faltungsreaktion. Deren Rate nimmt mit steigender Harnstoffkonzentration leicht ab und erreicht ein Minimum bei ungefähr 2,7 M Harnstoff, bevor sie wieder zunimmt (s. Abb. 2-2a). Diese langsame Reaktion wird in Gegenwart von Prolylisomerasen, wie SlyD und Cyclophilin, stark beschleunigt (s. Abb. 3c, Teilarbeit C), was darauf hindeutet, dass ihre Rate durch eine Prolylisomerisierung limitiert wird. Um zu überprüfen ob diese Isomerisierung tatsächlich durch Pro161 verursacht wird (dem einzigen *cis*-Prolin in N2), wurde die Faltung der P161A-Variante untersucht. Diese zeigte eine deutlich vereinfachte Faltung. Es war nur noch eine konformationelle Entfaltungs- und Rückfaltungsphase vorhanden (s. Abb. 4e, Teilarbeit C), deren Minimum in der Chevron-

Auftragung sich mit dem des Gleichgewichtsübergangs deckt (s. Abb. 2c, Teilarbeit C). Die Ergebnisse zeigen, dass nur eine Aminosäure, Pro161, für die komplexe Faltung des Wildtypproteins verantwortlich ist.

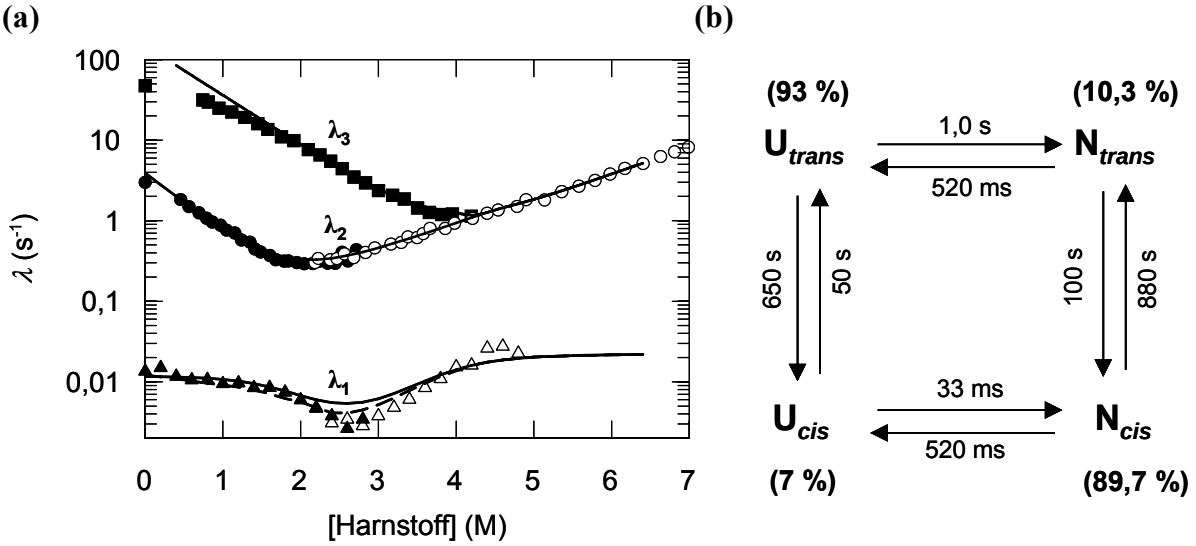


Abbildung 2-2: (a) Entfaltungs- (offene Symbole) und Rückfaltungskinetiken (ausgefüllte Symbole) von N2'. Die durchzogenen Linien entsprechen den simulierten Daten anhand des Boxmodells (rechte Abbildung); die gestrichelte Linie entspricht den Daten für λ_1 unter Berücksichtigung einer leichten Harnstoffabhängigkeit. (b) Kinetisches Modell der Faltung und Prolylisomerisierung in N2'. Die *cis* und *trans* Anteile in entfalteter und nativer N2' sind Teilarbeit D entnommen und in Klammen angegeben. Die Werte über und unter den Pfeilen entsprechen den reziproken Ratenkonstanten. Für die $U_{cis} \leftrightarrow U_{trans}$ Reaktion sind die Werte für die Entfaltung in 5,0 M Harnstoff angegebenen; alle anderen Werte beziehen sich auf die Rückfaltung in 1,0 M Harnstoff.

2.3.1 Untersuchungen zur Prolylisomerisierung in der nativen und denaturierten N2-Domäne

Um den Faltungsmechanismus von N2' noch besser zu verstehen, wurde die *cis/trans*-Isomerisierung von Pro161 im nativen und im entfalteten Protein durch Doppelmischexperimente untersucht. Für die Charakterisierung der Isomerisierung in der denaturierten N2'-Domäne wird das native Protein für unterschiedlich lange Zeit entfaltet und anschließend rückgefaltet. Die Zunahme der Rückfaltungsamplitude ist dabei ein Maß für die Zahl der Moleküle, die nach einer bestimmten Entfaltungsdauer eine *cis*- oder eine *trans*-Konformation an Pro161 einnehmen. Die zeitabhängige Zu- oder Abnahme der Amplitude ergibt somit die Kinetik der Prolylisomerisierung im denaturierten Protein. Die Bildung der Moleküle, die auf dem prolinlimitierten Weg zurückfalten, sollte eine Verzögerungsphase aufweisen, da die Prolylisomerisierung im entfalteten Protein in einer sequentiellen Reaktion nach der konformationelle Entfaltung erfolgt. Diese Verzögerung sollte zu einer kinetischen Phase mit der Zeitkonstante der konformationellen Entfaltung (τ= 520 ms) und einer negativen Amplitude führen. Eine Reaktion mit dieser Zeitkonstante wird tatsächlich

beobachtet, aber ihre Amplitude ist positiv. Das zeigt, dass ein kleiner Teil der Moleküle schon im nativen Ausgangsprotein ein inkorrektes (*trans*) Prolin-Isomer aufweist. Deren konformationelle Entfaltung führt direkt zur Bildung der Moleküle die auf dem prolinlimitierten Weg zurückfalten und erklärt so das Fehlen der Verzögerungsphase (s. Abb. 5, Teilarbeit C).

Bei pH 2,0 erfolgt die konformationelle Entfaltung der N2' mit einer Zeitkonstante von <5 ms. Nach der kürzestmöglichen Entfaltungszeit des Doppelmischexperiments (20 ms) ist die N2-Domäne daher bereits vollständig entfaltet und aus den Amplituden der beobachteten Rückfaltung konnte direkt der Anteil von *cis*- und *trans*-Isomer an Pro161 im nativen Protein (89 %/11 %) bestimmt werden.

Die vergleichende kinetische Analyse mittels Einzel- und Doppelmischexperimenten zeigte, dass die Rückfaltung von N2' auf zwei parallelen Faltungswegen erfolgt. Um zu überprüfen ob die Faltung der *trans*-Spezies schon zu nativem Protein führt, wurde ein Zweischrittentfaltungstest für native Moleküle genutzt [76; 153]. 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 Bildungskinetik des vollständig gefalteten Proteins. Die entsprechenden Doppelmischexperimente ergaben, dass die N2-Moleküle mit Pro161 in der inkorrekten *trans*-Konformation, genauso langsam entfalten wie Moleküle mit einem *cis*-Pro161. Die Moleküle, die in der intermediären Phase λ_2 rückfalten, gehören also auch zum Ensemble der nativen Spezies. In diesem Experiment wurde ebenfalls die langsame Prolylisomerisierung beobachtet, da die Umwandlung der intermediären Form in das native Protein mit einer kleinen Fluoreszenzänderung einhergeht. Dies ist vermutlich auf die vier Tyrosine (Tyr151, Tyr165, Tyr166, Tyr168) zurückzuführen, die sich in den zur Pro161-Schleife führenden Faltblattsträngen befinden.

Die Einzel- und Doppelmischexperimente zeigten, dass Spezies mit dem *cis* und *trans* Isomer von Pro161 sowohl in der nativen, als auch in der denaturierten Form von N2' koexistieren, und dass die Ent- und Rückfaltung jeweils mit einer starken Verschiebung des *cis*/*trans*-Verhältnis an diesem Prolin einhergehen. Das Vier-Spezies Boxmodell in Abb. 2-2b bietet den einfachsten kinetischen Mechanismus um die Kopplung von konformationeller

Faltung und Prolylisomerisierung im gefalteten und entfalteten Zustand des Proteins zu beschreiben.

Wenn die bisher gezogen Schlüsse und Interpretationen der kinetischen Experimente richtig sind, so sollte der Faltungsmechanismus der N2 sehr gut durch ein Boxmodell beschrieben werden. Dieses besteht aus vier Reaktionen mit acht Ratenkonstanten (Abb. 2-2b). Die geschlossene Anordnung des Modells liefert eine Beziehung zwischen diesen Ratenkonstanten und reduziert die Anzahl an unabhängigen Variablen auf sieben. Jedoch sind nur sechs Parameter (drei λ-Werte und drei Amplituden) zugänglich und das System ist damit unterbestimmt. Die zusätzlichen benötigten Informationen stammen aus den Doppelmischexperimenten. Dadurch erhält man, sowohl für den N- als auch den U-Zustand die Gleichgewichtskonstanten für die *cis* $\leftrightarrow$ *trans* Reaktionen (die $U_{trans} \leftrightarrow U_{cis}$ und $N_{trans} \leftrightarrow N_{cis}$ Reaktionen in dem Box Modell, siehe Abb. 2-2b). Die Übereinstimmung der experimentellen und der simulierten Werte ist für λ_3 und λ_2 fast perfekt. Nur bei geringen Harnstoffkonzentrationen, weicht der log λ_3 leicht von der Linearität ab, was vermutlich auf ein vorgelagertes Intermediat unter diesen Bedingungen zurückzuführen ist (Abb. 2-2a). Die Übereinstimmung für λ_1 in den Basislinien ist ebenfalls fast perfekt, und kleine Abweichungen sind nur im Übergangsbereich bei 2,6 M Harnstoff zu erkennen. Das deutet darauf hin, dass die Rate der Pro161 *trans*$\rightarrow$*cis* Isomerisierung im nativen Protein leicht denaturierungsmittelabhängig ist. Berücksichtigt man eine leichte Verlangsamung der Rückfaltung, so wird tatsächlich die Übereinstimmung zwischen Simulation und Messdaten noch besser (Abb. 2-2b).

Mit Hilfe des Boxmodells wurde auch der Amplitudenverlauf der einzelnen Phasen und der Gleichgewichtsübergang berechnet. Die Simulationen stimmen ebenfalls hervorragend mit den experimentellen Daten überein (s. Abb. 8, Teilarbeit C) und bestätigen, dass das Box-Modell die Kopplung zwischen konformationeller Faltung und Prolylisomerisierung von N2' adäquat beschreibt.

Bisher wurde angenommen, dass Prolylpeptidbindungen in gefalteten Proteinen hoch geordnet sind, da der native Zustand nur mit einem *trans* oder *cis* Isomer kompatibel ist. Es sind nur sehr wenige Ausnahmen bekannt, in denen konformationelle Heterogenität von Prolinen mittels NMR Spektroskopie identifiziert wurde [67; 68; 69; 154; 155; 156; 157; 158; 159; 160; 161]. Die N2'-Domäne ist das erste Protein in dem die energetische Kopplung von Faltung und Prolylisomerisierung detailliert untersucht wurde. Die Ergebnisse demonstrierten, dass beide Spezies zum Ensemble des nativen Zustands gehören.

Während der Faltung wird der *cis*-Anteil von 7 % im denaturierten Protein auf 89 % im nativen Protein erhöht, was einem Energieaufwand von 10 kJ mol^{-1} entspricht. Obwohl Pro161 exponiert an der Spitze einer Schleife befindet, werden 50 % der insgesamt verfügbaren Konformationsenergie des nativen Proteins verwendet, um das *cis*/*trans*-Gleichgewicht an Pro161 zu verändern. Momentan gibt es noch keine Evidenz, dass die zwei Isomere unterschiedliche biologische Funktionen in N2' haben.

Dennoch gibt es keinen Zweifel, dass die Prolylisomerisierung über die Proteinfaltung hinaus wichtig ist [71; 93; 116; 161]. Alternative funktionelle Zustände eines Proteins können im Prinzip durch Prolylisomerisierung für eine definierte Zeit eingefroren werden, da die gegenseitige Umwandlung zwischen den Isomeren intrinsisch langsam ist. Die Kopplung zwischen konformationeller Faltung und Prolylisomerisierung bietet einen einfachen Weg um ankommende Signale zu propagieren, sei es ausgelöst durch Ligandenbindung oder durch kovalente Modifikation [66; 90; 162; 163; 164; 165; 166; 167]. Dadurch kann ein intrinsisch schnelles Signal in eine langsame Antwort ungewandelt werden. Erste experimentelle Evidenz für individuelle Komponenten solcher Signaltransduktionswege sind vorhanden [48; 69; 71; 93; 112; 116; 161; 168; 169; 170; 171].

2.4 Einfluss der lokalen Umgebung auf das *cis*/*trans*-Verhältnis an Pro161 in N2'

Die in 2.3 beschriebene Analyse des Faltungsmechanismus der N2-Domäne zeigte, dass dieser hervorragend durch ein Box-Model beschrieben ist, in dem die konformationelle Faltung und die Prolylisomerisierung an Pro161 miteinander koppeln. Aus dieser Analyse ergaben sich mehrere grundsätzliche Fragen: (i) Wie ist der native Zustand mit zwei Prolylisomeren kompatibel? (ii) Wie ist die lokale Konformation an den *cis*/*trans*-Isomerisierungszustand von Pro161 gekoppelt? (iii) Und wie wird während der Faltung konformationelle Energie akkumuliert und in der gefalteten N2' an die Spitze der Schleife nach Pro161 transferiert?

In Abwesenheit einer gefalteten Struktur, wie in kurzen Peptiden, wird das *cis*/*trans*-Gleichgewicht und die Isomerisierungskinetik von Xaa-Pro-Bindungen durch die chemische Natur des Restes Xaa vor Prolin bestimmt [63]. In der N2'-Domäne befindet sich Asp160 vor Pro161. Um die Bedeutung des Aminosäurerestes 160 für die Isomerisierung von Pro161 im nativen und denaturierten Protein zu untersuchen, wurde eine Bibliothek von N2'-Varianten mit allen 20 Aminosäuren an dieser Position hergestellt. Die Substitutionen haben nur einen geringen Effekt auf die Proteinstabilität, mit einer maximalen Differenz von 5,6 kJ mol^{-1} (s. Abb. 2, Teilarbeit D), was vermutlich auf die exponierte Position dieses Kettenbereiches

zurückzuführen ist (s. Abb. 1, Teilarbeit D). Die Analyse der Faltungskinetik aller 20 Varianten zeigte, dass ihre Entfaltung unabhängig von der Natur der Aminosäure an Position 160 ist. Lediglich für die Gly160-Variante sind in den drei Rückfaltungsphasen Unterschiede zu erkennen. Die beiden schnellen Phasen (λ_3 und λ_2) reflektieren die konformationelle Faltung ($U_{cis} \rightarrow N_{cis}$ bzw. $U_{trans} \rightarrow N_{trans}$) während die langsame Reaktion λ_1 der Isomerisierung an Pro161 im nativen Protein ($N_{trans} \rightarrow N_{cis}$) entspricht (Abb. 2-2a). Für die meisten N2'-Varianten variiert diese Rate weniger als zweifach, im Gegensatz zu Peptiden, bei denen die Isomeriserungsrate um bis zu 25-fach variiert [63]. Die Natur der Aminosäure vor Pro161 hat offensichtlich nur einen äußerst geringen Effekt auf die Isomerisierungskinetik in N2' (s. Tab. 2, Teilarbeit D).

Die beiden schnellen konformationellen Faltungsreaktionen (λ_3 und λ_2) können parallel in *stopped-flow*-Messungen verfolgt werden und sind praktisch identisch mit den $U_{cis} \rightarrow N_{cis}$ bzw. $U_{trans} \rightarrow N_{trans}$-Reaktionen. Die Rückfaltungsrate und der Unterschied in ΔG_D zwischen der Variante und dem Wildtypprotein können in einer Brønsted-ähnlichen Auftragung korreliert werden. Der Großteil der N2'-Varianten mit *cis*-Pro161 folgen einer Geraden mit einer Steigung von 1 (s. Abb. 3b, Teilarbeit D). Das deutet darauf hin, dass der Unterschied in ΔG_D zwischen der Variante und dem Wildtypprotein im Verlauf der Faltung bereits im Übergangszustand etabliert ist. Oder in anderen Worten, im Übergangszustand, befindet sich die Aminosäure 160 bereits in der nativähnlichen Konformation (Φ-Wert ≈ 1).

Demgegenüber ist die Rückfaltung der Moleküle mit einem *trans*-Pro161 ($U_{trans} \rightarrow N_{trans}$) kaum durch die Substitution an Position 160 beeinflusst, und für die meisten Varianten ergibt sich keine Korrelation zwischen der Gleichgewichts- und der Aktivierungsenthalpie (s. Abb. 3c, Teilarbeit D). Zusammenfassend zeigen die Brønsted-Auftragungen, dass im Übergangszustand die *cis*-Form schon nativähnlich geordnet ist, wohingegen dieser in der *trans*-Form ungeordnet ist.

Aus den Rückfaltungsexperimenten lässt sich anhand der Rückfaltungsamplituden auch direkt das *cis*/*trans*-Verhältnis der einzelnen Varianten im denaturierten Zustand bestimmen. In der entfalteten Form wird der geringste *cis*-Anteil für Ala160-Pro161 (7,5 %) und der höchste für Trp160-Pro161 (33,3 %) beobachtet. Der ermittelte *cis*-Anteil stimmt sehr gut mit den aus Pentapeptiden erhaltenen Daten überein (außer Pro), was darauf hindeutet, dass keine nichtlokalen Wechselwirkungen um Pro161 in der denaturierten N2'-Domäne existieren (s. Abb. 4a, Teilarbeit D). Dieses Ergebnis ist wichtig, da momentan evtl. vorhandene Reststrukturen und daraus resultierende konformationelle Beschränkungen im entfalteten Zustand von Proteinen intensiv untersucht werden [172; 173].

Das *cis*/*trans*-Verhältnis in nativer N2' kann durch die bereits angesprochene Doppelsprungmischtechnik (s. 2.3.1) bestimmt werden. Im nativen Zustand von N2 dominiert die *cis*-Form und alle Varianten mit Substitution von Pro161 haben einen sehr ähnlichen *cis*-Anteil von etwa 90 %. Ein relativ geringer Wert (75,5 %) wird nur für die Gly160-Variante beobachtet, bei der möglicherweise die lokale Flexibilität erhöht ist (s. Abb. 4b, Teilarbeit D). Es gibt praktisch keine Korrelation zwischen den *cis*-Anteilen im nativen und im denaturierten Protein. Dieser Befund zeigt, dass das *cis*/*trans*-Verhältnis in der nativer N2-Domäne nicht durch die Aminosäure vor Pro161 bestimmt wird, und dass die für die Verschiebung des *cis*/*trans*-Gleichgewichts benötigte Energie nicht aus lokalen Interaktionen in der unmittelbaren Umgebung von Pro161 stammt.

Pro161 befindet sich an der Spitze einer exponierten β-Schleife (Abb. 1-1b). Während der Faltung muss also konformationelle Energie vom gefalteten Teil des Moleküls zu dieser Schleife transferiert werden, damit sich das *cis*/*trans*-Verhältnis fast 100-fach verändern kann. Um den molekularen Weg für diesen Energietransfer zu identifizieren, wurden alle Aminosäurereste in den beiden Verbindungspeptiden, die sich links und rechts von Pro161 befinden, einzeln mutiert. Darüber hinaus wurden zusätzlich Insertionen und Deletionen in die verbindenden Schleifensegmente eingeführt. Alle 15 daraus resultierenden N2'-Varianten waren gefaltet und gut handhabbar. Aminosäureaustausche in den Verbindungspeptiden zwischen den β-Faltblattsträngen und Pro161 zeigten nur eine moderate Destabilisierung, wohingegen Substitutionen in den β-Faltblattsträngen N2' stark destabilisierten (s. Tab. 3, Teilarbeit D).

Für alle N2'-Varianten wurde das *cis*/*trans*-Gleichgewicht an Pro161 im nativen Protein durch unsere Doppelmischmethode bestimmt. Daraus wurde dann die Freie Enthalpie berechnet, die notwendig war, um dieses Gleichgewicht zu verschieben. Es zeigte sich, dass diese stark an die Stabilität der beiden Faltblattstränge geknüpft ist und speziell durch Mutationen an deren Enden verringert wird (s. Abb. 5, Teilarbeit D). Diese Kopplung der lokalen Stabilität eines zweisträngigen Faltblatts und des *cis*/*trans*-Gleichgewichts an Pro161 in der Schleife zwischen den beiden Strängen deutete darauf hin, dass ein Teil der Freien Energie während der Faltung von N2' durch die energetische Verknüpfung der beiden β-Faltblattstränge bereitgestellt wird.

Die geringste Verschiebung des *cis*/*trans*-Gleichgewichts an Pro161, und damit die schwächste Kopplung zwischen konformationeller Faltung und Pro161-Isomerisierung wurde für die Mutation Gly158Ala und die Alanin-Insertionen nach den Aminosäuren 157 und 162 beobachtet. Für diese stimmt die Verringerung der Kopplungsenergie annähernd mit dem

Stabilitätsverlust durch die Mutation überein. Vermutlich beeinflussen diese die Stabilität von N2', da sie direkt mit der Kopplung von konformationeller Stabilität des β-Faltblatts und der Prolylbindung an der Spitze der Pro161-Schleife interferieren. Parallel zu Mutationen die sich in der Nähe der Pro161-Schleife befinden, wurden auch stabilisierende oder destabilisierende Aminosäureaustausche, wie Asn138Gly und His129Gln untersucht, die sich weit entfernt von Pro161 befinden. Diese zeigten keine Veränderung des *cis*/*trans*-Anteils im Vergleich zum Wildtypprotein. Das impliziert, dass Veränderungen in der konformationellen Stabilität außerhalb dieses Faltblatts das *cis*/*trans*-Gleichgewicht nicht verändern.

Von allen N2'-Varianten wurden die Entfaltungs- und Rückfaltungskinetiken gemessen, um die Auswirkungen der Substitutionen auf die individuelle Stabilität und Faltung der *cis* und der *trans* Form zu untersuchen. Die Substitutionen in der Pro161-Schleife beeinflussen die Faltungskinetik sehr unterschiedlich (s. Abb. 6, Teilarbeit D). Einige Mutationen veränderten die Kinetik kaum, andere zeigten veränderte Faltungsraten sowohl für die *cis* als auch für die *trans* Form. Eine besonders interessante Klasse von Varianten sind solche wie Gly158Ala. Diese Substitution destabilisiert N2' ausschließlich durch eine stark verlangsamte Rückfaltung der *cis*-Form, wohingegen die Faltung der *trans*-Form nicht beeinflusst wird. Die Analyse der Faltungskinetik aller Varianten lieferte einen kompletten Satz an thermodynamischen und kinetischen Parametern für die *trans* und *cis* Form von N2'. Die graphische Darstellung in Abb. 7 von Teilarbeit D veranschaulicht, dass der Effekt der Mutationen auf die Gleichgewichtsstabilität und die Aktivierungsenergie der Faltung gering für die *trans* aber groß für die *cis* Form von N2' ist.

Insbesondere die Substitutionen an den Positionen 158, 159, 162 in den Verbindungspeptiden destabilisieren die *cis*-Form stark, haben aber fast keinen Effekt auf die Stabilität der *trans*-Form. Dies verdeutlicht, dass die Verbindung zwischen Pro161 und den Faltblattsträngen wichtig ist für die konformationelle Stabilität der *cis*-Form, aber nicht der *trans*-Form von N2'. Die kinetische Analyse der *trans*-Form der Mutanten zeigt, dass die Verbindungspeptide im entfalteten Protein, im Übergangszustand und im gefalteten Zustand die gleiche Konformationen haben, und dass dies höchstwahrscheinlich eine entfaltete Konformation ist.

Die Mutationen in den β-Faltblattsträngen selbst destabilisieren N2' (außer T156C) und wirken dabei ebenfalls stark unterschiedlich auf die *trans* und *cis*-Form. Generell destabilisieren die individuellen Substitutionen die *cis*-Form stärker als die *trans*-Form. Die ermittelten Daten stimmen sehr gut mit denen zu den Veränderungen des *cis*/*trans*-Gleichgewichts in Abb. 5 (Teilarbeit D) überein. Die Ergebnisse zeigen einmal mehr die

energetische Kopplung von konformationeller Stabilität der β-Stränge und dem *cis*/*trans*-Verhältnis an Pro161 und verdeutlichen, dass die Interaktion zwischen dem β-Faltblatt verbessert ist, wenn sich Pro161 in der *cis*-Form befindet.

In der *cis*-Form zeigte die Schleife um Pro161 eine definierte native Konformation, wohingegen sie in der *trans*-Form entfaltet ist, da keine energetische Kopplung zu den β-Faltblattsträngen existiert. Die energetische Verbindung zwischen der *cis*-Form an Pro161 und dem β-Faltblatt ist bereits im Übergangszustand voll ausgebildet, da sie durch die Verbindungspeptide übertragen wird. Lokale Ent- und Rückfaltung wird hier also verwendet um die energetische Kommunikation zwischen einer Prolylisomerisierung (hier an Pro161) und Sekundärstrukturelementen (in diesem Fall, ein zweisträngiges β-Faltblatt) zu brechen und wiederherzustellen. Solch eine Verknüpfung ist möglicherweise nicht nur auf späte Faltungsschritte beschränkt, sondern kann auch als langsamer molekulare Schalter verwendet werden, um die Funktion regulatorischer Proteine zu modulieren [48; 69; 71; 93; 112; 116; 161; 168; 169; 170; 171]. Die N2-Domäne bietet ein einfaches System um zu verstehen, wie molekulare Kommunikationswege durch Prolylisomerisierung unterbrochen und wieder regeneriert werden können.

2.4.1 Φ-Wert-Analyse von N2'

Die Faltungsrate eines Proteins wird durch die Höhe der Aktivierungsbarriere bestimmt, das heißt, dem Unterschied in der Freien Energie zwischen dem entfalteten Zustand und dem Übergangszustand. Um eine Proteinfaltungsreaktion vollständig zu verstehen, ist es notwendig den Beitrag individueller Reste und deren Interaktionen zur Stabilität des Übergangszustands zu ermitteln. Experimentell werden dafür Proteinvarianten und die Φ-Wert-Analyse verwendet. Die Φ-Wert-Analyse vergleicht die energetischen Interaktionen von Aminosäuren im Übergangszustand mit denen im nativen Zustand, in beiden Fällen jeweils relativ zum entfalteten Zustand. Vom entfalteten Zustand nimmt man an, dass er durch die Mutation nicht beeinflusst wird. Ein Φ-Wert ist 1, wenn die mutierte Seitenkette sich im Übergangszustand bereits in einer nativähnlichen Umgebung befindet. In diesem Fall wird nur die Rückfaltungskinetik durch den Aminosäureaustausch verändert. Ein Φ-Wert ist 0, wenn die mutierte Seitenkette sich in einem entfalteten Bereich während des Übergangszustands befindet und erst danach ihre nativen Wechselwirkungen ausbildet. In einem solchen Fall wäre nur die Entfaltungskinetik im Vergleich zum Wildtypprotein verändert. Wenn sowohl die Kinetik von Ent- und Rückfaltung verändert ist, so werden

gebrochene Φ-Werte erhalten. Diese sind ein Hinweis darauf, dass die Aminosäure schon im Übergangszustand teilweise ihre native Struktur ausgebildet hat [39; 174; 175].

Viele kleine Proteine wurden mit dieser Methode untersucht, wobei zwei Typen identifiziert wurden: Proteine mit diffusen oder delokalisierten Übergangszuständen, in denen die meisten Interaktionen nur teilweise ausgebildet sind und Proteine mit polarisierten Übergangszuständen, in denen bestimmte Interaktionen schon fast vollständig ausgebildet sein müssen, um die Überquerung des Übergangszustands zu ermöglichen [42; 176; 177; 178; 179; 180; 181; 182; 183; 184; 185].

Die thermodynamische und kinetische Analyse der N2-Varianten zeigte, dass die Pro161-Schleife und die beiden Faltblattstränge 4 und 5 links und rechts davon (As156-165), ihre native Konformation größtenteils schon im Übergangszustand der Faltung ausgebildet haben (s. Teilarbeit D). Um zu überprüfen, ob auch andere Bereiche von N2' ihre nativen Wechselwirkungen schon früh während der Faltung entwickeln, wurden 45 Einzelmutationen (in der Regel zu Alanin), verteilt über das gesamten Protein, eingeführt.

Ein Großteil dieser N2'-Varianten konnte in ausreichender Menge produziert werden. Dabei zeigte sich eine deutliche Korrelation von Proteinstabilität und Ausbeute der Reinigung, wobei instabilere Proteinvarianten zu einer geringeren Proteinmenge führten. Fünf N2'-Varianten, in denen hydrophobe Aminosäuren ersetzt wurden, aggregierten während der Reinigung irreversibel. Diese gehörten zum zentralen Kern des Proteins, was den Beitrag von hydrophoben Wechselwirkungen für die Stabilität von N2' verdeutlicht. Darüber hinaus wurden viele stark destabilisierende Mutationen auch in anderen Bereichen des Proteins gefunden, was darauf hindeutet, dass auch weit entfernte Bereiche wichtig für die thermodynamische Stabilität von N2' sind.

Die meisten der untersuchten Proteinvarianten zeigten eine ähnliche Rückfaltung wie N2', aber aufgrund ihrer zumeist geringeren Stabilität im Vergleich zum Wildtypprotein, eine beschleunigte Entfaltung. Die entsprechenden mutierten Aminosäuren haben also einen geringen Φ-Wert (0 - 0.3), was bedeutet, dass deren Seitenkette ihre nativen Interaktionen erst nach dem Übergangszustands ausbildet. Bereiche mit geringen Φ-Werten sind zumeist N- (As102-130) oder C-terminal (As 170-205) in der N2' zu finden (s. Abb. 2-3).

Hohe Φ-Werte (0,6 – 1,0) wurden für G153 in β-Faltblatt 4 und für G146 in Turn 2 gefunden. Substitutionen an diesen Positionen verändern hauptsächlich die Rückfaltungskinetik, was zeigt, dass deren Interaktionen im Übergangszustand bereits sehr nativähnlich sind. Diese Bereiche schließen sich an die bereits untersuchte Region mit hohen Φ-Werten an. Intermediäre Φ -Werte (0,3 – 0,6) zeigten Aminosäuren, die sich in den β-

Faltblattsträngen 4 und 5 befinden, oder Reste, die mit diesen Bereichen interagieren (Leu125, Val150, Tyr151). Die N2-Domäne besitzt also einen gut definierten Faltungsnukleus, der im Wesentlichen nur zwei Faltblattstränge und die dazwischen liegende Pro161-Schleife beinhaltet (s. Abb. 2-3).

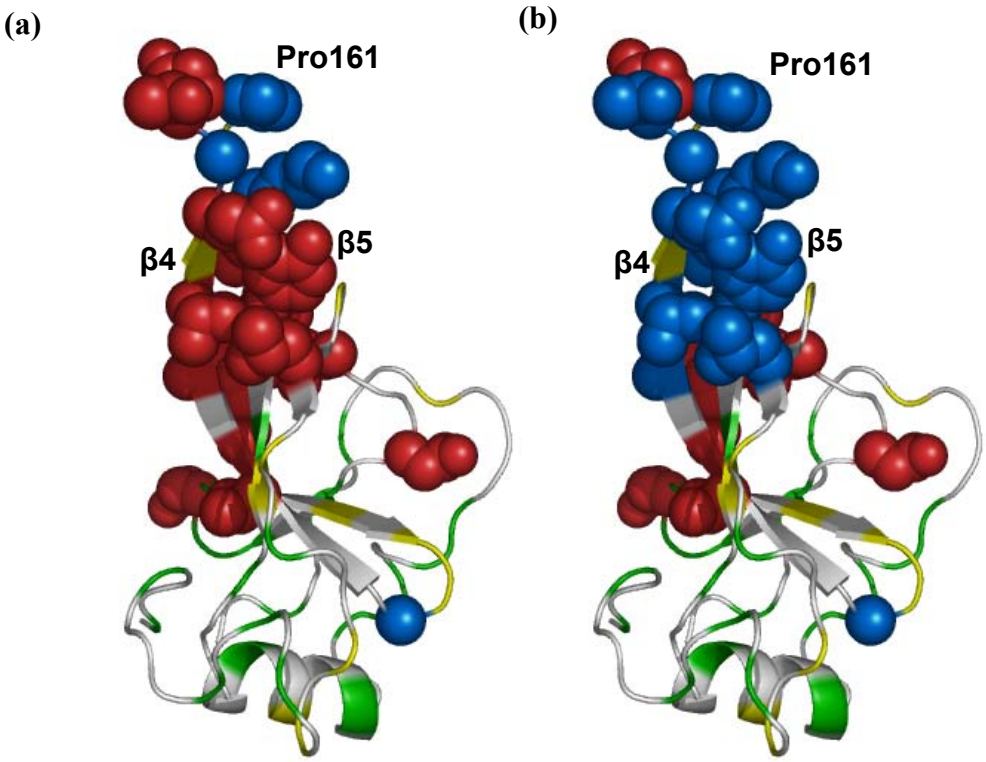


Abbildung 2-3: Φ-Werte für **(a)** die *trans*- und **(b)** die *cis*-Form, die auf die Struktur von N2' aufgetragen sind. Aminosäurereste sind farbkodiert nach ihren Φ-Werten: grün, $0{,}0 < \Phi < 0{,}3$; rot, $0{,}3 < \Phi < 0{,}6$; blau, $0{,}6 < \Phi < 1{,}0$; gelb, Reste mit einem zu geringen Stabilitätsunterschied zum Wildtypprotein ($\Delta\Delta G_{UN} < 2.0$ kJ mol^{-1}). Die Reste mit den höchsten Φ-Werten sind als raumfüllendes Modell gezeigt.

Im Rahmen dieser Arbeit wurde von allen neuen Varianten das *cis*/*trans*-Verhältnis an Pro161 in der nativen Form von N2' durch Doppelmischexperimente ermittelt. Die Ergebnisse waren sehr eindeutig: Alle Aminosäuren, außer V155, verändern das *cis*/*trans*-Gleichgewicht praktisch nicht. Die Mutation V155A befindet sich in Faltblattstrang 4 in der Nähe von Pro161 und verringert den *cis*/*trans*-Anteil von 89 % auf 79 % *cis*. Diese Daten bestätigen, dass das *cis*/*trans*-Verhältnis an Pro161 tatsächlich nur durch die Interaktionen der β-Faltblattstränge bestimmt wird, da die energetische Kopplung nur lokal ausgebildet ist.

2.4.2 Untersuchungen zur Aktivierungsenergie der Isomerisierung an Pro161 in nativer und entfalteter N2'

Aufgrund des partiellen Doppelbindungscharakters der Peptidbindung ist die Aktivierungsenergie für die *cis/trans*-Isomerisierung sehr hoch (ca. 80 kJ mol^{-1}), weshalb Prolylisomerisierungen langsame Reaktionen mit Zeitkonstanten im Bereich von 10 bis 100 s bei 25 °C sind [73; 186]. Die Temperaturabhängigkeit der Reaktion (zur Bestimmung der Aktivierungsenergie) wurde bisher praktisch nur in kurzen Peptiden analysiert [63; 65]. Ob die Isomerisierung in denaturierten und nativen Proteinen wie in Peptiden verläuft, wurde bisher kaum untersucht [187; 188; 189].

Die *cis*→*trans*-Isomerisierung an Pro161 in der entfalteten N2-Domäne (6 M Harnstoff) kann durch Doppelmischexperimente (s. Teilarbeit C), anhand der Amplitudenänderung in Abhängigkeit von der Entfaltungszeit bestimmt werden. Aus der Eyring-Auftragung (Abb. 2-4a) wird aus der Steigung der Regressionsgeraden die Aktivierungsenthalpie der Reaktion $\Delta H^{\ddagger}$ ermittelt. Der erhaltene Wert von $\Delta H^{\ddagger} = 94{,}7$ kJ mol^{-1} stimmt sehr gut mit den bisherigen Werten aus Peptiden überein [64]. Kürzlich wurde die Faltung von N2' mittels intramolekularem Förster-Resonanz-Energie-Transfer (FRET) untersucht [190]. An verschiedenen Positionen in N2' wurden Cysteine eingeführt und an diese spezifisch das Dansylderivat AEDANS gekoppelt. Auf diese Weise konnte die Faltung der N2' anhand des Energietransfers von Tryptophanresten auf den Dansylrest untersucht werden. Dabei wurde die Variante N2'-T156C-AEDANS identifiziert, mit der die Prolylisomerisierung auch direkt anhand der Fluoreszenzänderung des AEDANS-Labels untersucht werden kann. Der so ermittelte Wert von $\Delta H^{\ddagger} = 93{,}1$ kJ mol^{-1} stimmt sehr gut mit dem aus den Doppelsprüngen ermittelten Wert überein, und verdeutlicht, dass die Aktivierungsenthalpie in Peptiden ähnlich zu der in denaturierten Proteinen ist (Daten nicht gezeigt).

Im Gegensatz zur Entfaltung kann die Temperaturabhängigkeit der Rückfaltung von N2' (in 0 M Harnstoff) direkt anhand einer Trp-Fluoreszenzzunahme oder einer geringen CD-Änderung bei 222 nm untersucht werden, und entspricht praktisch der $N_{trans} \rightarrow N_{cis}$ Reaktion. Das über die Temperaturabhängigkeit dieser Reaktion ermittelte $\Delta H^{\ddagger} = 66{,}9$ kJ mol^{-1} für die Asp-Pro Bindung in wildtypischer N2'-Domäne ist signifikant kleiner, als das der Entfaltung (s. Abb. 2-4b). Die Varianten mit den anderen 19 Xaa-Pro-Bindungen zeigen sehr ähnliche $\Delta H^{\ddagger}$-Werte (65-75 kJ mol^{-1}, Abb. 2-4b). Die 20 N2'-D160X weisen also nicht nur eine vergleichbare Aktivierungsbarriere auf, sondern auch eine ähnliche Rate für die $N_{trans} \rightarrow N_{cis}$-Reaktion (Tabelle 2, Teilarbeit D). Die Ursachen für den signifikanten Unterschied in den

$\Delta H^{\ddagger}$-Werten der Entfaltung und der Isomerisierung an Pro161 im gefalteten Protein bleibt unklar. Möglicherweise wird durch die konformationelle Faltung der β-Faltblattsränge Energie an die Spitze der Pro161-Schleife transferiert, was die Aktivierungsenergie für die Isomerisierung verringert.

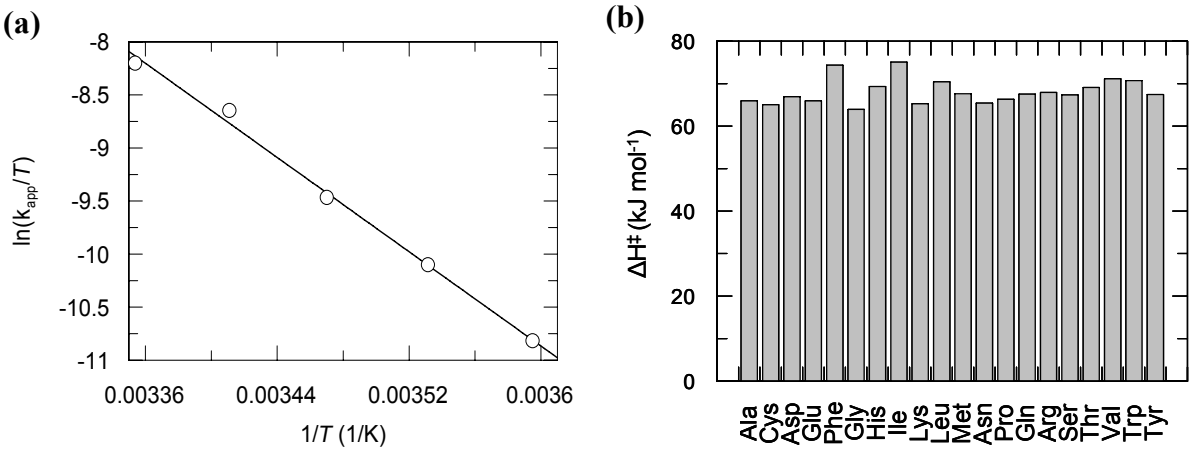


Abbildung 2-4: (a) Eyring Plot für die Prolylisomerisierung von Pro161 im entfalteten Wildtypprotein. Die Geschwindigkeitskonstante wurde durch Doppelmischexperimente bestimmt (s. Teilarbeit C). Dafür wurde das Protein zuerst in 100 mM Glycin pH 2,0 entfaltet und nach unterschiedlich langer Entfaltung in 100 mM K-Phosphat, pH 7,0 wieder zurückgefaltet. Ein nichtlinearer Angleich an die Rückfaltungsamplitude ergab die Geschwindigkeitskonstante bei der jeweiligen Temperatur. Mit Hilfe eines linearen Angleichs an die Daten (durchgezogene Linie) wurde ein $\Delta H^{\ddagger}$ =94,7 kJ mol^{-1} ermittelt (b) Auftragung der Aktivierungsenergie für die $N_{trans} \rightarrow N_{cis}$-Reaktion für die Aminosäure Xaa-Pro in N2', die anhand der Änderung der Tryptophanfluoreszenz bei 0 M Harnstoff ermittelt wurde.

2.5 Die N2'-Domäne als Substrat für Prolylisomerasen

Peptidyl-Prolylisomerisierungen bestimmen die Faltungsraten vieler Proteine [73; 75; 186] und können durch Prolylisomerasen (PPIasen) katalysiert werden [80; 191; 192; 193; 194]. Die Prolylisomerisierung im Verlauf der $N_{trans} \rightarrow N_{cis}$-Reaktion von N2' lässt sich sehr gut durch Cyclophilin und SlyD katalysieren (s. Abb. 3c, Teilarbeit C), mit einer ähnlich hohen katalytischen Effizienz wie die reduzierte und carboxymethylierte Form der RNase T1 (RCM-T1), dem bisherigen Standardproteinsubstrat für PPIasen. Die N2-Domäne ist vermutlich ein hervorragendes Substrat für PPIasen, weil sich Pro161 in einer oberflächenexponierten Schleife befindet, die sehr gut für diese Enzyme zugänglich ist.

Der klassische Enzymtest für PPIasen nutzt die Substratspezifität von Proteasen (wie Chymotrypsin oder Trypsin), die keine Peptide mit *cis*-Prolyl-Bindung als Substrat akzeptieren [191]. Mit Hilfe dieses Tests wurde herausgefunden, dass Mitglieder der FKBP-PPIase-Familie eine hohe Substratspezifität für hydrophobe Aminosäuren an der Xaa-Position

vor Prolin haben [64; 65; 95]. Die Charakterisierung von humanem FKBP12 zeigte dabei, dass es die Isomerisierung von Phe-Pro 1000-fach besser katalysiert, als die von Glu-Pro [95].

Da dieser Test nicht für proteaseempfindliche PPIasen anwendbar ist, wurde vor kurzem ein sehr sensitiver proteasefreier Peptidtest entwickelt (Zoldák *et al*., unveröffentlicht). Dafür werden Tetrapeptide verwendet, die eine Aminobenzoylgruppe am Aminoterminus und eine Paranitroanilingruppe am Carboxyterminus tragen. In diesen Peptiden ist die Fluoreszenz empfindlich auf den Isomerisierungszustand der Prolylbindung. Um die Substratspezifität von PPIasen zu untersuchen, wurde eine Bibliothek von Tetrapeptiden synthetisiert in der die Position vor dem Prolin durch alle 20 natürlichen Aminosäuren ersetzt ist (Zoldák *et al*., unveröffentlicht). Eine entsprechende Proteinbibliothek als Substrat für PPIasen existiert bereits für die N2-Domäne (siehe Teilarbeit C). Diese N2'-Bibliothek wurde verwendet um den Einfluss der Aminosäure auf das *cis*/*trans*-Gleichgewicht an Pro161 im nativen und denaturierten Protein zu charakterisieren. Mit Hilfe dieser beiden Bibliotheken auf Peptid- und auf Proteinbasis wurde die Substratspezifität von Proteinen der FKBP12-Familie umfassend charakterisiert. Dabei waren speziell komplex aufgebaute Prolylisomerasen interessant, wie z.B. SlyD und Triggerfaktor (s. Abb. 1, Teilarbeit E). Diese Proteine besitzen zusätzlich zur PPIase-Domäne noch eine Chaperondomäne, deren Einfluss auf die Aktivität und die Substratspezifität noch nicht verstanden ist.

Der proteasefreie Peptidtest bestätigte die hohe Substratspezifität von FKBP12 (s. Abb. 4a, Teilarbeit E). Es besteht ein etwa 500-facher Unterschied in der katalytischen Effizienz zwischen Leu-Pro (der Besten) und Asp/Glu-Pro (die Schlechtesten). FKBP12 beschleunigt die Prolylisomerisierung in Proteinfaltungsreaktionen sehr schlecht [104; 195], da ihm eine Chaperondomäne fehlt. Seine katalytische Effizienz k_{cat}/K_M konnte etwa 200-fach verbessert werden, nachdem die Chaperondomäne von SlyD (IF-Domäne, für *insert in flap*) in einer Schleife von FKBP12 nahe dem aktiven Zentrum eingepflanzt worden war [104]. Dieses chimäre Protein wird als FKBP12-IF bezeichnet. Die Insertion der Chaperondomäne hat dabei keinen Einfluss auf die hohe Substratspezifität von FKBP12 gegenüber Peptiden (s. Abb. 4b, Teilarbeit E).

Für die N2'-Domäne als Substrat ergibt sich im Vergleich zu den Peptiden ein völlig anderes Bild. FKBP12 katalysiert die Isomerisierung in Proteinen generell schlecht, etwa 100-1000-fach schlechter als in den entsprechenden Peptiden. Die hohe Substratspezifität bezüglich der Aminosäure vor dem Prolin bleibt aber erhalten. Dies ändert sich drastisch, wenn die IF-Domäne mit ihrer Chaperonfunktion in FKB12 eingefügt wird. Nun reichen schon 1-10 nM dieses Enzyms um die Rückfaltung aller Mitglieder der N2'-Protein

Bibliothek stark zu beschleunigen (s. Abb. 4b, Teilarbeit E). Das zeigt, dass in Gegenwart der IF-Domäne nicht nur die katalytische Effizienz deutlich erhöht ist (bis zu 1000-fach), sondern dass sie für alle Mitglieder der N2'-Bibliothek fast identisch ist. FKBP12-IF1 zeigt für die meisten N2'-Varianten einen k_{cat}/K_M-Wert von $2 \cdot 10^6$ und $5 \cdot 10^6$ $M^{-1}s^{-1}$. Dies deutet darauf hin, dass in der Gegenwart der IF-Domäne die Art der Aminosäure vor dem Prolin für die katalytische Effizienz von FKBP12 unwichtig ist. Ganz im Gegensatz zu FKBP12 ohne die IF-Domäne. Hier war die sehr geringe PPIase-Aktivität sehr stark von der Aminosäure vor Pro161 abhängig.

Die Chaperondomäne des chimären Proteins FKBP12-IF stammt aus der Prolylisomerase SlyD aus *E. coli*. Um die Funktion der IF-Domäne in ihrer natürlichen Umgebung zu untersuchen, wurde die Substratspezifität von SlyD* ebenfalls mit Hilfe der Peptid- und Proteinbibliotheken charakterisiert. SlyD* ist eine C-terminal verkürzte Form von SlyD, die nur aus der Prolylisomerase und der Chaperondomäne besteht. SlyD* hat bei der Katalyse von Prolylisomerisierungen in Peptidsubstraten eine ähnlich hohe Spezifität wie FKBP12-IF, und bevorzugt ebenfalls sehr stark hydrophobe Reste vor dem Prolin. Genauso wie bei FKBP12-IF war auch bei SlyD diese hohe Substratspezifität verschwunden, wenn die Bibliothek an N2-Varianten als Proteinsubstrate eingesetzt wurde.

Zusätzlich zu SlyD* wurde auch Triggerfaktor untersucht, der strukturell keine Ähnlichkeit zu SlyD und FKBP12-IF hat. Dieser ist mit dem bakteriellen Ribosom nahe der Austrittstelle der naszierenden Proteinkette assoziiert, und wahrscheinlich in die ersten Schritte der *de-novo* Proteinfaltung involviert [101; 196; 197]. Er besteht aus einer Ribosomenbindungsdomäne, einer PPIase-Domäne vom FKBP12-Typ, sowie einer Chaperondomäne, und katalysiert prolinlimitierte Proteinfaltungsreaktionen *in vitro* sehr gut [198; 199; 200]. Wie alle anderen FKBPs, so zeigt auch Triggerfaktor eine hohe Substratspezifität bezüglich der Xaa-Pro-Bindung in Peptiden und ähnelt in seinem Aktivitätsprofil FKBP12-IF (s. Abb. 4bd, Teilarbeit E). Im Proteinfaltungstest mit der Bibliothek an N2'-Varianten ist die Aktivität von Triggerfaktor extrem hoch und, wie in den Fällen von FKBP12-IF und SlyD, ist sie praktisch unabhängig von der Aminosäure vor Prolin.

Die k_{cat}/K_M-Werte, wie sie im vorausgegangenen Abschnitt bestimmt wurden, charakterisieren die Effizienz eines Enzyms bei geringer Substratkonzentration. Um die kinetischen Parameter k_{cat} und K_M separat zu ermitteln, wurde die Geschwindigkeit der katalysierten Faltungsreaktion als Funktion der Substratkonzentration bestimmt. Beide, die katalysierte und unkatalysierte Faltung des Substratproteins laufen in Gegenwart eines Faltungsenzyms ab, was in der Analyse berücksichtigt werden muss. Dazu wurde eine

Methode verwendet, die zuerst für die Prolylisomerisierung in Peptiden und später für Katalyse der Proteinfaltung adaptiert wurde [102; 104; 201].

Die Enzymkinetik der katalysierten Faltung wurde für drei N2'-Varianten (Ala, Asp, Leu an Position160) als Substrate und für drei PPIasen als Enzyme (FKBP12-IF, SlyD*, Triggerfaktor) gemessen (s. Abb. 6, Teilarbeit E). Die Ergebnisse sind für alle drei Substrate und alle drei PPIasen bemerkenswert ähnlich. Die k_{cat}-Werte liegen jeweils im Bereich von 2 bis 3 s^{-1}, und K_M im Bereich von 0,6 bis 3,8 µM. Für alle drei PPIasen ist der K_M-Wert für Ala und Leu ähnlich, und für Asp lediglich zweifach höher. Die aus den individuellen Parametern k_{cat} und K_M berechneten Quotienten k_{cat}/K_M sind den zusammengesetzten k_{cat}/K_M-Werten sehr ähnlich, die aus den Experimenten bei geringer Substratkonzentration ermittelt wurden (s. Abb. 3 und Tab. 2, Teilarbeit E). Für FKBP12 ohne Chaperondomäne konnte kein Sättigungsverhalten in katalysierten Faltungsexperimenten mit N2'-D160L festgestellt, was darauf hindeutet, dass die Substrataffinität von FKBP12 für faltenden Proteine sehr gering ist.

Alle drei PPIasen mit Chaperondomäne zeigten eine ähnlich geringe Substratspezifität. Die Chaperondomäne hat also zwei wichtige Konsequenzen für die katalysierte Proteinfaltung. Sie führt zu einer hohen Affinität für nichtnative Proteine (K_M-Wert ~1 µM) in der katalysierten Faltung und die katalytische Effizienz wird dadurch unabhängig von der Sequenzumgebung der zu isomerisierenden Proline. Dieser Einfluss der Chaperondomänen ist offenbar ein generischer Effekt. Die Chaperondomänen von SlyD und Triggerfaktor unterscheiden sich deutlich in Größe und Form (s. Abb. 1, Teilarbeit E), beeinflussen aber die katalytische Eigenschaften in der gleichen Art und Weise.

Der ermittelte K_M-Wert von ~1 µM stimmt sehr gut mit den bisher bestimmten K_M- und K_I-Werten für das andere Substratprotein RCM-RNase T1 überein ([198; 199], Zoldák *et al*., unveröffentlicht). Eine abgestimmte moderate Affinität ist offensichtlich für eine optimale Funktion der Chaperondomäne in Faltungsenzymen sehr wichtig. Eine zu geringe Affinität würde die Substraterkennung beeinträchtigen, eine zu hohe Affinität würde mit dem Substrattransfer zur Prolylisomerasestelle interferieren.

Die ermittelte Umsatzzahl (2-3 s^{-1}) von FKBP12-IF1, SlyD und Triggerfaktor für die katalysierte Faltung stimmt ebenfalls sehr gut mit vorher bestimmten Werten für RCM-T1 überein [104; 199]. Für die Katalyse der Isomerisierung von Leu-Pro in einem Tetrapeptid durch FKBP12 wurden Werte von 600 s^{-1} für k_{cat} und 500 µM für K_M abgeschätzt [202]. Das bedeutet, dass PPIasen mit Chaperondomänen Proteinsubstrate etwa 100-1000-fach besser binden, aber die Umsatzzahl um den Faktor 100-1000 geringer ist. Diese geringen k_{cat}-Werte reflektieren also nicht die katalysierte *cis*/*trans*-Isomerisierung selbst. Wir vermuten, dass die gemessenen

k_{cat}-Werte durch die Substratdissoziation von der Chaperondomäne und die Effizienz der produktiven Bindung an der Prolylisomerasestelle bestimmt werden. Tatsächlich dissoziieren entfaltete Substratproteine von Triggerfaktor und SlyD mit einer Rate im Bereich von 8 s^{-1} ([203], Zoldák *et al*., unveröffentlicht), was im Bereich der beobachteten k_{cat}-Werte liegt.

Zusammen führt dies zu einem einfachen Modell für den Mechanismus von Prolylisomerasen, die Chaperondomänen für die Substratbindung verwenden. Zu Beginn bindet die Chaperondomäne an nichtnative Proteine. Das Bindungsgleichgewicht ist hoch dynamisch, was einen effizienten intramolekularen Substrattransfer zur Prolylisomerasestelle ermöglicht. Produktiver Transfer zum aktiven Zentrum der PPIase ist ein stochastischer Prozess, der mit unproduktiver Bindung und Rückbindung an die Chaperonstelle konkurriert. Die enzymkinetischen Daten deuten darauf hin, dass beide gemessenen kinetischen Parameter der katalysierten Faltung k_{cat} und K_M durch die Chaperondomäne bestimmt werden. K_M reflektiert die Affinität für das rückfaltende Protein und k_{cat} wird vermutlich durch die Dissoziation des Substratproteins von der Chaperondomäne kontrolliert.

Diese Ergebnisse verdeutlichen, dass SlyD und Triggerfaktor generelle Faltungshelfer sind. Der Triggerfaktor ist der erste Faltungshelfer, der mit naszierenden Proteinen interagiert. Bislang war unverstanden, wie eine PPIase mit einer sehr hohen Substratspezifität eine derart generelle Funktion erfüllen kann. Unsere Ergebnisse mit der N2'-Domäne als Substratprotein zeigen, dass die bisherigen Spezifitätsanalysen mit Tetrapeptiden artifiziell sind, da für Proteinsubstrate die Chaperondomänen eine generell hohe Affinität für entfaltete Proteine sicherstellen und damit die katalytische Effizienz der PPIasen bestimmen.

Die Bibliothek an N2'-Varianten mit allen Aminosäuren vor Pro161 hat sich also als ein sehr gutes Substrat zur Untersuchung der Spezifität von PPIasen erwiesen, wodurch in Zukunft sicherlich auch Mitglieder der anderen beiden PPIase-Familien, Parvulin und Cyclophilin charakterisiert werden können. Darüber hinaus gibt es weitere chimäre Prolylisomerasen mit Chaperondomänen, wie FkpA und SurA, die sich strukturell und funktionell stark von den bisher untersuchten PPIasen SlyD und Triggerfaktor unterscheiden.

Im Rahmen der Projektarbeit von Max von Delbrück konnte eine N2'-Variante identifiziert werden (N2'-T156C-AEDANS), die eine zehnfach größere Signaländerung bei der $N_{trans} \rightarrow N_{cis}$ Prolylisomerisierung wie das Wildtypprotein zeigte, und mit ähnlich hoher Aktivität von PPIasen katalysiert wird. Dadurch kann die Rate der katalysierten Prolylisomerisierung noch exakter bestimmt werden. N2'-T156C-AEDANS weist auch bei direkter Anregung des AEDANS-Moleküls eine gut messbare Signaländerung bei der *trans*→*cis*-Isomerisierung auf. Dadurch wird es jetzt möglich sein die Katalyse der

Prolylisomerisierung direkt in Zellextrakten oder die Aktivität von ribosomenassoziiertem Triggerfaktor zu untersuchen.

2.6 Optimierung von Schleifen und Turns in der N2'-Domäne

Die bisherigen Ergebnisse haben gezeigt, dass die N2'-Domäne eine unabhängige Faltungseinheit und Pro161 für die Faltung wichtig ist. Der native Zustand besteht aus zwei gefalteten Formen, die sich in ihrer Konformation an Pro161 unterscheiden (89 % *cis* und 11 % *trans*). Um das *cis*/*trans*-Verhältnis von 7 % *cis* im denaturierten Protein zu 89 % *cis* im nativen zu verschieben, bedarf es 10 kJ mol^{-1}. Diese Energie stammt hauptsächlich aus Interaktionen zwischen den beiden Strängen des β-Faltblatts, das in die Schleife um Pro161 übergeht (s. Abb. 1-1b). Die Stränge verbessern ihre stabilisierenden Interaktionen wenn Pro161 in *cis* ist. Diese Stabilisierung wird zu Pro161 weitergeleitet, weil die verbindenden Peptide zwischen den β-Strängen und Pro161 nativ gefaltet sind, wenn Pro161 in *cis* ist.

Pro161 befindet sich in einer irregulären, sieben Aminosäuren langen Schleife. Es stellt sich die Frage, ob es möglich ist Pro161 zu eliminieren und diese Schleife zu klassischen Turn-Sequenzen zu verkürzen und dabei die 10 kJ mol^{-1} an stabilisierender Energie zu erhalten. Die beiden Enden der β-Stränge überkreuzen sich im Übergang in die Schleife (Abb. 2-5), wobei es abhängig vom Winkel an den Enden ist, ob sie optimal durch fünf (häufig G1-Bulge) oder vier Aminosäure (häufig Typ I'-Turn) verbunden werden können. Da bei N2 ein intermediärer Winkel (~ 40°) vorliegt, wurde die Schleife As157-163 (Schleife 3) sowohl durch mehrere G1-Bulge und Typ I'-Turn-Sequenzen, basierend auf Konsensussequenzanalysen [22; 23], ersetzt.

Alle vier N2'-Varianten, in denen die Pro161-Schleife durch ein G1-Bulge ersetzt wurde, waren gefaltet, aber nur zwei waren stabilisiert. Dagegen waren alle Varianten mit einem Typ I' β-Turn stark stabilisiert, zwischen 7,3 und 9,0 kJ mol^{-1}. Am stabilsten war die Variante mit dem Turn der Sequenz VNGK, was tatsächlich der statistisch am häufigsten vorkommenden Typ I' β-Turn Sequenz entspricht (s. Abb. 2a, Teilarbeit F). Typ I' β-Turns sind aufgrund der Komplementarität zwischen Turn-Geometrie und der rechtshändigen Drehung von β-Faltblattstränge weit verbreitet in Protein β-Turns. Der Turn wird durch die Sequenz (–B1) – (L1) – (L2) – (+B1) beschrieben [23], wobei L1 und L2 die Turn-Reste und –B1 und +B1 die verbindenden Reste in den β-Faltblattsträngen bezeichnen. Die Rückgratwinkel Φ und Ψ für L1 und L2 liegen im linksgängigen (L-α) Bereich der Ramachandran-Auftragung. Deshalb wird an L1 Asn, Asp und Gly in 60 % der Turns gefunden. An der Position L2 sind die Anforderungen strikter (Φ ~ 90° and Ψ ~ 0°), sodass

Gly mit Abstand (90 %) am häufigsten ist. Die statistische Analyse von Typ I' Turns zeigt auch eine Präferenz für die Aminosäurereste in der –B1 und +B1 Position. Zum Beispiel kommt Lysin mit großer Häufigkeit an der +B1 Position vor, da diese vermutlich Wasserstoffbrücken mit umgebenden Aminosäuren eingehen kann. An der Position –B1 gibt es eine klare Präferenz für β-verzweigte Aminosäuren, mit Valin in 20 % aller Typ I' Turns [16; 204].

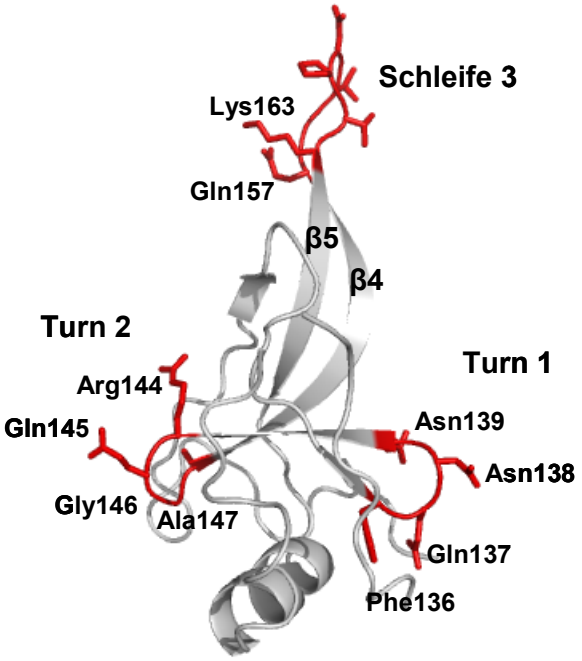


Abbildung 2-5: Tertiärstruktur der N2'-Domäne von G3P (Koordinaten von Holliger *et al.* (1999))[51]. Die Seitenketten von Turn 1 und 2, sowie Schleife 3 sind als Stabmodell gezeigt. Die Abbildung wurde mit Hilfe von Pymol erstellt.

Um den Ursprung der starken Stabilisierung durch die Schleifenoptimierung zu verstehen, wurde die Faltungskinetik einiger Turn 3-N2'-Varianten untersucht. In allen Fällen ist die Entfaltungskinetik unverändert (s. Abb. 3a, Teilarbeit F) und die Stabilisierung stammt ausschließlich aus einer beschleunigten Rückfaltung der Proteine (s. Abb. 3b, Teilarbeit F). Dies deutet darauf hin, dass die stabilisierenden Interaktionen in den optimierten Turns schon sehr früh im Faltungsprozess – vor dem Übergangszustand – ausgebildet werden. Diese Ergebnisse stimmen sehr gut mit denen zum Faltungsmechanismus von N2' überein (s. 2.3).

Die N2'-Domäne besitzt zwei weitere Typ I'-Turns. Turn 1 (As136-139) ist vollständig im Protein vergraben, wohingegen Turn 2 (As144-147) größtenteils oberflächenexponiert ist (s. Abb. 2-5). Die Aminosäuresequenzen von Turn 1 FQNN und Turn 2 RQGA weichen stark von der Konsensussequenz VNGK ab. Deshalb wurden sieben Turn 2 und sechs Turn 1-Varianten mit einzelnen, oder multiplen Aminosäureaustauschen

hergestellt, und deren Einfluss auf die Stabilität und Faltung untersucht. In Turn 1 weichen die Aminosäuren an allen vier Positionen vom Konsensus ab und wurden daher zunächst einzeln durch die statistisch am häufigsten vorkommende Aminosäure ersetzt. Lediglich ein einziger Austausch davon, Asn138Gly an Position L2, war stabilisierend ($\Delta\Delta G_D$ = 3,0 kJ mol^{-1}). Das liegt dran, dass Glycin die einzige Aminosäure ist, für die die Rückgratwinkel an dieser Stelle erlaubt sind. Alle anderen Mutationen im Turn 1 wirken destabilisierend, teilweise drastisch (F136V, $\Delta\Delta G_D$ = -15,5 kJ mol^{-1}), weil alle Seitenketten der Aminosäuren (außer N138) in vielfältige van der Waals oder Wasserstoffbrückenbindungen involviert sind, und die Mutation hin zur statistisch bevorzugten Aminosäure nicht die Zerstörung dieser Wechselwirkungen kompensieren kann (Abb. 3a, Teilarbeit F).

Der Turn 2 (As144-147) ist größtenteils oberflächenexponiert und das Gly an Position L2 stimmt bereits mit dem Konsensus überein. Durch den Austausch der anderen Positionen zur jeweils präferierten Aminosäure konnte N2' deutlich stabilisiert werden. Die N2'-Variante mit der optimierten Turnsequenz VNGV ist um 6 kJ mol^{-1} stabiler als das Wildtypprotein. Interessanterweise stabilisieren alle Mutationen in Turn 1 und 2 dadurch, dass sie die Rückfaltung beschleunigen, genauso wie die Turn 3-Varianten (s. Abb. 4, Teilarbeit F). Sie haben also einen Φ-Wert von etwa 1, was darauf hindeutet, dass die Turn-Optimierung die Interaktionen der β-Faltblattstränge im Faltungsprozess erleichtern.

Die stabilsten Turnsequenzen FQGN für Turn 1, VNGV für Turn 2 und VNGK für Turn 3 wurden in einer Proteinvariante 3T-N2' kombiniert. Der T_M dieses Proteins ist stark um 17 °C erhöht (T_M = 54,9 °C; Abb. 6a, Teilarbeit F), und die erreichte Stabilisierung entspricht nahezu der Summe der Beiträge der Einzelmutationen, was darauf hindeutet, dass sich die konsekutiven Turns kaum beeinflussen. Die starke Stabilisierung von 3T-N2' stammt ausschließlich aus der Beschleunigung der Rückfaltung, was mit den Ergebnissen für die N2-Varianten mit den individuellen Turns in Einklang steht.

Die natürliche Funktion der N2-Domäne ist die Bindung an den F-Pilus, im ersten Schritt der Infektion von *E. coli* durch filamentöse Phagen. Die stabilsten Turn-Sequenzen wurden in den fd-Phagen eingebracht und deren Infektiosität bestimmt. Substitutionen in Turn 1 und 3 haben keinen Einfluss, was verdeutlicht, dass das *cis*/*trans*-Gleichgewicht an Pro161 im Wildtyp keine biologische Funktion hat. Mutationen im Turn 2 hingegen verringern die Infektiosität 200-fach. Die erste Aminosäure im Turn 2, Arg144, ist Teil der Bindungsstelle für den F-Pilus [114]. Die Mutation zu Valin reduziert wahrscheinlich die Affinität und verschlechtert dadurch die Infektion. Die Schleifenoptimierung von Turn 2 hat folglich eine Kehrseite, nämlich die Eliminierung der biologischen Funktion. Turn 2 ist

offensichtlich für die Bindung optimiert und nicht auf Stabilität. Ähnliche Stabilitäts/Funktions-Beziehungen wurden wiederholt gefunden. Zum Beispiel optimiert ein veränderter Turn in der WW Domäne von Pin1 die Stabilität und Faltungsenergetik, er eliminiert aber auch die Fähigkeit zur Bindung an phosphorylierte Substrate [205]. Der Phage mit den kombinierten drei stabilsten Turns, ist ungefähr 2000-fach weniger infektiös als der Wildtypphage. Das ist noch einmal 10-fach geringer, als für den Phagen mit Mutationen nur in Turn 2, obwohl die Optimierung in Turn 1 und Turn 3 alleine die Infektion kaum beeinträchtigt. Möglicherweise interferiert eine insgesamt zu hohe Stabilität der N2'-Domäne ebenfalls mit der Infektion. [112].

Das rationale Design von Schleife 3 (mit dem wichtigen Pro161) zeigte, dass die irreguläre Schleife aus sieben Resten am besten durch eine klassische vier Aminosäuren lange Typ I' Turn-Sequenz VNGK ersetzt wird. Mit Hilfe von gerichteter Proteinoptimierung kann nur ein kleiner Bereich des möglichen Sequenzraums untersucht werden. Es stellt sich somit die Frage, ob diese Sequenz VNGK die stabilste ist, oder ob es nicht noch andere Turntypen und/oder bessere Substitutionen gibt. Aus diesem Grund wurden Bibliotheken von Phagen fd hergestellt in denen an den vier Turnpositionen alle 20 Aminosäuren zugelassen waren. Diese Bibliothek wurden mittels *Proside in vitro* durchmustert und dabei zwei Sequenzen etwa gleich häufig gefunden: INGR und IDGR. Diese Sequenzen stimmen sehr gut mit den aus der statistischen Analyse von Typ I' Turns erhaltenen Sequenzen überein: An der ersten Stelle die β-verzweigte Aminosäure Ile, an der zweite Position ein Asp oder Asn, die beide die erforderlichen Rückgratwinkel einnehmen können. An der L2 Position jeweils ein Glycin, das an dieser Stelle am häufigsten zu finden ist (90 %) und an der letzten Position das positive geladene Arginin, das ähnlich wie Lysin Wasserstoffbrückenbindungen zu Aminosäuren im umliegenden Bereich eingehen kann [23]. Beide Turn-Sequenzen stabilisieren die isolierte N2-Domäne ähnlich stark ($\Delta\Delta G_D$ = 10 kJ mol^{-1}) wie die VNGK-Variante aus dem rationalen Design.

Diese Ergebnisse zeigen, zusammen mit Publikationen anderer Arbeitsgruppen, dass durch die Sequenzoptimierung von β-Turns die Proteinstabilität sehr stark erhöht werden kann [15; 16; 17; 18; 19; 20; 206; 207; 208]. Die Präferenzen einzelner Positionen in Sekundärstrukturelementen ist nicht nur für β-Turns, sondern auch für α-Helices sehr gut untersucht [21; 24; 25; 27]. Es gibt mittlerweile viele Beispiele, in denen Proteine durch gerichtete Helix-Optimierung stabilisiert werden konnten [24; 206; 209; 210]. Deshalb sollte es heute möglich sein jedes Protein, für das eine Kristallstruktur vorhanden ist, durch rationales Proteindesign zu stabilisieren.

2.7 Faltung von N2' im Zweidomänenprotein G3P*

Die N2'-Domäne des Phagen fd bildet im G3P*-Gesamtprotein eine kooperative Einheit mit der N1-Domäne. Die beiden vorgefalteten Domänen N1 und N2 assemblieren in einer langsamen Domänenassoziationsreaktion [52; 53], wodurch die N2'-Domäne stark stabilisiert und ihr T_M-Wert um 12 °C erhöht wird ($\Delta\Delta G_D \sim 10$ kJ mol^{-1}, s. Abb. 1, Teilarbeit C). Wie wirken sich nun diese Wechselwirkungen im Gesamtprotein auf die Faltungskinetik von N2' und das *cis*/*trans*-Gleichgewicht an Pro161 aus?

Die N1-Domäne von G3P* ist erheblich stabiler als die N2-Domäne und bleibt daher bei Harnstoffkonzentrationen gefaltet, bei denen N2' entfaltet werden kann. Entfaltung und Rückfaltung von N2' kann also unter Bedingungen untersucht werden, bei denen die N1-Domäne immer gefaltet vorliegt und so nicht zur gemessenen Signaländerung beiträgt. Die Rückfaltung der isolierten N2' ist eine komplexe Reaktion, die in drei Phasen abläuft. Die schnelle Phase mit einer kleinen Amplitude ist die $U_{cis} \rightarrow N_{cis}$ Reaktion. Danach faltet der Großteil der Moleküle in der $U_{trans} \rightarrow N_{trans}$ Phase etwa 20-fach langsamer. Anschließend erfolgt die Isomerisierung an Pro161 mit einer kleinen Signaländerung ($N_{trans} \rightarrow N_{cis}$). Diese drei Phasen für N2' werden auch im Zweidomänenprotein G3P* beobachtet, mit praktisch identischen Zeitkonstanten (Abb. 2-6a). Die Faltung von N2' verläuft in G3P* also unabhängig von der strukturierten, nativ vorliegenden N1-Domäne und der eventuell parallel verlaufenden Faltungsvorgänge in der Gelenkdomäne. Diese Ergebnisse zeigen, dass die gesamte Stabilisierung von N2' in G3P* durch eine verlangsamte Entfaltung verursacht wird, da im nativen Protein die beiden Domänen fest aneinander gekoppelt sind und nur als kooperative Einheit entfalten [52; 53].

Diese Kopplung und damit der Schutz vor Entfaltung findet erst in der finalen Assemblierungsreaktion statt, in der Pro213 isomerisiert, die Gelenksubdomäne vollständig faltet und ihre Interaktionen mit der N1 ausbildet. Die Gelenksubdomäne besteht aus zwei β-Faltblattsträngen, die fest mit der N1 interagieren und in die N2 eingeschoben ist. Die beiden β-Stränge fungieren offensichtlich als eine Art Klammer, die die Enden der N2 zusammenhält und dadurch thermodynamisch vor der Entfaltung schützt.

Pro161 befindet sich in Schleife 3, weit entfernt von der Domänengrenzfläche im G3P*. Um zu untersuchen ob sich das *cis*/*trans*-Gleichgewicht an Pro161 durch die stabilisierenden Wechselwirkungen mit der N1-Domäne verändert, wurden die bereits ausführlich unter 2.3.1 beschriebenen Doppelmischexperimente durchgeführt. Nach kurzzeitigem Entfalten des Proteins in pH 2,0 ist die N2' in G3P* schon vollständig entfaltet, das *cis*/*trans*-Gleichgewicht an Pro161 aber unverändert, sodass anhand der Amplituden der

schnellen und langsamen Rückfaltungsreaktionen das *cis*/*trans*-Verhältnis an Pro161 im nativen Protein bestimmt werden konnte. Bereits nach kurzzeitigem Entfalten werden zwei Rückfaltungsphasen beobachtet, was zeigt, dass auch in G3P* eine Heterogenität an Pro161 existiert. Die Werte von 14 % *trans* und 86 % *cis* stimmen sehr gut mit den Ergebnissen für die isolierte N2 überein (11 % trans/89 % cis). Die Zu- oder Abnahme der Rückfaltungsamplitude mit zunehmender Entfaltungszeit für die *trans*- und die *cis*-Form von Pro161 ergibt die Kinetik der Prolylisomerisierung im denaturierten Protein (Abb. 2-6b). Die ermittelte Zeitkonstante von 50 s ist ebenfalls ähnlich zu dem für die isolierte N2'-Domäne erhaltenen Wert.

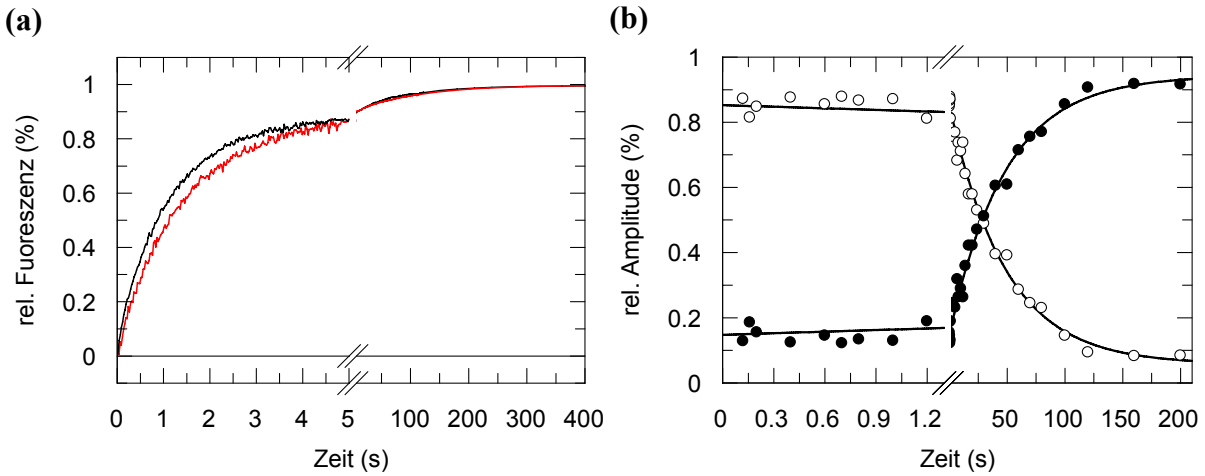


Abbildung 2-6. (a) Übereinstimmende Rückfaltungsfaltungskinetiken der isolierten N2 (schwarz, τ_1 = 0,033 s, τ_2 = 1,08 s, τ_3 = 98 s) und N2 in G3P* (rot, τ_1 = 0,038 s, τ_2 = 1,35 s, τ_3 = 69 s) bei 0,5 M Harnstoff, ausgehend von denaturiertem Protein. Für eine bessere Vergleichbarkeit wurden die Reaktionskinetiken so normiert, dass Start- und Endwerte übereinstimmen. **(b)** *Stopped-flow* Doppelmischexperimente um in G3P* die Isomerisierung von Pro161 in entfaltetem N2' zu untersuchen. Relative Amplituden der schnellen λ_3 (○) und der intermediären λ_2 (●) Rückfaltungsreaktion (τ = 50 s) als eine Funktion der Entfaltungszeit in 100 mM Glycin (pH 2,0). Alle Messungen erfolgten wie in Teilarbeit C beschrieben.

Die Zeitkonstante für die Isomerisierung an Pro161 in nativem G3P* ($N_{trans} \rightarrow N_{cis}$ Reaktion) ist mit 67 s dem entsprechenden Wert für die isolierte N2-Domäne (56 s) ebenfalls sehr ähnlich. Die Doppelmischexperimente verdeutlichten, dass die Faltung der N2-Domäne und die Isomerisierung an Pro161 im G3P* unabhängig von den anderen Domänen abläuft. Darüber hinaus zeigten sie auch, dass die deutliche Stabilisierung der N2'-Domäne in G3P* erst in der langsamen Domänenassoziationsreaktion erreicht wird.

2.8 Die Domänenassemblierung im Gen-3-Protein wird durch Prolylisomerisierung und Stabilisierung eines Wasserstoffbrückennetzwerks kontrolliert

Wie bereits erläutert umfasst die Faltung des G3P-Gesamtproteins einen sehr großen Zeitbereich der sich vom Millisekunden- bis in den Stundenbereich erstreckt. Zuerst falten die

N1-Domäne und die globuläre N2-Domäne unabhängig voneinander, dann assoziieren sie in der langsamen Domänenassemblierungsreaktion, die durch die Rate der *trans→cis* Isomerisierung von Pro213 in der Gelenkregion limitiert wird. Diese Isomerisierung wirkt sehr stark verlangsamend auf die Faltung, führt jedoch zu einer festen Kopplung und starken Stabilisierung der beiden Domänen [52; 53].

Innerhalb eines Zeitbereiches von mehreren Stunden ist dabei ein Intermediat mit einem *trans* Pro213 populiert, über dessen Eigenschaften bisher wenig bekannt war. Deshalb wurde in Zusammenarbeit mit der Gruppe von Jochen Balbach (Universität Halle) mit Hilfe von H/D-Austausch und hochaufgelöster Echtzeit-NMR die Struktur und die Stabilität dieses Intermediats untersucht. Dafür wurde eine G3P* Variante mit vier Substitutionen (T13I, T101I, Q129H und D209Y) verwendet [211], da diese deutlich stabiler als das Wildtypprotein ist und sich deshalb besser für Rückfaltungsstudien eignet. Durch die Experimente konnte geklärt werden wie sich, ausgelöst durch die Pro213 *trans→cis* Isomerisierung, die Struktur von G3P* und die lokale Stabilität während des Übergangs vom lose assoziierten Faltungsintermediat zum gedockten, voll gefalteten Zustand verändert.

Das nicht native *trans* Isomer von Pro213 im lose assoziiertem Faltungsintermediat beeinflusst die einzelnen Strukturelemente von G3P* in unterschiedlicher Weise. Die N2-Domäne hat praktisch schon eine Konformation mit chemischen Verschiebungen wie im vollständig gefalteten Protein erreicht, bevor die Domänenassemblierungsreaktion stattfindet (s. Abb. 2, Teilarbeit G). In der Gelenkregion und in dem Bereich von N1, der dieser Region im nativen Zustand kontaktiert, sind hingegen noch große Bereiche von nicht nativer Struktur vorhanden. Die Faltung der Gelenkregion und ihre Assoziation mit N1 ist somit an die Pro213 *trans* → *cis* Isomerisierung gekoppelt. Im gefalteten G3P* sind tatsächlich der Großteil der Interdomäneninteraktionen in der Gelenkregion und der N1 lokalisiert [51]. Die Schleifen Asp24-Asp28 in N1 und Phe136-Asn139 in N2 verändern ebenfalls ihre Struktur während der langsamen Verriegelungsreaktion. Diese Schleifen sind von der Gelenkregion weit entfernt und in lokale Interaktionen zwischen N1 und N2 im gefalteten Protein involviert. Die Interdomänenkontakte in dieser weit entfernten Region werden offenbar auch durch die Isomerisierung an Pro213 kontrolliert.

Die Gleichgewichtsentfaltung von G3P* erfolgt in zwei Stufen [211]. In der ersten Stufe dissoziieren die Domänen und N2 entfaltet, die N1 bleibt nativ ähnlich gefaltet. Sie entfaltet erst in einem zweiten Übergang, der bei deutlich höherer Temperatur oder Denaturierungsmittelkonzentration erfolgt. Ihr Übergangsmittelpunkt entspricht dem der

isolierten N1-Domäne. N2 hingegen ist als isolierte Domäne deutlich destabilisiert (s. Teilarbeit C).

Amid-NH Austauschexperimente lieferten positionsspezifische Informationen über die lokale Stabilität des vollständig gefalteten Proteins mit *cis* Pro213. Amid-NH mit Schutzfaktoren von $\geq 10^6$ wurden in allen drei Bereichen von G3P lokalisiert, was zeigt, dass N1, N2 und die Gelenkregion eine kooperative Einheit bilden. Der feste Zusammenhalt liefert offenbar den Großteil der Stabilität für das gesamte Protein und bestimmt die globale Entfaltung unter nativen Bedingungen.

Die N2-Domäne profitiert dabei am stärksten von diesen kooperativen Interaktionen. Schutzfaktoren im Bereich der N2-Domäne von G3P* sind bis zu 10^4 höher als in der isolierten N2-Domäne (s. Teilarbeit C). In der gefalteten Form von G3P* werden die Enden der N2-Domäne durch zwei antiparallele β-Faltblattstränge der Gelenkdomäne zusammengehalten, und die entsprechende vorteilhafte Entropieänderung ist vermutlich für den Großteil der starken Stabilisierung von N2 verantwortlich. Ein Großteil dieses entropischen Stabiliserungseffektes auf die N2-Domäne ist dabei schon ausgebildet, wenn die Gelenkregion erst teilweise strukturiert, und Pro213 noch *trans* ist (s. Abb. 3b, Teilarbeit G). Diese Ergebnisse stimmen sehr gut mit den kinetischen Daten zur Faltung von N2' in G3P* überein (s. 2.7).

Insgesamt konnten die NH-Schutzfaktoren für 48 Reste im Faltungsintermediat von G3P* bestimmt werden. Eine Reihe dieser Reste zeigten Schutzfaktoren im Intermediat, die schon so hoch, oder fast so hoch waren, wie im nativen Protein. Sie befinden sich zum Großteil in der N2-Domäne und im oberen Teil von N1 (Abb. 2-7a). Ihre lokalen Stabilitäten sind also im Wesentlichen vom Isomerisierungszustand an Pro213 unbeeinflusst, was damit übereinstimmt, dass sie auch nativähnliche chemische Verschiebung im Intermediat zeigen (Abb. 2c, Teilarbeit G).

Amid-NHs für die sich der Schutzfaktor 10-100-fach während der Pro213-limitierten Domänenassoziationsreaktion erhöht, befinden sich sowohl in N1 als auch in N2. Viele von ihnen sind in der Nähe der Domänengrenzfläche oder im Kern von N2 lokalisiert, was darauf hindeutet, dass diese Bereiche zwar nativ gefaltet sind, aber ihre Struktur durch die Domänenassemblierungsreaktion stabilisiert wird. Die Amid-NHs, deren Schutzfaktor sich 100- oder teilweise mehr als 1000-fach erhöht, wenn die Pro213 *trans*→*cis* Isomerisierung erfolgt, befinden sich alle in der Gelenkregion oder in Positionen in der N1-Domäne, die diesem Bereich gegenüber stehen (Abb. 2-7a).

Amidprotonen des Proteinrückgrats werden vor dem Austausch hauptsächlich durch Wasserstoffverbrückungen mit ihren Akzeptoren geschützt und erhöhte Schutzfaktoren implizieren, dass die entsprechenden H-Brücken verstärkt sind. Die Wasserstoffbrücken, die als Folge der *trans→cis* Isomerisierung an Pro213 besonders stark stabilisiert wurden, bilden eine Kette aus, die die Grenzfläche zwischen N1 und der Gelenkregion zweimal überbrückt (Abb. 2-7a). Diese Kette beginnt bei Val215 in der Gelenkregion. Es bildet eine H-Brücke mit Ser16 in N1 aus, und von dort pflanzt sich die Kette über Asn15, Gly55 fort zu Tyr54 und Val45, immer in N1, und von da aus wieder über Tyr100 und Tyr102 zurück, in die Gelenkregion (Abb. 2-7b). Die starke Stabilisierung des Netzes an Wasserstoffbrücken zwischen den Resten 44-46 in N1 und den Resten 99-102, sowie 206-208 in der Gelenkregion liefert vermutlich die strukturelle Basis und die energetische Triebkraft für die feste Domänenassemblierung, die durch das Umschalten von *trans* → *cis* an Pro213 ausgelöst wird.

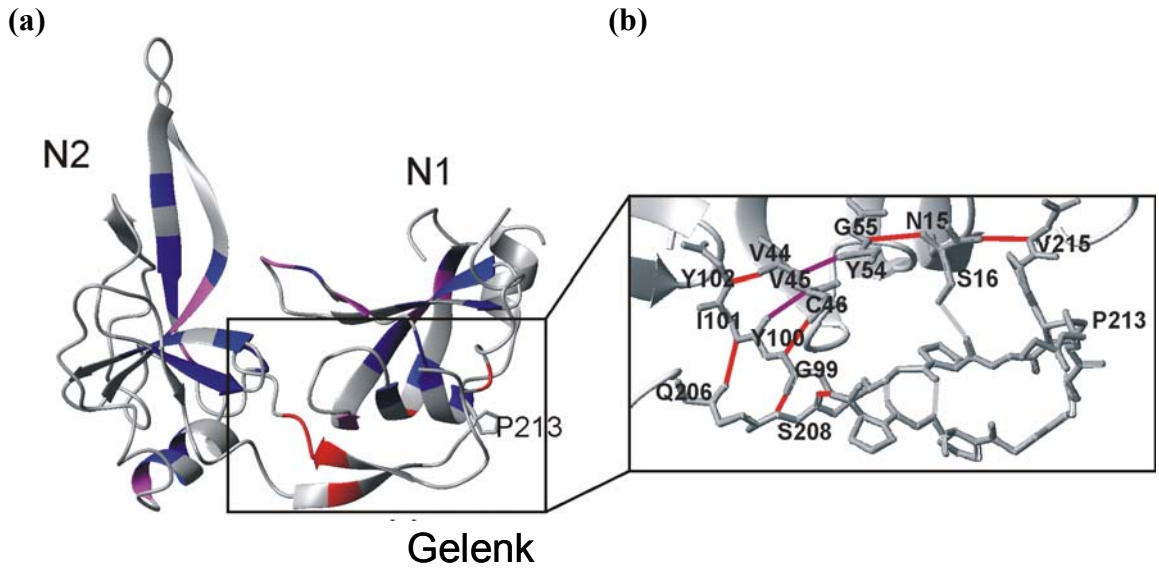


Abbildung 2-7: Stabilität von H-Brücken in N und I. (a) Zunahme der Amid-Schutzfaktoren während der Domänenassemblierungsreaktion von G3P*. Reste in blau haben einen ähnlichen Schutz im nativen Protein und im Faltungsintermediat oder eine weniger als 10-fach Zunahme der Schutzfaktoren. Reste in magenta zeigen eine 10-100-fache Zunahme des Schutzfaktors während der Faltung. Reste in rot haben eine mehr als 100-fach höheren Schutz gegenüber dem Austausch während der Domänenassoziationsreaktion. (b) Wasserstoffbrückennetzwerk in der Gelenkregion und zwischen der Gelenkdomäne und N1. Alle Aminosäurereste sind im Stabmodell gezeigt. Aminosäuren, die in Wasserstoffbrücken involviert sind, wurden markiert. H-Brücken, deren NH-Schutzfaktoren während der I → N Reaktion >100-fach zunehmen sind rot, und solche die 10-100-fach zunehmen magenta eingefärbt. Grau gekennzeichnete H-Brücken, sind in N nicht geschützt.

Die stark stabilisierten Rückgratwasserstoffbrücken in Abb. 2-7b zeigen den molekularen Weg, der die räumliche Spezifität von H-Brücken und den Unterscheid von lokalen Stabilitäten verwendet, um das Signal weiterzuleiten, das durch das Umschalten an Pro213 generiert wurde. Die Verkürzungen zwischen Pro213 und den Anknüpfungspunkten der

Wasserstoffbrücken sind reich an Prolinresten und β-verzweigten Aminosäuren. Das erhöht wahrscheinlich die Kettenrigidität und hilft so, das Signal der Isomerisierung an Pro213 zur anderen Seite der Gelenkregion weiterzuleiten.

Während der Infektion von Phagen wird die Rückreaktion, also die Pro213 *cis→trans* Isomerisierung, genutzt um die Bindungsstelle für den TolA-Rezeptor zu exponieren. Diese Aktivierung verwendet wahrscheinlich den gleichen, umgekehrten Mechanismus. Der letzte Schritt der Faltung wird durch die Bindung des F-Pilus an N2 revertiert, weil dadurch der Cluster an Wasserstoffbrücken zwischen der N1-Domäne und der Gelenkregion geschwächt wird, um die Domänen zu disassemblieren.

Die Ergebnisse zur Kopplung zwischen *cis*/*trans* Isomeren an Pro213 und der Struktur und Stabilisierung der Domäneninteraktion in G3P* entsprechen konzeptionell dem Mechanismus der N2-Domäne, in der die *cis*/*trans* Isomere an Pro161 an die Stabilisierung des β-Faltblatts gekoppelt sind, das in die Pro161-Schleife mündet. Hier wurde deutlich, dass die Faltblattstränge ihre Interaktionen besser ausbilden, wenn Pro161 in *cis* ist, und dabei sowohl Seitenketteninteraktionen, als auch Proteinrückgratwechselwirkungen wichtig sind. Für diese Energieübertragung sind die Verbindungspeptide von Pro161 zu den Faltblattsrängen entscheidend, analog zu den rigiden Bereichen um Pro213, die für die Signalweiterleitung wichtig sind. Durch beide Arbeiten beginnen wir langsam zu verstehen, wie Veränderungen im Isomerenzustand an Prolinresten in einer spezifischen und gerichteten Art und Weise zu anderen Regionen im Protein weitergeleitet werden können, während der Faltung oder bei einem Schaltprozess in einem gefalteten Protein.

2.9 Die Prolylisomerasedomäne von PpiD aus *E .coli* hat eine Parvulin-Faltung aber keine katalytische Aktivität

Während der Biogenese von filamentösen Phagen werden alle Hüllproteine in die innere Membran von *E. coli* eingelagert, bevor sie durch den Gen-1/Gen-4/Gen-11-Proteinkomplex assembliert werden [212; 213; 214; 215; 216]. Dies gilt auch für das Gen-3-Protein, das über seinen C-terminalen hydrophoben Bereich (As396-420) in die innere Membran insertiert, wobei die drei löslichen Domänen ins Periplasma gerichtet sind. Der Faltungsprozess des Gen-3-Proteins erstreckt sich von Millisekunden bis Stunden. Die feste Assoziation der beiden N-terminalen Domänen in der abschließenden Faltungsreaktion ist sehr langsam, da sie durch die *trans→cis* Isomerisierung der Gln212–Pro213 Peptidbindung in der Gelenkregion kontrolliert wird [52; 53]. Der intermediär populierte Zustand ist thermodynamisch instabil und proteolyseempfindlich [217]. Vermutlich wird die Isomerisierung an Pro213 durch

periplasmatische Prolylisomerasen katalysiert, und so die stark stabilisierend wirkende Domänenassemblierung beschleunigt.

Im Periplasma gibt es eine Vielzahl an Faltungshelfern. Die dabei vermutlich wichtigste Klasse sind Proteine, die sowohl Chaperon- als auch Prolylisomeraseeigenschaften besitzen, wie FkpA, SurA und PpiD [218; 219; 220; 221]. SurA ist hauptsächlich an der Reifung von Außenmembranproteinen beteiligt, FkpA ist vermutlich für die korrekte Faltung löslicher periplasmatische Proteine wichtig und PpiD soll in die Maturation von inneren Membranproteinen involviert sein. PpiD wurde in einem Screen als *multicopy suppressor* in einem SurA Deletionsstamm gefunden [222], da beide ähnliche Substratspezifitäten haben [223].

PpiD ist über eine einzelne N-terminale Membranhelix in der inneren Membran (As13-35) in der Nähe des SecYEG-Translocons verankert [224]. Es besteht wahrscheinlich aus drei Domänen, wobei für die Erste (As36-265) und die Dritte (As356-623) eine α-helikale Struktur vorhergesagt ist. Die zweite Domäne ähnelt Prolylisomerasen vom Parvulin-Typ. Alle drei Domänen sind, wie G3P, ins Periplasma orientiert. Parvuline sind neben den Cyclophilinen und den FKBP-Proteinen, die dritte strukturelle Klasse von Prolylisomerasen [225; 226; 227]. Sie sind ubiquitär, ungefähr 100 Aminosäuren groß und falten in ein viersträngiges antiparalleles β-Faltblatt, das von vier α-Helices umgeben ist, was als Parvulin-Faltung bezeichnet wird [228; 229].

Um die Funktion der Parvulindomäne von PpiD (As266-355) zu verstehen, wurde sie isoliert produziert und ihre Stabilität, Struktur, Faltung, Peptidbindungsstelle und PPIase-Eigenschaften charakterisiert. Die isolierte Parvulindomäne von PpiD wird im Folgenden als PpiD* bezeichnet. Das Protein ist korrekt gefaltet und zeigt eine hohe Stabilität gegenüber thermischer und harnstoffinduzierter Denaturierung. Seine Gleichgewichtsstabilität und schnelle Faltungskinetik wird sehr gut durch einen einfachen Zweizustandsmechanismus beschrieben (s. Abb. 1, Teilarbeit H). Mit Hilfe von NMR Spektroskopie wurde eine hervorragend definierte Struktur von PpiD* erhalten, die eine klassische Parvulin-Faltung zeigt (s. Abb. 2, Teilarbeit H). Auch die NMR-dynamischen Eigenschaften sind sehr ähnlich zu anderen homologen Parvulinen (s. Abb. 3, Teilarbeit H).

Dartigalongue und Raina (1998) haben eine sehr hohe Prolylisomeraseaktivität von bis zu $3 \cdot 10^9$ M^{-1} s^{-1} für PpiD im klassischen proteasegekoppelten Peptidtest berichtet [230]. Um die Prolylisomeraseaktivität von PpiD* zu untersuchen, wurden verschiedene Peptid- und Proteinsubstrate verwendet. Jedoch waren wir nicht in der Lage die berichtete Prolylisomeraseaktivität zu bestätigen: PpiD* zeigte bis zu Konzentrationen von 20 μM keine

Katalyse (s. Abb. 5, Teilarbeit H). Auch das Gesamtprotein PpiD war als Prolylisomerase inaktiv.

PpiD* unterscheidet sich an allen sieben Aminosäureresten des katalytischen Zentrums von Pin1, dem bislang am besten charakterisierten Parvulin, das eine sehr hohe PPIase-Aktivität zeigt [170; 228; 231]. Versuche PpiD* zu aktivieren, indem diese sieben Schlüsselreste des katalytisch aktiven Homologen in PpiD* transplantiert wurden, einzeln oder in Kombination, blieben erfolglos. Keine der konstruierten 17 Varianten zeigte eine PPIase-Funktion, so dass es unwahrscheinlich erscheint, dass PpiD* eine latente Aktivität besitzt, die durch unterschiedliche Substrate oder durch Mutationen aktiviert werden kann. Dennoch interagiert PpiD* mit Peptiden und die identifizierte Bindungsstelle ist identisch mit dem katalytischen Zentrum der anderen Parvuline.

Was ist die natürliche Funktion von PpiD*? Es ist ubiquitär in Gram-negativen Bakterien, und die Reste an der Substratbindungsstelle sind hochkonserviert (s. Abb. S2, Teilarbeit H). Die Domäne ist sehr wahrscheinlich wichtig für die biologische Funktion von PpiD. Dartigalongue und Raina (1998) haben in einem Screen Positionen identifiziert, die für die PpiD-Funktion wichtig sind [230]. Vier davon sind in der Parvulindomäne, wobei zwei Reste, G312 und G313, vermutlich für die strukturelle Integrität wichtig sind. Die anderen beiden Positionen, G347 und I350, befinden sich in der Peptidbindungsstelle, was darauf hindeutet, dass die Substratbindung der Parvulindomäne an der Funktion von PpiD beteiligt ist [170; 230]. PpiD wurde ursprünglich als *mutlicopy suppressor* von SurA identifiziert. Tatsächlich zeigt die Parvulindomäne von PpiD die höchste strukturelle Ähnlichkeit zur ersten Parvulindomäne von SurA. SurA besteht aus drei globulären Domänen und einer carboxyterminalen Helix. Die N-terminale α-helikale Domäne, die Parvulindomäne und die C-terminale Helix bilden eine kompakte strukturelle Einheit, die als Chaperon aktiv ist. Die zweite Parvulindomäne steht von dieser Einheit ab und ist über zwei lange, vermutlich flexible Linker angeknüpft [99]. Diese ist aktiv als Prolylisomerase, aber offenbar für die SurA-Funktion *in vivo* nicht notwenig. Die erste Parvulindomäne, ist strukturell ähnlich zu der von PpiD, und ebenfalls inaktiv als PPIase in intaktem SurA und als isolierte Domäne [219]. Als ein Teil der Chaperoneinheit scheint sie mit der N-terminalen Domäne von SurA zu kooperieren. Es wurde vorgeschlagen, dass diese Domäne als ein generelles Bindungsmodul für entfaltete Proteine fungiert, und dass die erste Parvulindomäne von SurA eine Spezifität für entfaltete äußere Membranproteine verleiht.

Die Ähnlichkeit zwischen PpiD und SurA erstreckt sich auch auf die helikale aminoterminale Domäne der beiden Proteine. Die Aminosäuren 86-191 in PpiD haben 31 % Sequenzidentität und 50 % Sequenzähnlichkeit zur Region 65-168 in SurA, was dem Großteil

der Chaperoneinheit dieses Proteins entspricht (As21-168). Die Region 54-85 in PpiD hat zwar nur eine geringe Sequenzähnlichkeit zu den Aminosäuren 21-65 in SurA, aber es ist ebenfalls vorhergesagt, dass sie zwei kurze β-Faltblattstränge und eine lange α-Helix an derselben Stelle wie in SurA bildet. Diese lange Helix ist in die Substratbindung von SurA involviert, was möglicherweise auch in PpiD der Fall sein könnte. Die hohe Ähnlichkeit der aminoterminalen Region deutet darauf hin, dass beide eine identische Faltung aufweisen. Zusammen mit der hochkonservierten inaktiven Parvulindomäne, bilden sie wahrscheinlich homologe Chaperoneinheiten mit ähnlicher Architektur und Funktion.

Die Bindungsstelle in der Parvulindomäne von PpiD ist unpolar und ähnelt sehr der entsprechenden Bindungsstelle in der inaktiven Parvulindomäne in SurA. Das erklärt, warum beide ähnliche Peptidsubstrate binden, und warum die Bindung sensitiv gegenüber Triton X-100 ist. Durch die Verwendung einer Vielzahl an Peptidsubstraten, haben Stymest und Klappa (2008) herausgefunden, dass die Interaktion von PpiD eher längerer Peptidketten bedarf als SurA, und dass PpiD keine so hohe Spezifität aufweist [223]. Dieser Unterschied reflektiert möglicherweise die verschiedenen biologischen Funktionen der beiden Proteine. PpiD ist in der inneren Membran verankert, nahe am SecYEC-Translokon [224] und kann mit translozierenden Protein nach der Freisetzung ins Periplasma quervernetzt werden. PpiD ist somit vermutlich der periplasmatische Pförtner an der Austrittsstelle des SecYEC-Translokons und an der Faltung vieler exportierter Proteine beteiligt. SurA dagegen ist ein lösliches periplasmatisches Protein, das an der Reifung von äußeren Membranproteinen beteiligt ist [219; 232; 233; 234; 235].

Der Faltungshelfer Triggerfaktor im Zytosol und SurA und PpiD im Periplasma von *E. coli* zeigen interessante Gemeinsamkeiten in ihrem molekularen Aufbau. Durch eine Ribosomenbindungsdomäne ist der Triggerfaktor an der Austrittsstelle für naszierende Proteine am Ribosom lokalisiert, PpiD befindet sich durch seinen Membrananker neben der Austrittsstelle des SecYEG Translokons. Die Chaperondomänen von Trigger Factor, SurA und vermutlich auch PpiD besitzen eine gemeinsame α-helikale Faltung, und in Triggerfaktor und SurA ist die flexible Prolylisomerasedomäne in einen lange Schleife der Chaperondomäne eingefügt. Dieses Prolylisomerasemodul ist eine FKBP-Domäne in Triggerfaktor und eine Parvulindomäne in SurA. PpiD fehlt eine solche flexibel angeknüpfte aktive Prolylisomerasedomäne, aber es hat mit SurA die inaktive Parvulindomäne gemeinsam, die vermutlich eine integrale Einheit mit der Chaperondomäne in beiden Proteinen bildet. Die funktionelle Verbindung basierend auf dieser strukturellen Homologie ist sicherlich noch

nicht gut verstanden, deshalb sind Strukturuntersuchungen für die beiden anderen Domänen von PpiD notwendig.

2.10 Struktur, Stabilität, Dynamik und Faltung des Gen-3-Proteins des Phagen IKe

Filamentöse Phagen sind eine große Familie von bakteriellen Viren. Am besten untersucht ist die Infektion von *E. coli* durch Ff-Phagen (f1, M13, fd). Der Phage IKe ist nahe verwandt zu den Ff-Phagen und infiziert *E. coli*-Zellen, die N-Pili tragen. Die Infektion wird ebenfalls durch ein Gen-3-Protein (IKe-G3P) vermittelt, in dem allerdings die Reihenfolge der Domänen vertauscht ist. Die erste N-terminale Domäne dieses Proteins (As1-110), IKe-N1, ist funktionell homolog zur N2-Domäne aus dem Phagen fd und initiiert die Infektion durch Bindung an die Spitze des N-Pilus. Sie ist durch einen kurzen Linker mit der zweiten N-terminalen Domäne IKe-N2 (As114-180) verbunden, die im zweiten Infektionsschritt, wie die N1-Domäne des Phagen fd mit der C-terminalen Domäne von TolA, dem finalen Rezeptor wechselwirkt. Die beiden pilusbindenden Domänen der Phagen fd und IKe besitzen keine Sequenzhomologie, wohingegen die beiden TolA-bindenden Domänen eine hohe Ähnlichkeit (55 %) aufweisen [127; 128].

Der Faltungsmechanismus des Gen-3-Proteins des Phagen fd [52; 53], ist, wie oben gezeigt, sehr stark an den Infektionsprozess gekoppelt [112; 116], wobei ein *in vitro*-Faltungsintermediat essentiell für die biologische Aktivität ist. Es stellt sich die Frage, ob dieser Mechanismus bei anderen filamentösen Phagen konserviert ist. Deshalb wurde die Stabilität und der Faltungsmechanismus der isolierten Domänen IKe-N1 und IKe-N2 und des IKe-G3P Zweidomänenkonstrukts (IKe-G3P*) charakterisiert. Parallel dazu wurde versucht die Kristallstruktur von IKe-G3P zu lösen um den strukturellen Aufbau der beiden Gen-3-Proteine zu vergleichen, und gemeinsame oder unterschiedliche Bauprinzipien zu identifizieren.

Die Kristallstruktur des Gesamtproteins konnte nicht gelöst werden, dafür aber die der einzelnen Domänen. Die IKe-N2-Domäne hat eine sehr ähnliche Struktur wie die homologe Domäne N1 des Phagen fd und die Rückgratatome überlagern mit einer *rmsd* von 1.2 Å (Abb. 2-8a). Die beiden strukturell wichtigen Disulfidbrücken sind konserviert, ebenso ein Großteil der Reste die an der TolA-III Bindung involviert sind.

Die N-Pilus bindende Domäne IKe-N1 zeigt weder zur F-pilusbindenden fd-N2-Domäne, noch zu irgendeiner anderen Domäne in der *Protein Data Bank* strukturelle Ähnlichkeit (s. Abb. 2-8a). Das verdeutlicht, dass sich das Protein evolutionär an den N-Pilus angepasst hat. Da dieser nur eine geringe Ähnlichkeit zum F-Pilus (25 % Sequenzidentität)

aufweist, sind beide Phagen-Rezeptorproteine strukturell unterschiedlich. IKe-N1 hat eine ungewöhnliche langgestreckte Struktur, die in zwei Teile unterteilt werden kann: Eine größere Subdomäne (~ 70 As), die die drei langen β-Faltblattsträngen 1, 4, 6 und eine lange gebogene α-Helix beinhaltet, und eine kleine Subdomäne (~ 35 As), die aus den drei kurzen β-Faltblattsträngen 2, 3, 5 besteht und hohe B-Faktoren in der Kristallstruktur zeigt (s. Abb. 2-8a). Hohe B-Faktoren sind häufig ein Hinweis darauf, dass es sich um flexible Bereiche des Proteins handelt. Dies konnte in NMR-Experimenten nicht bestätigt werden. Nur die Spitze dieser Domäne ist flexibel, und der übrige Bereich ist ähnlich rigide wie die andere Subdomäne.

In fd-G3P besitzen beide isolierte Domänen eine unterschiedliche Stabilität. Während die TolA-bindende fd-N1 sehr stabil ist, ist die pilusbindende fd-N2 nur marginal stabil, und wird erst im Gesamtprotein durch die Domänenassoziation stark stabilisiert (s. Teilarbeit H). Im Gegensatz dazu sind beide isolierte Domänen von IKe-G3P thermodynamisch sehr stabil. Besonders bemerkenswert dabei ist, dass die kleine IKe-N2 Domäne erst durch 8 M GdmCl vollständig denaturiert werden kann. Es zeigte sich, dass das Protein durch Salz sehr stark stabilisiert wird, da es eine ungleichmäßige Ladungsverteilung besitzt.

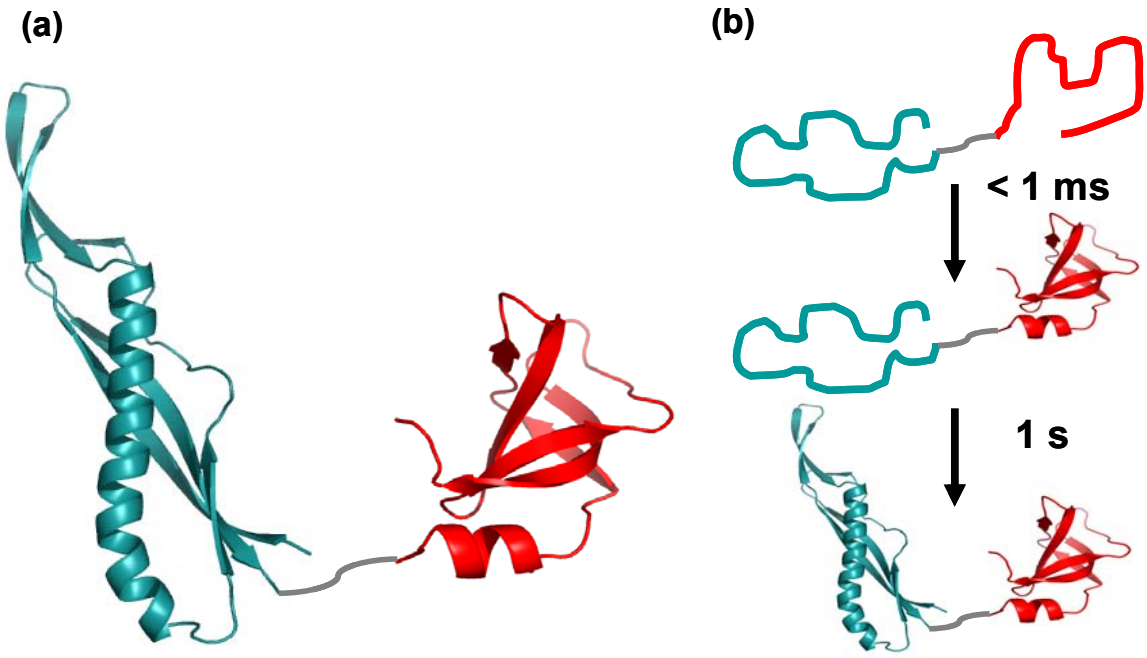


Abbildung 2-8: (a) Strukturmodell von IKe-G3P basierend auf den Kristallstrukturen der isolierten IKe-N1 (hellblau) und IKe-N2 (rot), die durch einen kurzen, sieben Aminosäuren langen Linker (grau) verbunden sind. **(b)** Faltungsmodell von IKe-G3P in 1 M GdmCl.

IKe-N2 und fd-N1 sind nicht nur in ihrer Struktur sehr ähnlich, sondern auch in ihrer Faltungskinetik. Die IKe-N2-Domäne faltet sehr schnell, innerhalb weniger Millisekunden, genauso wie die homologe fd-N1-Domäne (Abb. 2-8b). Das zeigt, dass neben der Struktur

und Funktion auch die thermodynamischen und kinetischen Eigenschaften konserviert sind. Die Faltungskinetik von IKe-N1 ist 100-fach langsamer als die von IKe-N2, wird aber ebenfalls sehr gut durch ein Zweizustandsmodell beschrieben (Abb. 2-8b).

Im Gen-3-Protein des Phagen fd interagieren beide N-terminalen Domänen über eine große Interdomänengrenzfläche. Durch die Assemblierung über die *trans*→*cis* Isomerisierung an Pro213 werden fd-N1 und fd-N2 thermodynamisch aneinander gekoppelt und dabei stark stabilisiert. Eine solche Interaktion wird im IKe-G3P nicht beobachtet. Die beiden Domänen zeigen die gleiche Stabilität und Faltungskinetik als isolierte Proteine und als Teil von IKe-G3P*. Deshalb hatte auch eine Verlängerung des Linkers zwischen den beiden Domänen keinen Einfluss auf deren Eigenschaften. Diese Ergebnisse wurden mittels NMR verifiziert. Das HSQC-Spektrum von IKe-G3P entspricht der Summe der beiden einzelnen Domänen, und bestätigt, dass trotz des sehr kurzen (6-7 As) Linkers zwischen beiden Domänen keine Wechselwirkung erfolgt.

Die Stabilität und Faltungskinetik von IKe-G3P unterscheidet sich somit grundlegend von dem funktionell homologen fd-G3P. IKe-G3P entspricht der Summe der einzelnen Teile, die sich unabhängig voneinander, wie Perlen an einer Schnur verhalten. Ein ähnliches Verhalten wurde auch für Titin gefunden, wo viele Module aneinandergereiht sind, aber nicht interagieren [236; 237]. Im fd-G3P dagegen, wie im Kapitel 2.8 beschrieben, ist die *in-vitro* Faltung viel komplexer und stark mit dem Infektionsprozess verknüpft. Im assemblierten Zustand ist fd-G3P sehr stabil und inert gegen Proteolyse. Während der Infektion wird die Verriegelung beider Domänen durch die Bindung von fd-N2 an den F-Pilus getrennt, wodurch die Bindungsstelle von fd-N1 an TolA-III zugänglich wird. Dafür ist die geringe Stabilität von fd-N2 notwenig [112]. Dieser Infektionsprozess ist hocheffizient, sodass bei einem Phagenüberschuss innerhalb von fünf Minuten praktisch alle *E. coli*-Zellen einer Lösung infiziert werden. Beim Phagen IKe ist offensichtlich eine Kopplung von Entfaltung und Infektion nicht notwendig. Er verwendet ein einfacheres Prinzip. Beide Domänen sind intrinsisch stabil und die Bindungsstellen für die beiden Rezeptoren sind (hochwahrscheinlich) die ganze Zeit zugänglich. Neben dem Ff- und IKe-Phagen gibt es noch einen dritten Typ von filamentösen Phagen der *E. coli* infiziert, der Phage IF1. Dessen Gen-3-Protein ist in seiner Stabilität und seiner Faltungskinetik sehr ähnlich zum IKe-G3P (Stefan Lorenz & Roman Jakob, unveröffentlicht). Für das Verständnis des IKe-G3P sind noch weitere Experimente notwendig, insbesondere Untersuchungen zur Wechselwirkung mit Rezeptorproteinen und zur Infektiosität des IKe-Phagen.

Die molekularen Infektionsprinzipien der Phagen Fd, IKe und IF1 sind sehr ähnlich, dennoch gibt es auch einige Unterschiede. Alle drei verwenden einen Pilus als primären Rezeptor um an *E. coli*-Zellen anzudocken und anschließend als finalen Rezeptor das TolA-III-Protein. Die TolA-III-bindende Domäne ist in allen Phagen hoch konserviert, wohingegen ihre pilusbindenden Domänen sehr unterschiedlich sind, da sie spezifisch an den jeweiligen Wirt adaptiert wurden. Der Fd-Phage nutzt für die Infektion einen molekularen Prolinschalter, der das Gen-3-Protein aus seiner stabilen, nicht aktiven Form in seine instabile, aber infektiöse Form umwandelt. Dieser Mechanismus wird nicht von den Phagen IKe und IF1 verwendet.

3 Abkürzungen

0SS-G3P* stab	disulfidfreies G3P* mit 14 zusätzlichen stabilisierenden Substitutionen
3SS-G3P* stab	G3P* mit drei Disulfidbrücken und 14 zusätzlichen stabilisierenden Substitutionen
AEDANS	5-(((Acetylamino)ethyl)amino)-naphthalin-1-sulfonsäure
As	Aminosäure
CD	Circulardichroismus
ΔC_p	Änderung der Wärmekapazität während der Entfaltung bei konstantem Druck
Csp	Kälteschockprotein
CT	C-terminale Domäne des Gen-3-Proteins
D	Denaturierungsmittel
E_A	Aktivierungsenthalpie einer Reaktion
E. coli	*Escherichia coli*
FKBP	FK-506 Bindeprotein
FPLC	*fast protein liquid chromatography*
FRET	Förster-Resonanz-Energie-Transfer
Gb1	β-Domäne des Streptokokken-Proteins G
ΔG_D	Gibb´sche Freie Enthalpie der Entfaltung, Gibbs-Energie, Freie Enthalpie
$\Delta\Delta G_D$	Differenz der freien Enthalpien der Entfaltung verschiedener Proteinvarianten
G3P	Gen-3-Protein filamentöser Phagen
G3P*	N-terminales Fragment des Gen-3-Proteins filamentöser Phagen mit den Domänen N1 und N2 (Reste 1 - 217 + ProSerGly$(His)_6$)
G8P	Gen-8-Protein filamentöser Phagen
GdmCl	Guanidiumhydrochlorid
$\Delta H^{\ddagger}$	Aktivierungsenthalpie der Reaktion
ΔH_D	van´t Hoff Enthalpie der Denaturierung
Hsp	Hitzeschockprotein
HSQC	*heteronuclear single quantum coherence*
IIHY-G3P*	G3P* mit den Mutationen T13I, T101I, Q129H und D209Y
k_{cat}	Umsatzzahl
K_I	Inhibitionskonstante
K_M	Michaliskonstante
λ	apparente Geschwindigkeitskonstante
m	Kooperativitätsparameter bei denaturierungsmittelinduzierter Entfaltung
N	nativer Zustand eines Proteins
N_{cis} oder N_{trans}	nativer Zustand eines Proteins mit einem *cis* oder *trans* Prolinisomer
N1	N1-Domäne des Gen-3-Proteins
N2	N2-Domäne des Gen-3-Proteins
NMR	*nuclear magnetic resonance*
Φ-Wert	Maß für die Nativähnlichkeit einer Wechselwirkung im Übergangszustand der Faltung
PPIase	Peptidyl-Prolyl *cis*/*trans* Isomerase oder Prolylisomerase
Proside	*protein stability increased by directed evolution*
RCM-T1	reduzierte und carboxymethylierte Form der S54G/P55N Doppelmutante von Ribonuclease T_1
rmsd	*root mean square deviation*
SlyD	Product des slyD (*sensitive-to lysis*) Gens
ssDNA	einzelsträngige DNA
τ	Zeitkonstante einer Reaktion
T_M	Übergangsmittelpunkt thermisch induzierter Entfaltung
TPR	*tetratrico peptide repeat*
U	entfalteter Zustand eines Proteins
U_{cis} oder U_{trans}	entfalteter Zustand eines Proteins mit einem *cis* oder *trans* Prolinisomer

4 Literaturverzeichnis

1. Becktel, W. J. & Schellman, J. A. (1987). Protein stability curves. *Biopolymers* **26**, 1859-77.
2. Kauzmann, W. (1959). *Advances in Protein Chemistry* **14**, 1-63.
3. 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**, 561-602.
4. Darby, N. & Creighton, T. E. (1995). Disulfide bonds in protein folding and stability. *Methods Mol Biol* **40**, 219-52.
5. Makhatadze, G. I. & Privalov, P. L. (1995). Energetics of protein structure. *Advances in Protein Chemistry* **47**, 307-425.
6. Pace, C. N., Shirley, B. A., McNutt, M. & Gajiwala, K. (1996). Forces contributing to the conformational stability of proteins. *Faseb J* **10**, 75-83.
7. Honig, B. & Yang, A. S. (1995). Free energy balance in protein folding. *Adv Protein Chem* **46**, 27-58.
8. Gilson, M. K. & Honig, B. H. (1986). The dielectric constant of a folded protein. *Biopolymers* **25**, 2097-119.
9. Perl, D. & Schmid, F. X. (2001). Electrostatic stabilization of a thermophilic cold shock protein. *J. Mol. Biol.* **313**, 343-357.
10. 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**, 380-383.
11. Martin, A., Kather, I. & Schmid, F. X. (2002). Origins of the high stability of an in vitro-selected cold-shock protein. *J Mol Biol* **318**, 1341-1349.
12. Nakamura, H. (1996). Roles of electrostatic interaction in proteins. *Q Rev Biophys* **29**, 1-90.
13. Mancusso, R., Cruz, E., Cataldi, M., Mendoza, C., Fuchs, J., Wang, H., Yang, X. & Tasayco, M. L. (2004). Reversal of negative charges on the surface of Escherichia coli thioredoxin: pockets versus protrusions. *Biochemistry* **43**, 3835-43.
14. Sun, D. P., Sauer, U., Nicholson, H. & Matthews, B. W. (1991). Contributions of engineered surface salt bridges to the stability of T4 lysozyme determined by directed mutagenesis. *Biochemistry* **30**, 7142-53.
15. Simpson, E. R., Meldrum, J. K., Bofill, R., Crespo, M. D., Holmes, E. & Searle, M. S. (2005). Engineering enhanced protein stability through beta-turn optimization: insights for the design of stable peptide beta-hairpin systems. *Angew Chem Int Ed Engl* **44**, 4939-44.
16. Simpson, E. R., Meldrum, J. K. & Searle, M. S. (2006). Engineering diverse changes in beta-turn propensities in the N-terminal beta-hairpin of ubiquitin reveals significant effects on stability and kinetics but a robust folding transition state. *Biochemistry* **45**, 4220-30.
17. Trevino, S. R., Schaefer, S., Scholtz, J. M. & Pace, C. N. (2007). Increasing protein conformational stability by optimizing beta-turn sequence. *J Mol Biol* **373**, 211-8.
18. Lee, J., Dubey, V. K., Longo, L. M. & Blaber, M. (2008). A logical OR redundancy within the Asx-Pro-Asx-Gly type I beta-turn motif. *J Mol Biol* **377**, 1251-64.
19. Ohage, E. C., Graml, W., Walter, M. M., Steinbacher, S. & Steipe, B. (1997). Beta-turn propensities as paradigms for the analysis of structural motifs to engineer protein stability. *Protein Sci* **6**, 233-41.
20. Kim, J., Brych, S. R., Lee, J., Logan, T. M. & Blaber, M. (2003). Identification of a key structural element for protein folding within beta-hairpin turns. *J Mol Biol* **328**, 951-61.
21. Richardson, J. S. & Richardson, D. C. (1988). Amino acid preferences for specific locations at the ends of a- helices. *Science* **240**, 1648-1652.
22. Hutchinson, E. G. & Thornton, J. M. (1994). A revised set of potentials for beta-turn formation in proteins. *Protein Sci* **3**, 2207-16.
23. Chan, A. W., Hutchinson, E. G., Harris, D. & Thornton, J. M. (1993). Identification, classification, and analysis of beta-bulges in proteins. *Protein Sci* **2**, 1574-90.
24. Munoz, V. & Serrano, L. (1995). Helix design, prediction and stability. *Curr Opin Biotechnol* **6**, 382-6.
25. Serrano, L. & Fersht, A. R. (1989). Capping and alpha-helix stability. *Nature* **342**, 296-9.
26. Serrano, L., Neira, J. L., Sancho, J. & Fersht, A. R. (1992). Effect of alanine versus glycine in alpha-helices on protein stability. *Nature* **356**, 453-5.
27. Serrano, L., Sancho, J., Hirshberg, M. & Fersht, A. R. (1992). Alpha-helix stability in proteins. I. Empirical correlations concerning substitution of side-chains at the N and C-caps and the replacement of alanine by glycine or serine at solvent-exposed surfaces. *J Mol Biol* **227**, 544-59.
28. Levinthal, C. J. (1968). Are there pathways for protein folding? *J. Chim. Phys.* **65**, 44-45.
29. Dobson, C. M., Sali, A. & Karplus, M. (1998). Protein folding - a perspective from theory and experiment. *Angewandte Chemie* **37**, 868-893.
30. Wolynes, P. G. (1997). Folding funnels and energy landscapes of larger proteins within the capillarity approximation. *Proc Natl Acad Sci U S A* **94**, 6170-5.
31. Dill, K. A. & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nat Struct Biol* **4**, 10-9.

32. Dinner, A. R., Sali, A., Smith, L. J., Dobson, C. M. & Karplus, M. (2000). Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem Sci* **25**, 331-9.
33. Vendruscolo, M., Paci, E., Dobson, C. M. & Karplus, M. (2001). Three key residues form a critical contact network in a protein folding transition state. *Nature* **409**, 641-5.
34. Shea, J. E. & Brooks, C. L., 3rd. (2001). From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. *Annu Rev Phys Chem* **52**, 499-535.
35. Dobson, C. M. & Hore, P. J. (1998). Kinetic studies of protein folding using NMR spectroscopy. *Nat Struct Biol* **5**, 504-7.
36. Jackson, S. E. (1998). How do small single-domain proteins fold? *Fold Des* **3**, R81-91.
37. Fersht, A. R. (2000). Transition-state structure as a unifying basis in protein-folding mechanisms: contact order, chain topology, stability, and the extended nucleus mechanism. *Proc Natl Acad Sci U S A* **97**, 1525-9.
38. Daggett, V. & Fersht, A. R. (2003). Is there a unifying mechanism for protein folding? *Trends in Biochemical Sciences* **28**, 18-25.
39. Fersht, A. R. (1995). Characterizing transition states in protein folding: an essential step in the puzzle. *Curr Opin Struct Biol* **5**, 79-84.
40. Matouschek, A., Kellis, J. T., Serrano, L., Bycroft, M. & Fersht, A. R. (1990). Transient folding intermediates characterized by protein engineering. *Nature* **346**, 440-445.
41. Lindorff-Larsen, K., Rogen, P., Paci, E., Vendruscolo, M. & Dobson, C. M. (2005). Protein folding and the organization of the protein topology universe. *Trends Biochem Sci* **30**, 13-9.
42. Itzhaki, L. S., Otzen, D. E. & Fersht, A. R. (1995). The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J Mol Biol* **254**, 260-88.
43. Garcia-Mira, M. M., Boehringer, D. & Schmid, F. X. (2004). The folding transition state of the cold shock protein is strongly polarized. *J Mol Biol* **339**, 555-69.
44. Bolen, D. W. & Rose, G. D. (2008). Structure and energetics of the hydrogen-bonded backbone in protein folding. *Annu Rev Biochem* **77**, 339-62.
45. Netzer, W. J. & Hartl, F. U. (1997). Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature* **388**, 343-9.
46. Jaenicke, R. (1999). Stability and folding of domain proteins. *Prog Biophys Mol Biol* **71**, 155-241.
47. Forman, J. R. & Clarke, J. (2007). Mechanical unfolding of proteins: insights into biology, structure and folding. *Curr Opin Struct Biol* **17**, 58-66.
48. Vogel, C., Bashton, M., Kerrison, N. D., Chothia, C. & Teichmann, S. A. (2004). Structure, function and evolution of multidomain proteins. *Curr Opin Struct Biol* **14**, 208-16.
49. Han, J. H., Batey, S., Nickson, A. A., Teichmann, S. A. & Clarke, J. (2007). The folding and evolution of multidomain proteins. *Nat Rev Mol Cell Biol* **8**, 319-30.
50. Batey, S., Nickson, A. A. & Clarke, J. (2008). Studying the folding of multidomain proteins. *Hfsp J* **2**, 365-77.
51. 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.
52. Martin, A. & Schmid, F. X. (2003). The folding mechanism of a two-domain protein: folding kinetics and domain docking of the gene-3-protein of phage fd. *J. Mol. Biol.* **329**, 599-610.
53. Martin, A. & Schmid, F. X. (2003). A proline switch controls folding and domain interactions in the gene-3-protein of the filamentous phage fd. *J. Mol. Biol.* **331**, 1131-1140.
54. 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**, 140-147.
55. Jabs, A., Weiss, M. S. & Hilgenfeld, R. (1999). Non-proline cis peptide bonds in proteins. *Journal of Molecular Biology* **286**, 291-304.
56. Macarthur, M. W. & Thornton, J. M. (1991). Influence of proline residues on protein conformation. *Journal of Molecular Biology* **218**, 397-412.
57. Ramachandran, G. N. & Mitra, A. K. (1976). An explanation for the rare occurrence of cis peptide units in proteins and polypeptides. *J. Mol. Biol.* **107**, 85-92.
58. Stewart, D. E., Sarkar, A. & Wampler, J. E. (1990). Occurrence and role of cis peptide bonds in protein structures. *Journal of Molecular Biology* **214**, 253-260.
59. 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.
60. Eckert, B. & Schmid, F. X. (2006). Prolin als molekularer Schalter. *BIOspektrum* **2/2006**, 151-153.
61. Cheng, H. N. & Bovey, F. A. (1977). Cis-trans equilibrium and kinetic studies of acteyl-L-proline and glycyl-L-proline. *Biopolymers* **16**, 1465-1472.

62. Grathwohl, C. & Wüthrich, K. (1981). NMR studies of the rates of proline cis-trans isomerization in oligopeptides. *Biopolymers* **20**, 2623-2633.
63. 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**, 449-60.
64. Stein, R. L. (1993). Mechanism of enzymatic and nonenzymatic prolyl cis- trans isomerization. *Advances in Protein Chemistry* **44**, 1-24.
65. Harrison, R. K. & Stein, R. L. (1992). Mechanistic Studies of Enzymic and Nonenzymic Prolyl Cis-Trans Isomerization. *J.Am.Chem.Soc.* **114**, 3464-3471.
66. 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.
67. Chazin, W. J., Kördel, J., Drakenberg, T., Thulin, E., Brodin, P., Grundström, T. & Forsén, S. (1989). Proline isomerism leads to multiple folded conformations of calbindin D9k: direct evidence from two-dimensional NMR spectroscopy. *Proc. Natl. Acad. Sci. USA* **86**, 2195-2198.
68. Feng, Y., Hood, W. F., Forgey, R. W., Abegg, A. L., Caparon, M. H., Thiele, B. R., Leimgruber, R. M. & McWherter, C. A. (1997). Multiple conformations of a human interleukin-3 variant. *Protein Sci* **6**, 1777-82.
69. 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. *Nat Struct Biol* **9**, 900-905.
70. 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 U S A* **99**, 1899-904.
71. Andreotti, A. H. (2003). Native state proline isomerization: An intrinsic molecular switch. *Biochemistry* **42**, 9515-9524.
72. Sarkar, P., Reichman, C., Saleh, T., Birge, R. B. & Kalodimos, C. G. (2007). Proline cis-trans isomerization controls autoinhibition of a signaling protein. *Mol Cell* **25**, 413-26.
73. 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.
74. Kern, D., Kern, G., Scherer, G., Fischer, G. & Drakenberg, T. (1995). Kinetic analysis of cyclophilin-catalyzed prolyl cis/trans isomerization by dynamic NMR spectroscopy. *Biochemistry* **34**, 13594-13602.
75. Schmid, F. X. & Baldwin, R. L. (1978). Acid catalysis of the formation of the slow-folding species of RNase A: evidence that the reaction is proline isomerization. *Proc. Natl. Acad. Sci. U.S.A.* **75**, 4764-4768.
76. Schmid, F. X. (1983). Mechanism of folding of ribonuclease A. Slow refolding is a sequential reaction via structural intermediates. *Biochemistry* **22**, 4690-4696.
77. Calabrese, M. F., Eakin, C. M., Wang, J. M. & Miranker, A. D. (2008). A regulatable switch mediates self-association in an immunoglobulin fold. *Nat Struct Mol Biol* **15**, 965-71.
78. Fanghanel, J. & Fischer, G. (2004). Insights into the catalytic mechanism of peptidyl prolyl cis/trans isomerases. *Front Biosci* **9**, 3453-78.
79. Schiene, C. & Fischer, G. (2000). Enzymes that catalyse the restructuring of proteins. *Current Opinion in Structural Biology* **10**, 40-45.
80. Fischer, G., Tradler, T. & Zarnt, T. (1998). The mode of action of peptidyl prolyl cis/trans isomerases in vivo: binding vs. catalysis. *FEBS Lett* **426**, 17-20.
81. 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.
82. Buchner, J. (1996). Supervising the fold: Functional principles of molecular chaperones. *FASEB J.* **10**, 10-19.
83. Landry, S. J. & Gierasch, L. M. (1994). Polypeptide interactions with molecular chaperones and their relationship to in vivo protein folding. *Annu.Rev.Biophys.Biomol.Struc.* **23,** 645-669.
84. Wandinger, S. K., Richter, K. & Buchner, J. (2008). The Hsp90 chaperone machinery. *J Biol Chem* **283**, 18473-7.
85. Kampinga, H. H. (2006). Chaperones in preventing protein denaturation in living cells and protecting against cellular stress. *Handb Exp Pharmacol*, 1-42.
86. Wegele, H., Muller, L. & Buchner, J. (2004). Hsp70 and Hsp90--a relay team for protein folding. *Rev Physiol Biochem Pharmacol* **151**, 1-44.
87. Nagradova, N. (2007). Enzymes catalyzing protein folding and their cellular functions. *Curr Protein Pept Sci* **8**, 273-82.
88. Gothel, S. F. & Marahiel, M. A. (1999). Peptidyl-prolyl cis-trans isomerases, a superfamily of ubiquitous folding catalysts. *Cell Mol Life Sci* **55**, 423-36.
89. Xu, Y. X. & Manley, J. L. (2007). New insights into mitotic chromosome condensation: a role for the prolyl isomerase Pin1. *Cell Cycle* **6**, 2896-901.
90. Lippens, G., Landrieu, I. & Smet, C. (2007). Molecular mechanisms of the phospho-dependent prolyl cis/trans isomerase Pin1. *Febs J* **274**, 5211-22.

91. Mesa, A., Somarelli, J. A. & Herrera, R. J. (2008). Spliceosomal immunophilins. *FEBS Lett* **582**, 2345-51.
92. Esnault, S., Shen, Z. J. & Malter, J. S. (2008). Pinning down signaling in the immune system: the role of the peptidyl-prolyl isomerase Pin1 in immune cell function. *Crit Rev Immunol* **28**, 45-60.
93. Lu, K. P., Finn, G., Lee, T. H. & Nicholson, L. K. (2007). Prolyl cis-trans isomerization as a molecular timer. *Nat Chem Biol* **3**, 619-29.
94. Shaw, P. E. (2007). Peptidyl-prolyl cis/trans isomerases and transcription: is there a twist in the tail? *EMBO Rep* **8**, 40-5.
95. Harrison, R. K. & Stein, R. L. (1990). Substrate specificities of the peptidyl prolyl cis-trans isomerase activities of cyclophilin and FK-506 binding protein: evidence for the existence of a family of distinct enzymes. *Biochemistry* **29**, 3813-3816.
96. Rahfeld, J. U., Rucknagel, K. P., Schelbert, B., Ludwig, B., Hacker, J., Mann, K. & Fischer, G. (1994). Confirmation of the existence of a third family among peptidyl- prolyl cis/trans isomerases - Amino acid sequence and recombinant production of parvulin. *FEBS Letters* **352**, 180-184.
97. Rahfeld, J.-U., Schierhorn, A., Mann, K.-H. & Fischer, G. (1994). A novel peptidyl-prolyl cis/trans isomerase from Escherichia coli. *FEBS Letters* **343**, 65-69.
98. Saul, F. A., Arie, J. P., Vulliez-le Normand, B., Kahn, R., Betton, J. M. & Bentley, G. A. (2004). Structural and functional studies of FkpA from Escherichia coli, a cis/trans peptidyl-prolyl isomerase with chaperone activity. *J Mol Biol* **335**, 595-608.
99. Bitto, E. & McKay, D. B. (2002). Crystallographic structure of SurA, a molecular chaperone that facilitates folding of outer membrane porins. *Structure* **10**, 1489-98.
100. Weininger, U., Haupt, C., Schweimer, K., Graubner, W., Kovermann, M., Bruser, T., Scholz, C., Schaarschmidt, P., Zoldak, G., Schmid, F. X. & Balbach, J. (2009). NMR solution structure of SlyD from Escherichia coli: spatial separation of prolyl isomerase and chaperone function. *J Mol Biol* **387**, 295-305.
101. Ferbitz, L., Maier, T., Patzelt, H., Bukau, B., Deuerling, E. & Ban, N. (2004). Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature* **431**, 590-6.
102. Scholz, C., Eckert, B., Hagn, F., Schaarschmidt, P., Balbach, J. & Schmid, F. X. (2006). SlyD proteins from different species exhibit high prolyl isomerase and chaperone activities. *Biochemistry* **45**, 20-33.
103. Maier, R., Scholz, C. & Schmid, F. X. (2001). Dynamic association of trigger factor with protein substrates. *J Mol Biol* **314**, 1181-90.
104. Knappe, T. A., Eckert, B., Schaarschmidt, P., Scholz, C. & Schmid, F. X. (2007). Insertion of a chaperone domain converts FKBP12 into a powerful catalyst of protein folding. *J Mol Biol* **368**, 1458-68.
105. Model, P. & Russel, M. (2006). Filamentous Bacteriophage. In *The Bacteriophages* (Calendar, R., ed.), Vol. 2, pp. 146-160. Plenum Publishing, New York.
106. Model, P. & Russel, M. (1988). Filamentous Bacteriophage. In *The Bacteriophages* (Calendar, R., ed.), Vol. 2, pp. 375-456. Plenum Publishing, New York.
107. Marvin, D. A. (1998). Filamentous phage structure, infection and assembly. *Curr Opin Struct Biol* **8**, 150-158.
108. 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.
109. 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.
110. 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**, 5200-5204.
111. 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.
112. Eckert, B. & Schmid, F. X. (2007). A conformational unfolding reaction activates phage fd for the infection of Escherichia coli. *J Mol Biol* **373**, 452-61.
113. Chatellier, J., Hartley, O., Griffiths, A. D., Fersht, A. R., Winter, G. & Riechmann, L. (1999). Interdomain interactions within the gene 3 protein of filamentous phage. *Febs Letters* **463**, 371-374.
114. 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.
115. Deng, L. W., Malik, P. & Perham, R. N. (1999). Interaction of the globular domains of pIII protein of filamentous bacteriophage fd with the F-pilus of *Escherichia coli*. *Virology* **253**, 271-7.
116. Eckert, B., Martin, A., Balbach, J. & Schmid, F. X. (2005). Prolyl isomerization as a molecular timer in phage infection. *Nature Structural & Molecular Biology* **12**, 619-623.
117. Spada, S. & Plückthun, A. (1997). Selectively infective phage (SIP) technology: A novel method for in vivo selection of interacting protein-ligand pairs. *Nature Medicine* **3**, 694-696.

118. Sieber, V., Plückthun, A. & Schmid, F. X. (1998). Selecting proteins with improved stability by a phage-based method. *Nature Biotechnology* **16**, 955-960.
119. 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.
120. Martin, A., Schmid, F. X. & Sieber, V. (2003). Proside: a phage based method for selecting thermostable proteins. *Methods in Molecular Biology* **230**, 57-70.
121. Wiseman, R. L., Dunker, A. K. & Marvin, D. A. (1972). Filamentous bacterial viruses. 3. Physical and chemical characterization of the If1 virion. *Virology* **48**, 230-44.
122. Romero, E. & Meynell, E. (1969). Covert fi- R factors in fi+ R+ strains of bacteria. *J Bacteriol* **97**, 780-6.
123. Peeters, B. P., Peters, R. M., Schoenmakers, J. G. & Konings, R. N. (1985). Nucleotide sequence and genetic organization of the genome of the N-specific filamentous bacteriophage IKe. Comparison with the genome of the F-specific filamentous phages M13, fd and f1. *J Mol Biol* **181**, 27-39.
124. Peeters, B. P., Schoenmakers, J. G. & Konings, R. N. (1987). Comparison of the DNA sequences involved in replication and packaging of the filamentous phages IKe and Ff (M13, fd, and f1). *Dna* **6**, 139-47.
125. Bradley, D. E. (1979). Morphology of pili determined by the N incompatibility group plasmid N3 and interaction with bacteriophages PR4 and IKe. *Plasmid* **2**, 632-6.
126. Bradley, D. E., Coetzee, J. N. & Hedges, R. W. (1983). IncI2 plasmids specify sensitivity to filamentous bacteriophage IKe. *J Bacteriol* **154**, 505-7.
127. Endemann, H., Bross, P. & Rasched, I. (1992). The adsorption protein of phage IKe. Localization by deletion mutagenesis of domains involved in infectivity. *Mol Microbiol* **6**, 471-8.
128. Endemann, H., Gailus, V. & Rasched, I. (1993). Interchangeability of the adsorption proteins of bacteriophages Ff and IKe. *J Virol* **67**, 3332-7.
129. Bross, P., Bussmann, K., Keppner, W. & Rasched, I. (1988). Functional analysis of the adsorption protein of two filamentous phages with different host specificities. *J Gen Microbiol* **134**, 461-71.
130. Kather, I., Bippes, C. A. & Schmid, F. X. (2005). A stable disulfide-free gene-3-protein of phage fd generated by in vitro evolution. *J Mol Biol* **354**, 666-78.
131. Kather, I. (2007). Stabilisierung von Proteinen durch in vitro-Evolution – Selektion und biophysikalische Charakterisierung, Universität Bayreuth.
132. Otzen, D. E., Rheinnecker, M. & Fersht, A. R. (1995). Structural factors contributing to the hydrophobic effect: the partly exposed hydrophobic minicore in chymotrypsin inhibitor 2. *Biochemistry* **34**, 13051-8.
133. Jaenicke, R. & Böhm, G. (1998). The stability of proteins in extreme environments. *Curr. Opin. Struct. Biol.* **8**, 738-748.
134. Chen, J. & Stites, W. E. (2001). Packing is a key selection factor in the evolution of protein hydrophobic cores. *Biochemistry* **40**, 15280-9.
135. Wunderlich, M., Martin, A., Staab, C. A. & Schmid, F. X. (2005). Evolutionary protein stabilization in comparison with computational design. *J Mol Biol* **351**, 1160-8.
136. Max, K. E., Wunderlich, M., Roske, Y., Schmid, F. X. & Heinemann, U. (2007). Optimized variants of the cold shock protein from in vitro selection: structural basis of their high thermostability. *J Mol Biol* **369**, 1087-97.
137. Wunderlich, M., Martin, A. & Schmid, F. X. (2005). Stabilization of the cold shock protein CspB from Bacillus subtilis by evolutionary optimization of Coulombic interactions. *J Mol Biol* **347**, 1063-76.
138. Wunderlich, M. & Schmid, F. X. (2006). In vitro evolution of a hyperstable Gbeta1 variant. *J Mol Biol* **363**, 545-57.
139. Joris, B., Ghuysen, J. M., Dive, G., Renard, A., Dideberg, O., Charlier, P., Frere, J. M., Kelly, J. A., Boyington, J. C., Moews, P. C. & et al. (1988). The active-site-serine penicillin-recognizing enzymes as members of the Streptomyces R61 DD-peptidase family. *Biochem J* **250**, 313-24.
140. Huang, W., Petrosino, J., Hirsch, M., Shenkin, P. S. & Palzkill, T. (1996). Amino acid sequence determinants of beta-lactamase structure and activity. *J Mol Biol* **258**, 688-703.
141. Herzberg, O. (1991). Refined crystal structure of beta-lactamase from Staphylococcus aureus PC1 at 2.0 A resolution. *J Mol Biol* **217**, 701-19.
142. Laminet, A. A. & Plückthun, A. (1989). The precursor of beta-lactamase: purification, properties and folding kinetics. *EMBO Journal* **8**, 1469-1477.
143. Vanhove, M., Raquet, X. & Frere, J.-M. (1995). Investigation of the Folding Pathway of the TEM-1 b-Lactamase. *Proteins: Struct. Funct. Genet.* **22**, 110-118.
144. Vanhove, M., Guillaume, G., Ledent, P., Richards, J. H., Pain, R. H. & Frere, J. M. (1997). Kinetic and thermodynamic consequences of the removal of the Cys-77-Cys-123 disulphide bond for the folding of TEM-1 beta-lactamase. *Biochem J* **321**, 413-7.
145. Vanhove, M., Lejeune, A. & Pain, R. H. (1998). Beta-lactamases as models for protein-folding studies. *Cell Mol Life Sci* **54**, 372-7.

146. Jelsch, C., Mourey, L., Masson, J. M. & Samama, J. P. (1993). Crystal structure of Escherichia coli TEM1 beta-lactamase at 1.8 A resolution. *Proteins* **16**, 364-83.
147. Strynadka, N. C., Adachi, H., Jensen, S. E., Johns, K., Sielecki, A., Betzel, C., Sutoh, K. & James, M. N. (1992). Molecular structure of the acyl-enzyme intermediate in beta-lactam hydrolysis at 1.7 A resolution. *Nature* **359**, 700-5.
148. Minasov, G., Wang, X. & Shoichet, B. K. (2002). An ultrahigh resolution structure of TEM-1 beta-lactamase suggests a role for Glu166 as the general base in acylation. *J Am Chem Soc* **124**, 5333-40.
149. Michel, G., Chantalat, L., Duee, E., Barbeyron, T., Henrissat, B., Kloareg, B. & Dideberg, O. (2001). The kappa-carrageenase of P-carrageenovora features a tunnel-shaped active site: A novel insight in the evolution of clan-B glycoside hydrolases. *Structure* **9**, 513-525.
150. 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.
151. Jaenicke, R. (1999). Stability and folding of domain proteins. *Progress in Biophysics & Molecular Biology* **71**, 155-241.
152. Netzer, W. J. & Hartl, F. U. (1997). Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature* **388**, 343-349.
153. 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.
154. 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**, 266-8.
155. Higgins, K. A., Craik, D. J., Hall, J. G. & Andrews, P. R. (1988). Cis-trans isomerization of the proline residue in insulin studied by 13C NMR spectroscopy. *Drug Design and Delivery* **3**, 159-170.
156. 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 (cisPro43)isoform and the Pro43Gly mutant. *Biochemistry* **29**, 4400-4409.
157. Adjadj, E., Naudat, V., Quiniou, E., Wouters, D., Sautiere, P. & Craescu, C. T. (1997). Solution structure of Lqh-8/6, a toxin-like peptide from a scorpion venom--structural heterogeneity induced by proline cis/trans isomerization. *Eur J Biochem* **246**, 218-27.
158. Yuan, X., Downing, A. K., Knott, V. & Handford, P. A. (1997). Solution structure of the transforming growth factor beta-binding protein-like module, a domain associated with matrix fibrils. *Embo J* **16**, 6659-66.
159. Loh, S. N., Mcnemar, C. W. & Markley, J. L. (1991). Detection and Kinetic Characterization of a Novel Proline Isomerism in Staphylococcal Nuclease by NMR Spectroscopy. *Techniques in Protein Chemistry* **II**, 275-282.
160. Truckses, D. M., Somoza, J. R., Prehoda, K. E., Miller, S. C. & Markley, J. L. (1996). Coupling between trans/cis proline isomerization and protein stability in staphylococcal nuclease. *Protein Sci* **5**, 1907-16.
161. Sarkar, P., Reichman, C., Saleh, T., Birge, R. B. & Kalodimos, C. G. (2007). Proline cis-trans isomerization controls autoinhibition of a signaling protein. *Molecular Cell* **25**, 413-426.
162. Alexandrescu, A. T., Hinck, A. P. & Markley, J. L. (1990). Coupling between local structure and global stability of a protein: mutants of stapylococcal nuclease. *Biochemistry* **29**, 4516-4525.
163. Breheny, P. J., Laederach, A., Fulton, D. B. & Andreotti, A. H. (2003). Ligand Specificity Modulated by Prolyl Imide Bond Cis/Trans Isomerization in the Itk SH2 Domain: A Quantitative NMR Study. *J. Am. Chem. Soc.* **125**, 15706-15707.
164. Ng, K. K. & Weis, W. I. (1998). Coupling of prolyl peptide bond isomerization and Ca2+ binding in a C-type mannose-binding protein. *Biochemistry* **37**, 17977-89.
165. Ng, K. K., Park-Snyder, S. & Weis, W. I. (1998). Ca2+-dependent structural changes in C-type mannose-binding proteins. *Biochemistry* **37**, 17965-76.
166. Wulf, G., Finn, G., Suizu, F. & Lu, K. P. (2005). Phosphorylation-specific prolyl isomerization: is there an underlying theme? *Nat Cell Biol* **7**, 435-41.
167. Joseph, J. D., Yeh, E. S., Swenson, K. I., Means, A. R. & Winkler. (2003). The peptidyl-prolyl isomerase Pin1. *Prog Cell Cycle Res* **5**, 477-87.
168. Asciutto, E. K., Madura, J. D., Pochapsky, S. S., OuYang, B. & Pochapsky, T. C. (2009). Structural and dynamic implications of an effector-induced backbone amide cis-trans isomerization in cytochrome P450cam. *J Mol Biol* **388**, 801-14.
169. OuYang, B., Pochapsky, S. S., Dang, M. & Pochapsky, T. C. (2008). A functional proline switch in cytochrome P450cam. *Structure* **16**, 916-23.
170. Yaffe, M. B., Schutkowski, M., Shen, M. H., Zhou, X. Z., Stukenberg, P. T., Rahfeld, J. U., Xu, J., Kuang, J., Kirschner, M. W., Fischer, G., Cantley, L. C. & Lu, K. P. (1997). Sequence-Specific and Phosphorylation-Dependent Proline Isomerization - a Potential Mitotic Regulatory Mechanism. *Science* **278**, 1957-1960.

171. Fischer, G. & Aumüller, T. (2003). Regulation of peptide bond *cis*/*trans* isomerization by enzyme catalysis and its implication in physiological processes. *Reviews of Physiology, Biochemistry and Pharmacology* **148**, 105-150.
172. Bowler, B. E. (2007). Thermodynamics of protein denatured states. *Mol Biosyst* **3**, 88-99.
173. Shortle, D. (1996). The denatured state (the other half of the folding equation) and its role in protein stability. *Faseb J* **10**, 27-34.
174. Grantcharova, V., Alm, E. J., Baker, D. & Horwich, A. L. (2001). Mechanisms of protein folding. *Curr Opin Struct Biol* **11**, 70-82.
175. Sanchez, I. E. & Kiefhaber, T. (2003). Origin of unusual Phi-values in protein folding: Evidence against specific nucleation sites. *Journal of Molecular Biology* **334**, 1077-1085.
176. Milla, M. E., Brown, B. M., Waldburger, C. D. & Sauer, R. T. (1995). P22 arc repressor: Transition state properties inferred from mutational effects on the rates of protein unfolding and refolding. *Biochemistry* **34**, 13914-13919.
177. Lopez-Hernandez, E. & Serrano, L. (1996). Structure of the transition state for folding of the 129 aa protein CheY resembles that of a smaller protein, CI-2. *Fold.Des.* **1**, 43-55.
178. Villegas, V., Martinez, J. C., Aviles, F. X. & Serrano, L. (1998). Structure of the transition state in the folding process of human procarboxypeptidase A2 activation domain. *J Mol Biol* **283**, 1027-36.
179. Grantcharova, V. P., Riddle, D. S. & Baker, D. (2000). Long-range order in the src SH3 folding transition state. *Proc Natl Acad Sci U S A* **97**, 7084-7089.
180. Riddle, D. S., Grantcharova, V. P., Santiago, J. V., Alm, E., Ruczinski, I. & Baker, D. (1999). Experiment and theory highlight role of native state topology in SH3 folding. *Nature Structural Biology* **6**, 1016-1024.
181. Martinez, J. C. & Serrano, L. (1999). The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nat Struct Biol* **6**, 1010-1016.
182. Northey, J. G. B., Di Nardo, A. A. & Davidson, A. R. (2002). Hydrophobic core packing in the SH3 domain folding transition state. *Nature Structural Biology* **9**, 126-130.
183. Guerois, R. & Serrano, L. (2000). The SH3-fold family: Experimental evidence and prediction of variations in the folding pathways. *Journal of Molecular Biology* **304**, 967-982.
184. McCallister, E. L., Alm, E. & Baker, D. (2000). Critical role of beta-hairpin formation in protein G folding. *Nature Structural Biology* **7**, 669-673.
185. Srivastava, A. K., Sauer, R. T., Paci, E., Friel, C. T., Lindorff-Larsen, K., Radford, S. E., Karplus, M. & Vendruscolo, M. (2002). Mutational studies of protein stability and folding of the hyperstable MYL Arc repressor variant. *Biophys Chem* **101-102**, 35-42.
186. Schmid, F. X. (2002). Prolyl isomerases. *Adv.Protein Chem.* **59**, 243-282.
187. Kiefhaber, T. & Schmid, F. X. (1992). Kinetic coupling between protein folding and prolyl isomerization. II. Folding of ribonuclease A and ribonuclease T1. *J. Mol. Biol.* **224**, 231-240.
188. Tan, Y. J., Oliveberg, M., Otzen, D. E. & Fersht, A. R. (1997). The rate of isomerisation of peptidyl-proline bonds as a probe for interactions in the physiological denatured state of chymotrypsin inhibitor 2. *Journal of Molecular Biology* **269**, 611-622.
189. Jackson, S. E. & Fersht, A. R. (1991). Folding of Chymotrypsin Inhibitor-2 .2. Influence of Proline Isomerization on the Folding Kinetics and Thermodynamic Characterization of the Transition State of Folding. *Biochemistry* **30**, 10436-10443.
190. Delbrück, M. v. (2008). Analyse des Faltungsmechanismus der isolierten N2-Domäne des Gen-3-Proteins mittels Förster-Resonanz-Energietransfer. Projektarbeit, Universität Bayreuth.
191. 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.
192. Fischer, S., Dunbrack, R. L. & Karplus, M. (1994). Cis-trans imide isomerization of the proline dipeptide. *J.Am.Chem.Soc.* **116**, 11931-11937.
193. Balbach, J. & Schmid, F. X. (2000). Proline isomerization and its catalysis in protein folding. In *Mechanisms of Protein Folding* 2nd edit. (Pain, R. H., ed.), pp. 212-237. University Press, Oxford.
194. Schmid, F. X., Mayr, L. M., Mücke, M. & Schönbrunner, E. R. (1993). Prolyl isomerases: Role in protein folding. *Advances in Protein Chemistry* **44**, 25-66.
195. Scholz, C., Zarnt, T., Kern, G., Lang, K., Burtscher, H., Fischer, G. & Schmid, F. X. (1996). Autocatalytic folding of the folding catalyst FKBP12. *Journal of Biological Chemistry* **271**, 12703-12707.
196. Kramer, G., Boehringer, D., Ban, N. & Bukau, B. (2009). The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat Struct Mol Biol* **16**, 589-97.
197. Merz, F., Boehringer, D., Schaffitzel, C., Preissler, S., Hoffmann, A., Maier, T., Rutkowska, A., Lozza, J., Ban, N., Bukau, B. & Deuerling, E. (2008). Molecular mechanism and structure of Trigger Factor bound to the translating ribosome. *Embo J* **27**, 1622-32.

198. Scholz, C., Stoller, G., Zarnt, T., Fischer, G. & Schmid, F. X. (1997). Cooperation of enzymatic and chaperone functions of trigger factor in the catalysis of protein folding. *EMBO Journal* **16**, 54-58.
199. Scholz, C., Mücke, M., Rape, M., Pecht, A., Pahl, A., Bang, H. & Schmid, F. X. (1998). Recognition of protein substrates by the prolyl isomerase trigger factor is independent of proline residues. *J. Mol. Biol.* **277**, 723-732.
200. Stoller, G., Rücknagel, K. P., Nierhaus, K., Schmid, F. X., Fischer, G. & Rahfeld, J.-U. (1995). Identification of the peptidyl-prolyl cis/trans isomerase bound to the Escherichia coli ribosome as the trigger factor. *EMBO Journal* **14**, 4939-4948.
201. 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**, 6127-6134.
202. Perczel, A., Park, K. & Fasman, G. D. (1992). Deconvolution of the circular dichroism spectra of proteins: the circular dichroism spectra of the antiparallel b-sheet in proteins. *Proteins: Structure Function and Genetics* **13**, 57-69.
203. Maier, R., Scholz, C. & Schmid, F. X. (2001). Dynamic association of trigger factor with protein substrates. *Journal of Molecular Biology* **314**, 1181-1190.
204. Ramirez-Alvarado, M., Blanco, F. J., Niemann, H. & Serrano, L. (1997). Role of beta-turn residues in beta-hairpin formation and stability in designed peptides. *J Mol Biol* **273**, 898-912.
205. Jager, M., Zhang, Y., Bieschke, J., Nguyen, H., Dendle, M., Bowman, M. E., Noel, J. P., Gruebele, M. & Kelly, J. W. (2006). Structure-function-folding relationship in a WW domain. *Proc Natl Acad Sci U S A* **103**, 10648-53.
206. Rea, A. M., Simpson, E. R., Meldrum, J. K., Williams, H. E. & Searle, M. S. (2008). Aromatic residues engineered into the beta-turn nucleation site of ubiquitin lead to a complex folding landscape, non-native side-chain interactions, and kinetic traps. *Biochemistry* **47**, 12910-22.
207. Takano, K., Yamagata, Y. & Yutani, K. (2000). Role of amino acid residues at turns in the conformational stability and folding of human lysozyme. *Biochemistry* **39**, 8655-65.
208. Vega, M. C., Martinez, J. C. & Serrano, L. (2000). Thermodynamic and structural characterization of Asn and Ala residues in the disallowed II' region of the Ramachandran plot. *Protein Sci* **9**, 2322-8.
209. Villegas, V., Viguera, A. R., Aviles, F. X. & Serrano, L. (1996). Stabilization of proteins by rational design of alpha-helix stability using helix/coil transition theory. *Fold Des* **1**, 29-34.
210. Baase, W. A., Eriksson, A. E., Zhang, X. J., Heinz, D. W., Sauer, U., Blaber, M., Baldwin, E. P., Wozniak, J. A. & Matthews, B. W. (1992). Dissection of protein structure and folding by directed mutagenesis. *Faraday Discuss*, 173-81.
211. 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.
212. Opalka, N., Beckmann, R., Boisset, N., Simon, M. N., Russel, M. & Darst, S. A. (2003). Structure of the filamentous phage pIV multimer by cryo-electron microscopy. *J Mol Biol* **325**, 461-70.
213. Marciano, D. K., Russel, M. & Simon, S. M. (2001). Assembling filamentous phage occlude pIV channels. *Proc Natl Acad Sci U S A* **98**, 9359-64.
214. Feng, J. N., Model, P. & Russel, M. (1999). A trans-envelope protein complex needed for filamentous phage assembly and export. *Mol Microbiol* **34**, 745-55.
215. Haigh, N. G. & Webster, R. E. (1999). The pI and pXI assembly proteins serve separate and essential roles in filamentous phage assembly. *J Mol Biol* **293**, 1017-27.
216. Russel, M. (1993). Protein-protein interactions during filamentous phage assembly. *J Mol Biol* **231**, 689-97.
217. Eckert, B. (2007). Die Bedeutung der Stabilität und des Faltungsmechanismus des Gen-3-Proteins filamentöser Phagen für die Infektion von Escherichia coli. Dissertation, Universität Bayreuth.
218. Behrens, S. (2002). Periplasmic chaperones--new structural and functional insights. *Structure* **10**, 1469-71.
219. Behrens, S., Maier, R., de Cock, H., Schmid, F. X. & Gross, C. A. (2001). The SurA periplasmic PPIase lacking its parvulin domains functions in vivo and has chaperone activity. *Embo J* **20**, 285-94.
220. Ramm, K. & Plückthun, A. (2000). The periplasmic Escherichia coli peptidylprolyl cis,trans-isomerase FkpA - II. Isomerase-independent chaperone activity in vitro. *Journal of Biological Chemistry* **275**, 17106-17113.
221. Ramm, K. & Plückthun, A. (2001). High enzymatic activity and chaperone function are mechanistically related features of the dimeric E. coli peptidyl-prolyl-isomerase FkpA. *J Mol Biol* **310**, 485-98.
222. Dartigalongue, C. & Raina, S. (1998). A new heat-shock gene, ppiD, encodes a peptidyl-prolyl isomerase required for folding of outer membrane proteins in Escherichia coli. *Embo J* **17**, 3968-80.
223. Stymest, K. H. & Klappa, P. (2008). The periplasmic peptidyl prolyl cis-trans isomerases PpiD and SurA have partially overlapping substrate specificities. *Febs J* **275**, 3470-9.

224. Antonoaea, R., Furst, M., Nishiyama, K. & Muller, M. (2008). The periplasmic chaperone PpiD interacts with secretory proteins exiting from the SecYEG translocon. *Biochemistry* **47**, 5649-56.
225. Rahfeld, J. U., Rucknagel, K. P., Schelbert, B., Ludwig, B., Hacker, J., Mann, K. & Fischer, G. (1994). Confirmation of the existence of a third family among peptidyl- prolyl cis/trans isomerases - Amino acid sequence and recombinant production of parvulin. *FEBS Lett.* **352**, 180-184.
226. Rahfeld, J.-U., Schierhorn, A., Mann, K.-H. & Fischer, G. (1994). A novel peptidyl-prolyl cis/trans isomerase from Escherichia coli. *FEBS Letters* **343**, 65-69.
227. Göthel, S. F. & Marahiel, M. A. (1999). Peptidyl-prolyl cis-trans isomerases, a superfamily of ubiquitous folding catalysts [Review]. *Cellular & Molecular Life Sciences* **55**, 423-436.
228. Ranganathan, R., Lu, K. P., Hunter, T. & Noel, J. P. (1997). Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent. *Cell* **89**, 875-886.
229. Sekerina, E., Rahfeld, J. U., Muller, J., Fanghanel, J., Rascher, C., Fischer, G. & Bayer, P. (2000). NMR solution structure of hPar14 reveals similarity to the peptidyl prolyl cis/trans isomerase domain of the mitotic regulator hPin1 but indicates a different functionality of the protein. *J Mol Biol* **301**, 1003-17.
230. Dartigalongue, C. & Raina, S. (1998). A New Heat-Shock Gene, Ppid, Encodes a Peptidyl-Prolyl Isomerase Required For Folding of Outer Membrane Proteins in Escherichia Coli. *EMBO Journal* **17**, 3968-3980.
231. Lu, K. P., Hanes, S. D. & Hunter, T. (1996). A human peptidyl-prolyl isomerase essential for regulation of mitosis. *Nature* **380**, 544-547.
232. Bitto, E. & McKay, D. B. (2004). Binding of phage-display-selected peptides to the periplasmic chaperone protein SurA mimics binding of unfolded outer membrane proteins. *FEBS Lett* **568**, 94-8.
233. Xu, X., Wang, S., Hu, Y. X. & McKay, D. B. (2007). The periplasmic bacterial molecular chaperone SurA adapts its structure to bind peptides in different conformations to assert a sequence preference for aromatic residues. *J Mol Biol* **373**, 367-81.
234. Lazar, S. W., Almiron, M., Tormo, A. & Kolter, R. (1998). Role of the Escherichia coli SurA protein in stationary-phase survival. *J Bacteriol* **180**, 5704-11.
235. Tormo, A., Almiron, M. & Kolter, R. (1990). surA, an Escherichia coli gene essential for survival in stationary phase. *J Bacteriol* **172**, 4339-47.
236. Scott, K. A., Steward, A., Fowler, S. B. & Clarke, J. (2002). Titin; a multidomain protein that behaves as the sum of its parts. *J Mol Biol* **315**, 819-29.
237. Politou, A. S., Thomas, D. J. & Pastore, A. (1995). The folding and stability of titin immunoglobulin-like modules, with implications for the mechanism of elasticity. *Biophys J* **69**, 2601-10.

5 Publikationsliste

A. Kather, I.[1], **Jakob, R.P**.[1], Dobbek, H., Schmid, F.X. (2008) Changing the determinants of protein stability from covalent to non-covalent interactions by in vitro evolution: a structural and energetic analysis. *J Mol Biol*. **381**, 1040-54.

B. Kather I., **Jakob R.P.**, Dobbek, H. and Schmid, F.X. (2008). Increased folding stability of TEM-1 beta-lactamase by in vitro selection. *J Mol Biol.* **383**, 238-51.

C. **Jakob, R.P.** and Schmid, F.X. (2008). Energetic Coupling between Native-State Prolyl Isomerization and Conformational Protein Folding. *J. Mol. Biol.* **377**, 1560-75.

D. **Jakob, R.P.** and Schmid, F.X. (2009). Molecular determinants of a native-state prolyl isomerization. *J. Mol. Biol.* **387**, 1017-31.

E. **Jakob, R.P.**, Zoldák, G. and Schmid, F.X**.** Chaperone domains adapt prolyl isomerases for a generic function in protein folding*., Manuskript in Vorbereitung*

F. **Jakob, R.P.**, Zierer, B.K., Hofmann, S., Lorenz, S., Dobbek, H. and Schmid, F.X**.** Protein stabilization by elimination of a cis-proline-containing loop and optimization of type I' β-turns*., Manuskript in Vorbereitung*

G. Weininger, U.[1], **Jakob, R.P.**[1], Kovermann, M., Balbach, J. and Schmid, F.X**.** Structural and functional analysis of the parvulin domain of PpiD from E. coli., *Manuskript in Vorbereitung*

H. Weininger, U., **Jakob, R.P.**, Eckert, B., Schweimer, K., Schmid, F.X and Balbach, J. Control of a domain assembly reaction by remote prolyl isomerization via a hydrogen bonding network., *Proc. Natl. Acad. Sci. akzeptiert*

I. **Jakob, R.P.**, Geitner, A.J., Weininger, U., Balbach, J., Dobbek, H. and Schmid, F.X**.** Functional characterization of the gene-3-protein of the filamentous phage IKe and the implication on the infection mechanism*., Manuskript in Vorbereitung*

[1] geteilte Erstautorenschaft

Die Arbeiten A-H sind im Anhang angeführt. Im Rahmen dieser Arbeit entstanden außerdem folgende Publikationen, die nicht Teil der Dissertation sind:

J. Merkens, H., Kappl, R., **Jakob, R.P.**, Schmid, F.X. and Fetzner, S. (2008). Quercetinase QueD of Streptomyces sp. FLA, a monocupin dioxygenase with a preference for nickel and cobalt. *Biochemistry* **47**, 12185-96.

K. Griese, J.J. [1], **Jakob, R.P.** [1], Schwarzinger, S., Dobbek, H. (2006) Xenobiotic reductase A in the degradation of quinoline by Pseudomonas putida 86: physiological function, structure and mechanism of 8-hydroxycoumarin reduction. *J Mol Biol*. **361**, 140-52

6 Darstellung des Eigenanteils

A. Die Ergebnisse zur Selektion und thermodynamischen Charakterisierung der Gen-3-Protein-Varianten wurden von Insa Kather erarbeitet. Die Kristallisation und die Strukturbestimmung der stabilsten Gen-3-Protein-Variante, sowie die Charakterisierung der verschiedenen Varianten der isolierten N2-Domäne wurde von mir durchgeführt. Das Manuskript wurde von Franz X. Schmid, Insa Kather mir formuliert.

B. Die Ergebnisse zur Selektion und thermodynamischen Charakterisierung der TEM-1 β-Varianten wurden von Insa Kather erarbeitet. Die Kristallisation und die Strukturbestimmung der TEM-1 β-Varianten wurde von mir durchgeführt. Das Manuskript wurde von Franz X. Schmid, Insa Kather und mir verfasst.

C. Alle dargestellten Ergebnisse wurden von mir erarbeitet. 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 verfasst.

E. Die Untersuchung der Substratspezifität von SlyD und FKBP12 bezüglich einer Tetrapeptidbibliothek wurde von Gabriel Zoldák durchgeführt. Alle anderen in der Arbeit dargestellten Ergebnisse wurden von mir erarbeitet. Die Publikation wurde von Franz X. Schmid, Gabriel Zoldák und mir formuliert.

F. Die Selektion von Turn 3-Varianten wurde Bettina Zierer im Rahmen ihrer Diplomarbeit durchgeführt. Darüber hinaus hat sie N2-Varianten mit optimierten Turn 1 und Turn 2 analysiert. Einige N2-Varianten mit Substitutionen in Turn 3 wurden von Stefanie Hofmann im Rahmen ihrer Bachelorarbeit charakterisiert. Alle anderen in der Arbeit dargestellten Ergebnisse wurden von mir erarbeitet. Die Publikation wurde von Franz X. Schmid und mir verfasst.

G. Die Strukturbestimmung und die dynamische Charakterisierung von der Parvulin-Domäne von PpiD wurde mittels NMR-Spektroskopie von Ulrich Weininger durchgeführt. Alle anderen Ergebnisse wurden von mir erarbeitet. Die Publikation wurde von Franz X. Schmid, Ulrich Weininger und mir formuliert.

H. Alle Ergebnisse wurden von Ulrich Weininger erarbeitet. Alle in der Arbeit verwendeten Proteine wurden von mir hergestellt. Die Publikation wurde von Franz X. Schmid, Jochen Balbach, Ulrich Weininger und mir verfasst.

I. Die NMR-spektroskopischen Untersuchungen wurden von Ulrich Weininger durchgeführt. Anne-Juliane Geitner hat im Rahmen ihrer Diplomarbeit eine Variante des Gen-3-Proteins des Phagen IKe biophysikalisch charakterisiert. Die Publikation wurde von Franz X. Schmid, Ulrich Weininger und mir formuliert.

7 Teilarbeiten

Teilarbeit A:

Kather, I., **Jakob, R.P.**, Dobbek, H., Schmid, F.X. (2008)
Changing the determinants of protein stability from covalent to non-covalent interactions by in vitro evolution: a structural and energetic analysis. *J Mol Biol*. **381**. 1040-54.

doi:10.1016/j.jmb.2008.06.073 J. Mol. Biol. (2008) 381, 1040–1054

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Changing the Determinants of Protein Stability from Covalent to Non-Covalent Interactions by *In Vitro* Evolution: A Structural and Energetic Analysis

Insa Kather[1,2]†, Roman Jakob[1,2]†, Holger Dobbek[3] and Franz X. Schmid[1,2]*

[1]*Laboratorium für Biochemie, Universitat Bayreuth, 95440 Bayreuth, Germany*

[2]*Bayreuther Zentrum für Molekulare Biowissenschaften, Universitat Bayreuth, 95440 Bayreuth, Germany*

[3]*Laboratorium Proteinkristallographie, Universität Bayreuth, 95440 Bayreuth, Germany*

Received 9 May 2008; accepted 11 June 2008
Available online
2 July 2008

Edited by C. R. Matthews

The three disulfide bonds of the gene-3-protein of the phage fd are essential for the conformational stability of this protein, and it unfolds when they are removed by reduction or mutation. Previously, we used an iterative *in vitro* selection strategy to generate a stable and functional form of the gene-3-protein without these disulfides. It yielded optimal replacements for the disulfide bonds as well as several stabilizing second-site mutations. The best selected variant showed a higher thermal stability compared with the disulfide-bonded wild-type protein. Here, we investigated the molecular basis of this strong stabilization by solving the crystal structure of this variant and by analyzing the contributions to the conformational stability of the selected mutations individually. They could mostly be explained by improved side-chain packing. The R29W substitution alone increased the midpoint of the thermal unfolding transition by 14 deg and the conformational stability by about 25 kJ mol^{-1}. This key mutation (i) removed a charged side chain that forms a buried salt bridge in the disulfide-containing wild-type protein, (ii) optimized the local packing with the residues that replace the C46–C53 disulfide and (iii) improved the domain interactions. Apparently, certain residues in proteins indeed play key roles for stability.

Keywords: protein stabilization; disulfide bonds; evolutionary protein design; phage gene-3-protein; phage display

Introduction

Disulfide bonding provides a simple and efficient means to convey proteins with a high conformational stability. In fact, many secreted proteins are maintained in the folded state only by their disulfide bonds and unfold when they are cleaved. The covalent cross-links provided by the disulfide bonds are robust, unlike non-covalent interactions that are weak and easily destabilized by changes in the solvent or by an increase in temperature. Attempts to substitute disulfide bonds by pairs of other amino acids revealed that even optimal non-covalent interactions could compensate only for a small amount of the strong decrease in stability that originates from the loss of the covalent cross-links.[1,2]

In our previous work, we used an *in vitro* selection strategy to convert a protein with three disulfide bonds into a stable disulfide-free variant that was much more stable than the original disulfide-bonded wild-type protein.[3] This protein, the N1–N2 two-domain fragment of the gene-3-protein (G3P) of the filamentous phage fd, is essential for the infection of *Escherichia coli* cells, and, in phage display, it is commonly used as a fusion partner to present target proteins at the phage surface.[4] G3P consists of three domains, N1 (68 aa), N2 (131 aa) and the carboxyterminal domain CT (150 aa), which

*Corresponding author. E-mail address: fx.schmid@uni-bayreuth.de.

† I.K. and R.J. contributed equally to this work.

Abbreviations used: G3P, gene-3-protein; N1, N2, the two N-terminal domains of G3P; G3P*, a fragment of G3P that consists of domains N1 and N2 and contains the four stabilizing mutations T13I, T101I, Q129H and D209Y; 3SS, 0SS, prefixes denoting G3P variants with and without the disulfide bonds, respectively; 0SS-G3P*-stab, the most stable disulfide-free variant; GdmCl, guanidinium chloride; DSC, differential scanning calorimetry.

0022-2836/$ - see front matter

are connected by glycine-rich linkers.[5–7] CT anchors G3P in the phage coat; N1 and N2 form a bilobal structure (Fig. 1) that protrudes from the phage tip.[8,10] It mediates infection by binding first to the tip of an F pilus of *E. coli*[11,12] and then to TolA at the cell surface.[6,7,13] N1 contains two disulfide bonds, whereas N2 contains a single disulfide; the protein unfolds when these disulfides are broken by reduction or mutagenesis.

N1 and N2 must be folded to initiate the infection process. In a series of proteolytic *in vitro* selections, we had used the correlation between the conformational stability and the protease resistance of G3P[14–16] to find optimal replacements for the individual disulfide bonds in N1 and N2 and, in addition, to identify stabilizing mutations at other non-cysteine positions.[3] For these selections, libraries in which either individual cysteine pairs had been randomized or mutations had been introduced in the entire gene for N1–N2 by error-prone PCR were employed. By this iterative approach, variants of N1–N2 that were devoid of all three disulfide bonds but stable and functional could be evolved. The best variant contained 14 second-site mutations (Fig. 1), and it showed a much higher conformational stability compared with the disulfide-containing wild-type protein. Despite the loss of all three disulfide bonds, the midpoints of the thermal transitions were increased from 48.5 to 67.0 °C for the N2 domain and from 60.0 to 78.7 °C for the N1 domain. The major decrease in conformational stability caused by the removal of the disulfides was thus overcompensated by strongly improved non-covalent interactions.

Here, we investigated how the selected mutations increase the stability of the disulfide-free N1–N2 variant. To this end, we determined their individual contributions to stability and solved the crystal structure of the most stable disulfide-free variant of N1–N2.

Results and Discussion

The most stable disulfide-free variant from the *Proside* selections contained 14 stabilizing mutations at non-cysteine positions (Fig. 1).[3] To elucidate how they compensate for the destabilization caused by the loss of the three disulfide bonds, we introduced them individually into a reference protein without disulfides. The disulfide-free form of wild-type G3P (0SS-G3P) could not serve as such a reference because it is unstable and forms aggregates. Instead, we used 0SS-G3P*, a variant that is marginally stable without the disulfide bonds. It already contains four stabilizing mutations (T13I in N1, Q129H in N2 and T101I as well as D209Y in the hinge between the two domains) that had been identified in an initial *Proside* selection with the disulfide-containing protein.[17] In 0SS-G3P*, the two disulfide bonds in N1 are replaced by the mutations C7S/C36I and C46I/C53V, and the disulfide bond in N2 is replaced by the C188V/C201A mutation (Fig. 1).

When the disulfide bonds are present, the protein (3SS-G3P*) unfolds in two stages (Fig. 2). The first transition at 61.8 °C reflects domain disassembly coupled with the unfolding of N2, whereas the second transition at 66.1 °C reflects the unfolding of N1. The first transition during thermal unfolding is largely reversible; heat-induced aggregation occurs, however, when the protein variants are kept in the thermally unfolded state for more than 10 min. In

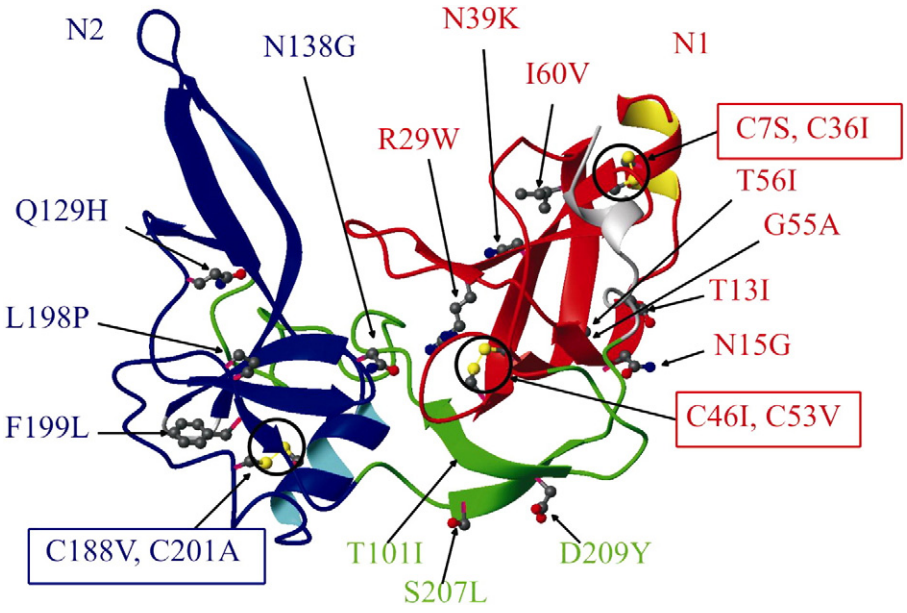


Fig. 1. Tertiary structure of G3P* (coordinates from Holliger *et al.*[8]). Domain N1 is shown in red; N2, in blue; and the hinge subdomain, in green. At the positions where mutations occurred, the side chains are shown in ball-and-stick representation and the selected stabilizing substitutions are indicated. The figure was prepared by using MOLMOL.[9]

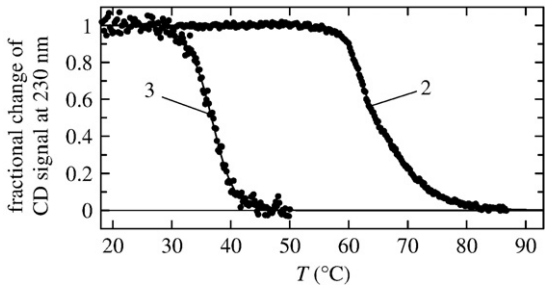


Fig. 2. Thermal transitions of 3SS-G3P* (2) and 0SS-G3P* (3) measured by CD at 230 nm. The fractional changes of the CD signal as obtained after a two-state analysis (3) or a three-state analysis (2) of the data are shown as a function of temperature. The transitions were measured with 4 μM protein in 100 mM potassium phosphate, pH 7.0, at a pathlength of 10 mm. The T_M and ΔH_D values are given in Table 1. Identical numbering is used in the figures and tables.

the absence of the disulfide bonds, 0SS-G3P* is only marginally stable and unfolds in a single cooperative transition with a midpoint at 36.7 °C (Fig. 2). This indicates that the domain dissociation is coupled with the unfolding of both domains, which are apparently unstable without the disulfide bonds and in the absence of the inter-domain interactions. The removal of the three disulfide bonds from G3P* thus led to a huge loss of stability, even though the Cys residues were replaced by optimized pairs of other residues.

Stabilizing mutations in domain N1

Two of the three disulfides of G3P* are in the N1 domain. After their removal, N1 becomes the least stable domain and limits the low overall stability of 0SS-G3P*. This explains why most of the stabilizing mutations that were found in the rounds of random mutagenesis and *in vitro* selection mapped to N1 (Fig. 1). It also explains why these mutations in N1 increased the stability of the entire protein.

In the structure of the N1–N2 fragment, the seven stabilizing mutations in N1 are labeled in red in Fig. 1. To determine their individual contributions to stability, we introduced them as single mutations into 0SS-G3P* and measured the thermal unfolding transitions of the corresponding variants (Fig. 3a). All the individual mutations are stabilizing, but unlike the transition of the disulfide-bonded wild-type protein (Fig. 2), their transitions appear to be monophasic. This indicates that for all these variants, the disassembly of the domains and the unfolding of the individual domains are thermodynamically coupled. The R29W mutation led to a particularly strong increase (by 14 °C) of the transition midpoint. The other single mutations in N1 increased T_M by 1–7 °C.

We calculated ΔG_D values under the assumption that the thermal transitions of 0SS-G3P* and its single mutants are cooperative two-state reactions. The extent of the coupling between domain dissociation and domain unfolding might differ for the individual mutants, as reflected in the variations of the ΔH_D values in Table 1. Therefore, such a two-state analysis is too simple, and the resulting ΔG_D values must not be overinterpreted. To minimize the errors introduced by this uncertainty, we calculated the ΔG_D values for these variants at 40 °C to avoid long extrapolations (Table 1). These calculations indicate that the R29W mutation stabilizes by about 24 kJ·mol^{-1}. Strong stabilizations (10–12 kJ·mol^{-1}) are also provided by the N15G and N39K mutations.

Stabilizing mutations in domain N2 and the hinge

In the wild-type form of G3P (3SS-G3P), N2 is the less stable domain, and above 40 °C, it is kept folded only by the inter-domain interactions with N1.[18] In our initial selections with the disulfide-bonded protein (3SS-G3P),[17] three of four stabilizing mutations thus mapped to N2 (Q129H) and to the hinge region (T101I, D209Y), which establishes most of the interactions with the N1 domain. In the selections, in the absence of the disulfide bond in N2 (C188–C201), three additional stabilizing mutations were found in N2 (N138G, L198P and F199L) and one was found in the hinge region (S207L). The corresponding single mutations in the reference protein 0SS-G3P* were all stabilizing (Fig. 3b) and increased the

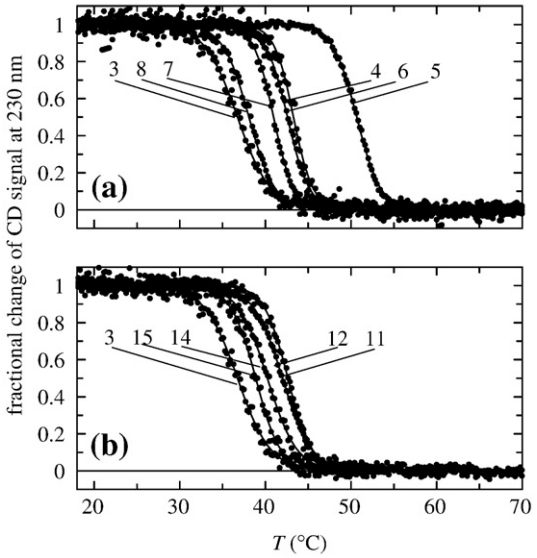


Fig. 3. Thermal transitions of the 0SS-G3P* variants. (a) 0SS-G3P* variants with single mutations in domain N1: 0SS-G3P* N15G (4), R29W (5), N39K (6), G55A (7) and T56I (8). (b) 0SS-G3P* variants with single mutations in domain N2: 0SS-G3P* N138G (11), L198P (12), F199L (14) and S207L (15). The transition of the reference protein 0SS-G3P* (3) is shown in both panels. The transitions were measured by CD at 230 nm as described in the legend to Fig. 2. The fractional change of the CD signal as obtained after a two-state analysis of the data is shown as a function of temperature. The T_M and ΔH_D values are given in Table 1. Identical numbering is used in the figures and tables.

Table 1. Stability data for wild-type G3P, 3SS-G3P*, and all 0SS-G3P* variants

No.	G3P* variant	T_M for N2		T_M for N1	ΔH_D for N2		ΔH_D for N1	ΔG_D for N2		ΔG_D for N1
1	3SS-wild type	48.5		60.0	621		213	14.2		−0.1
2	3SS-G3P*	61.8		66.1	640		228	33.8		4.0
3	0SS G3P*		36.7			480			−4.8	
4	N15G		42.6			649			5.3	
5	R29W		50.7			628			18.8	
6	N39K		43.3			743			7.7	
7	G55A		41.0			707			2.1	
8	T56I		38.1			534			−3.3	
9	I60V		38.9			539			−1.8	
10	P65H		37.3			568			−5.0	
11	N138G		42.6			598			4.9	
12	L198P		42.0			492			3.1	
13	L198V		40.5			553			0.9	
14	F199L		40.2			479			0.4	
15	S207L		39.5			560			−2.0	
16	R29W/N39K	53.2		63.0	557		129	19.8		1.2
17	R29W/N39K/T56I	53.5		67.1	762		120	28.7		2.5
18	R29W/N39K/T56I/L198P	57.6		68.4	790		104	37.3		2.5
19	R29W/T56I		50.3			644			18.8	
20	R29W/L198P		56.3			564			23.7	
21	R29W/N39K/F199L	56.4		64.7	689		197	30.1		2.7
22	N39K/T56I		41.9			656			4.0	
23	T56I/L198P		41.8			551			3.1	
24	R29W/N39K/G55A/I60V/F199L	58.1		70.4	607		203	28.2		6.0
25	N15G/R29W/N39K/G55A/F199L	60.1		75.2	670		167	34.1		7.0
26	N15G/R29W/N39K/T56I/N138G/L198P/F199L/S207L	66.0		75.5	777		212	49.3		9.1
27	N15G/R29W/N39K/G55A/T56I/I60V/N138G/L198P/F199L/S207L	67.0		78.7	852		135	56.6		6.7

The melting temperatures (T_M) are given in degrees Celsius, and the values for ΔG_D and ΔH_D are given in kilojoules per mole. The error ranges are ±0.5 deg for T_M and ±5% for ΔH_D. The ΔG_D values for the N2 domain refer to 40 °C; those for the N1 domain, to 60 °C. The errors in ΔG_D depend on the extent of extrapolation from the T_M to 40 or 60 °C. The T_M values for the N2 and N1 domains are the midpoints of the first transition and the second transition, respectively, of the biphasic unfolding transition as obtained after a three-state analysis. Variants 3–15, 19, 20, 22 and 23 show monophasic transitions, and their apparent T_M, ΔH_D and ΔG_D values are given. The transitions were measured as described in Figs. 2 and 3. The data were analyzed by non-linear regression with a constant heat capacity change ΔC_p of 10,000 J mol^{-1} K^{-1} for the first transition and that of 1000 J mol^{-1} K^{-1} for the second transition.[17] Identical numbering is used in the figures and tables. The values for 3SS-G3P* (variant 2) were taken from Ref. 17.

T_M values between 3 and 6 deg, equivalent to increases in ΔG_D between 5.2 and 9.7 kJ·mol^{-1} (Table 1). N138G is the most strongly stabilizing mutation, followed by L198P and F199L.

The four mutations were also introduced in a cumulative fashion into the isolated N2 domain with the C188–C201 disulfide bond intact (Table 2). The domain without the disulfide bond could not be used as a reference protein because it is unstable in isolation. The L198P mutation stabilized N2 to a similar extent as it stabilized 0SS-G3P*. The stabilization by N138G was smaller, and F199L was slightly destabilizing for the disulfide-containing isolated N2 domain. Possible explanations for these differences are subsequently discussed in the light of the crystal structure of the stabilized mutant.

Energetic interactions between the mutation sites

To search for energetic interactions between the mutated residues and to relate the single mutants with the combinations found in the selections, we produced a series of variants in which two or more mutations were combined. The flow diagram in Fig. 4 relates the single and multiple mutants. The corresponding stability data are included in Table 1.

We started by combining the two most stabilizing mutations in N1, R29W and N39K. This double mutant (variant 16; Table 1) was much more stable than the single mutants, and the two T_M values of its biphasic unfolding transition were both higher than the composite T_M values of the respective single mutants (variants 5 and 6; Table 1). The N1 domain remained folded after the domain dissociation (at 53.2 °C) and showed a separate thermal transition with a T_M of 63.0 °C (Table 1). The isolated 0SS-N1 domain with the R29W and N39K mutations shows the same T_M (62.6 °C; Table 2). Isolated 0SS-N1 without these mutations aggregated during purification, probably because it is unfolded.

The T56I mutation generally had small effects on stability. It was slightly stabilizing as a single mutation (variant 8; Table 1) or when combined with R29W/N39K (variant 17), but it was neutral or slightly destabilizing in combination with the single mutations R29W (variant 19) and N39K (variant 22).

L198P is located in the N2 domain, and, as a single mutation, it increased ΔG_D by about 7.9 kJ mol^{-1} (variant 12). When combined with mutations in N1 in several arrangements, similar stabilizations were observed (compare variants 17 and 18, variants 20 and 5 and variants 8 and 23; Table 1). This suggests

Table 2. Stability data for the isolated domains of G3P

Domain	T_M (°C)	ΔH_D (kJ mol^{-1})	ΔG_D (kJ mol^{-1})
N1 with disulfide bonds			
Wild type	62.4	241	1.8
G3P* (T13I)	67.9	235	5.4
N1 without disulfide bonds			
R29W	33.7	87	−8.5
R29W/N39K	62.6	142	1.1
R29W/N39K/G55A	67.9	172	3.9
N15G/R29W/N39K/G55A	77.5	188	8.9
N15G/R29W/N39K/G55A/ T56I/I60V	79.0	173	8.8
N2 with disulfide bonds			
Wild type	34.6	303	−5.7
G3P* (Q129H)	38.8	317	−1.2
Q129H/N138G	41.9	302	1.8
Q129H/N138G/F199L	40.0	314	0
Q129H/N138G/L198P	49.0	321	8.2
Q129H/N138G/L198P/F199L	47.7	354	7.9

For all proteins, the melting temperatures (T_M), the van't Hoff enthalpies of denaturation at T_M (ΔH_D) and the Gibbs free energies of denaturation (ΔG_D) are given. The error ranges are ±0.5 deg for T_M and ±5% for ΔH_D. As in Table 1, the ΔG_D values for the N2 domain refer to 40 °C; those for the N1 domain, to 60 °C. The errors in ΔG_D depend on the extent of extrapolation from the T_M to 40 or 60 °C. Thermal transitions were measured by CD at 230 nm (N1) or 222 nm (N2), with 4 μM protein in 100 mM potassium phosphate, pH 7.0, at a pathlength of 10 mm. The values for G3P* (T13I) were taken from Ref. 17.

that there is no interaction across the domain interface between L198P in N2 and the stabilizing mutations in N1. A similar result was obtained when the F199L mutation in N2 was combined with the R29W/N39K mutations in N1 (variants 16 and 21).

The variant with the R29W and N39K mutations in N1 and the F199L mutation in N2 was further combined with either the G55A/I60V (variant 24) or the N15G/G55A (variant 25) double mutation in N1, which had been found in our previous selections. The inclusion of the G55A/I60V and N15G/G55A mutations increased the T_M value by 5.7 and 10.5 deg, respectively (Table 1). These mutations were also introduced into the isolated N1 domain, and similar results were obtained (Table 2). Together, these analyses suggest that the contributions of the individual mutations are largely additive. This remains a semi-quantitative statement. As outlined above, the calculated ΔG_D values rest on the assumption that the thermal unfolding transitions of our variants are single two-state reactions or the sum of two two-state reactions.

The combination of all selected mutations (six in N1 and four in N2) resulted in variant 27 (Table 1). Its two transitions show midpoints at 67 and 78.7 °C, and its ΔG_D value is increased by about 70 kJ mol^{-1} relative to the 0SS-G3P* variant.

Unfolding measured by different probes

The unfolding transitions of the 0SS-G3P* variants were routinely measured by CD at 230 nm. At this wavelength, the first transition (domain disassembly coupled with N2 unfolding) is accompanied by a small decrease in signal; the second transition (unfolding of N1), by a large change. To examine the robustness of the T_M values derived from these curves, we employed additional probes to follow the thermal unfolding of several variants. Representative data for variant 25 (with the mutations N15G, R29W, N39K, G55A and F199L) are shown in Fig. 5.

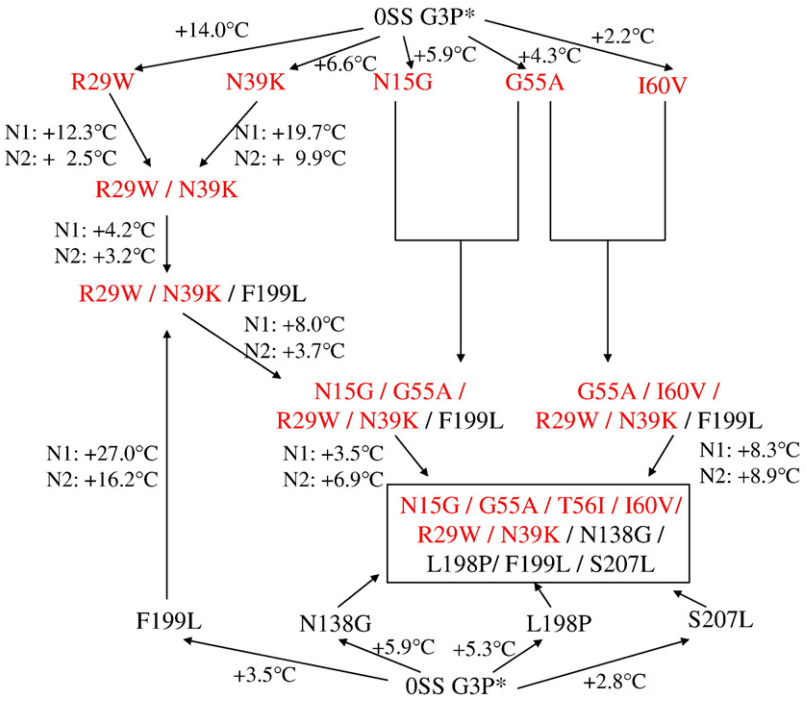


Fig. 4. Scheme for the analysis of the stabilizing effects of the selected second-site mutations in the N1 and N2 domains of 0SS-G3P*. The single mutations in domain N1 are shown at the top (in red); those in domain N2, at the bottom (in black). The 0SS-G3P* variant with all stabilizing mutations (0SS-G3P*-stab) is framed. The increases of the T_M values (ΔT_M) upon the successive introduction of the mutations into 0SS-G3P* are also given. For some variants, the transitions of N1 and N2 almost coincide, and only apparent T_M (and ΔT_M) values are given. They represent the temperatures at which half of the CD change has occurred. For 0SS-G3P* variants with biphasic unfolding transitions, ΔT_M values are given for both domains.

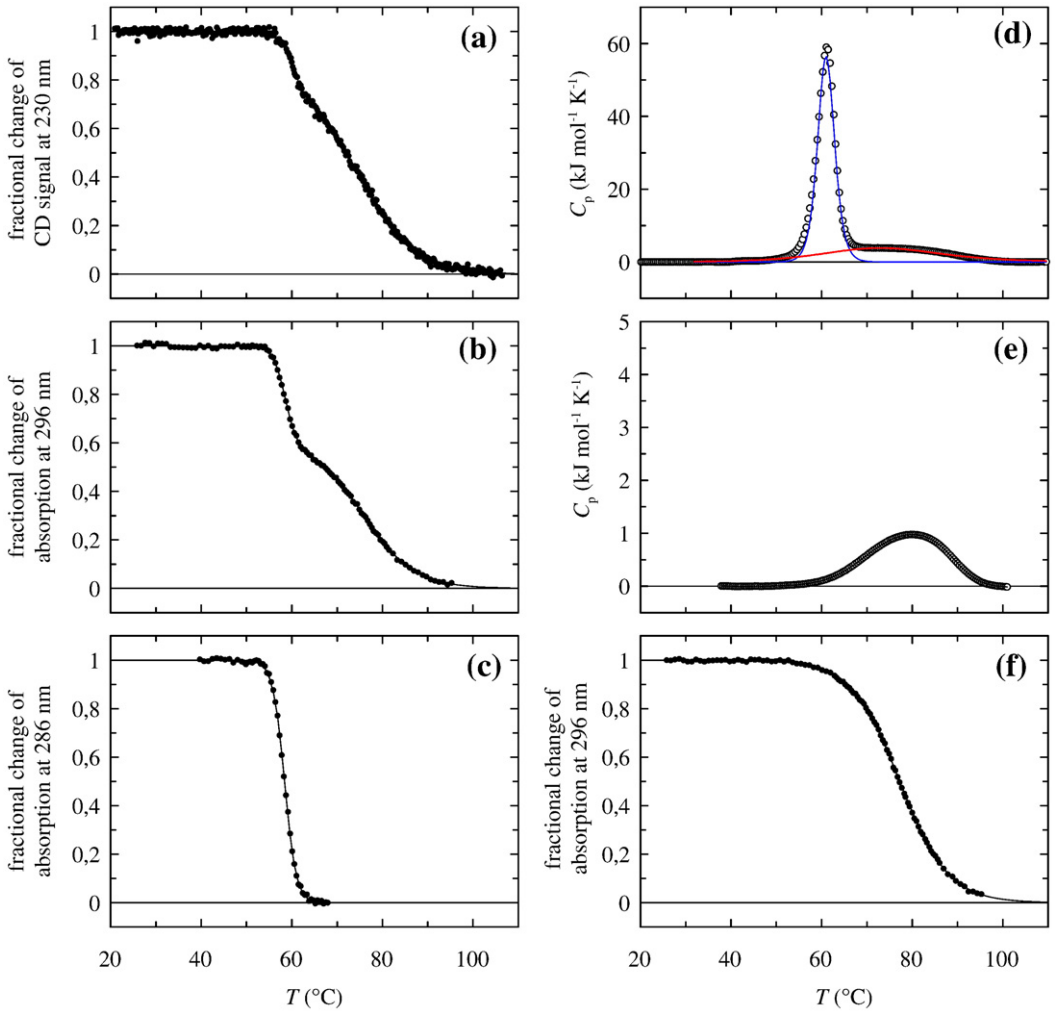


Fig. 5. Thermal unfolding of variant 25 (0SS-G3P* N15G/R29W/N39K/G55A/F199L) followed by different probes. (a) CD at 230 nm. (b) Absorption at 296 nm. (c) Absorption at 286 nm. (f) Unfolding of the isolated N1 domain of variant 25 by absorption at 296 nm. In (a), (b), (c) and (f), the fractional changes in signal are shown as a function of temperature. The protein concentrations were 4 μM (CD) or 10 μM (absorbance) in 100 mM potassium phosphate, pH 7.0. The continuous lines in (a)–(c) and (f) represent the fit according to a three-state model (a and b) and to a two-state model (c and f). (d and e) DSC measurements. The heat capacity changes during unfolding of (d) the N1–N2 fragment (50 μM) and (e) the isolated N1 domain (183 μM) in 100 mM potassium phosphate, pH 7.0, were measured as a function of temperature at a heating rate of 90 K/h. The continuous line represents the result of a Levenberg–Marquardt non-two-state least-squares analysis. The red and blue lines represent the deconvolution into two independent transitions of domains N1 and N2, respectively. The thermodynamic data are shown in Table 3.

The data derived from the analysis are summarized in Table 3.

CD at 230 nm (Fig. 5a) and Trp absorbance at 296 nm (Fig. 5b) follow both unfolding reactions, albeit with different relative amplitudes. The change in Tyr absorbance at 286 nm (Fig. 5c) monitors the first transition, because most of the Tyr residues are located in the N2 domain. Unfolding was also measured by differential scanning calorimetry (DSC; Fig. 5d). Here, the first transition gives a dominant peak because it contains the enthalpies of both domain disassembly and N2 unfolding. The second peak is small and broad because N1 is small. The contributions of N1 to the changes in Trp absorbance and to the enthalpy of unfolding were also measured in experiments with the isolated N1 domain of this variant (Fig. 5e and f). The comparison shows that the decomposition of the experimental unfolding curves into two transitions is rather robust, yielding similar values for the T_M and ΔH_D parameters. It also gives an estimate of the accuracy of these values. Multiparameter unfolding measurements were also performed with variants 1, 2, 26 and 27. They gave similar values and confirmed the results from the CD measurements in Table 1.

The key contribution of the R29W mutation

The replacement of Arg29 by Trp in the N1 domain provides a very strong contribution to stability (Table 1). Arg29 is in the N1 domain, and, in the disulfide-containing wild-type protein, it is close to the C46–C53 disulfide bond and to the

Table 3. Stability data measured by different probes for variant 25

Probe	Unfolding of N2		Unfolding of N1		Unfolding of isolated N1	
	T_M2 (°C)	ΔH_D2 (kJ mol^{-1})	T_M1 (°C)	ΔH_D1 (kJ mol^{-1})	T_M1 (°C)	ΔH_D1 (kJ mol^{-1})
CD at 230 nm	60.1	670	75.2	167	77.5	188
Trp absorption at 296 nm	58.6	681	75.8	181	77.3	191
Tyr absorption at 286 nm	58.4	707				
Calorimetry	61.0	674	77.0	141	79.2	194

The melting temperatures (T_M) and the enthalpies of unfolding at T_M (ΔH_D) are given for the N2 (label "2") and N1 (label "1") domains. They represent the midpoints of the first transition and the second transition, respectively, of the biphasic unfolding transition as obtained after a three-state analysis. The T_M1 value of the isolated N1 domain reflects the midpoint of the monophasic transition of this domain. The transitions were measured as described in the legends to Figs. 2 and 3 (CD measurements) and Fig. 5 (absorption and DSC measurements). The data were analyzed by non-linear regression with a fixed heat capacity change ΔC_p of 10,000 J mol^{-1} K^{-1} for the first transition and that of 1000 J mol^{-1} K^{-1} for the second transition.

domain interface (Fig. 1). The R29W mutation might thus increase the intrinsic stability of the N1 domain, compensate locally for the loss of the C46–C53 disulfide (which is replaced by the I46/V53 pair in 0SS-G3P*) and/or improve the domain interactions.

To gain further insight into the contributions of the R29W mutation, we also introduced it into the reference protein, 3SS-G3P*, and into the corresponding isolated N1 domain, both with intact disulfides. The N1 domain without the disulfide bonds could not be studied because it is unstable in isolation. All G3P* variants with intact disulfide bonds unfold in two consecutive transitions. In the first transition, the domains disassemble and N2 unfolds; in the second transition, N1 unfolds separately at elevated temperature. Both transitions are shifted to higher temperature by the R29W mutation (Fig. 6a). The shift found for the second transition is also observed when the isolated N1 domain with the R29W mutation is heated (Fig. 6b).

The analysis (Table 4) shows that the extent of stabilization by the R29W mutation is highest when the disulfide bonds are absent, as in the selections. The stabilization by 27.3 kJ mol^{-1} is higher than the value given in Table 1 (23.6 kJ mol^{-1}) because it is calculated at 60 °C (to facilitate the comparison with the more stable 3SS variants in Table 4) and because the apparent ΔH_D values are different. The 3SS protein is also stabilized, albeit by a smaller extent. Its N1 domain gains 6.5 kJ mol^{-1} in stability, and, by improving the inter-domain interactions, it stabilizes N2 by 5.8 kJ mol^{-1}. The sum of these contributions (12.3 kJ mol^{-1}) is significantly smaller than the stabilization observed in the absence of the disulfide bonds (Table 4).

The stability data for the four variants with and without the disulfide bonds containing either Arg29 or Trp29 can be arranged in a pseudo double-mutant cycle (Fig. 7). In this cycle, the horizontal reactions reflect the effect of the R29W mutation and the

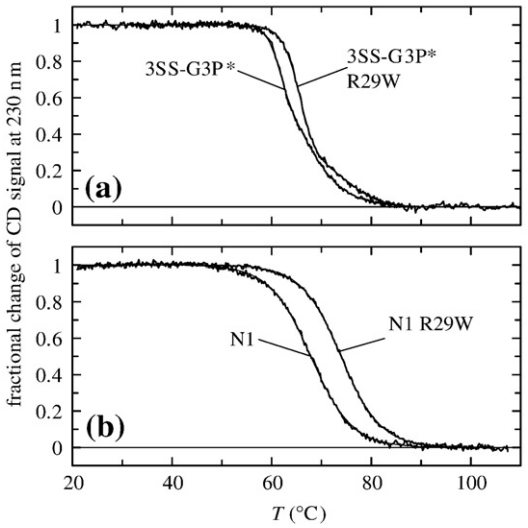


Fig. 6. Effect of the R29W mutation on the thermal unfolding of (a) 3SS-G3P* and (b) its isolated N1 domain. The transitions were measured by CD at 230 nm with 4 μM protein in 100 mM potassium phosphate, pH 7.0, at a pathlength of 10 mm. The fractional changes of the CD signal as obtained after (a) a three-state analysis or (b) a two-state analysis of the data are shown as a function of temperature. The parameters from the analysis are listed in Table 4.

Table 4. Stabilization of 0SS-G3P* and 3SS-G3P* by the R29W mutation

G3P* variant	T_M (°C)	ΔH_D (kJ mol^{-1})	ΔG_D (60 °C) (kJ mol^{-1})	$\Delta\Delta G$ R29W (kJ mol^{-1})
0SS-G3P*				
N1–N2	36.7	480	−45.4	
0SS-G3P* R29W				
N1–N2	50.7	628	−18.1	+27.3
Isolated N1	33.7	87	−8.5	–
3SS-G3P*				
N2	61.8	648	+2.9	
N1	68.6	336	+8.4	
Isolated N1	67.9	235	+5.4	
3SS-G3P* R29W				
N2	65.6	549	8.7	+5.8
N1	75.9	335	14.9	+6.5
Isolated N1	73.9	251	9.8	+4.4

For all protein variants, the melting temperatures (T_M), the enthalpy of unfolding ΔH_D at T_M and the Gibbs free energy of unfolding at 60 °C ΔG_D (60 °C) are given. For the 0SS-G3P* variants, N1 and N2 unfold in a single cooperative transition. For the 3SS-G3P*, separate transitions for N2 and N1 are observed. The transitions were measured as described in the legends to Figs. 2 and 3. The data were analyzed by non-linear regression with a fixed heat capacity change ΔC_p of 10,000 J mol^{-1} K^{-1} for the transition of N2 and that of 1000 J mol^{-1} K^{-1} for the transition of N1.

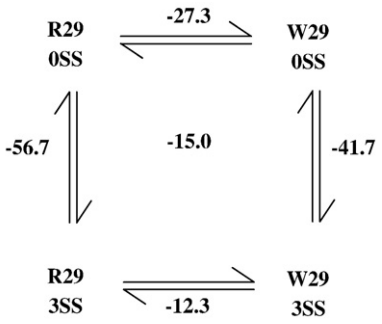


Fig. 7. Thermodynamic cycle for the linkage between the R29W mutation and the replacement of the disulfide bonds. The Gibbs free energies of folding (ΔG_N in kJ mol^{-1}) are given.

vertical reactions reflect the introduction of the three disulfide bonds. The calculated coupling energy of 15 kJ mol^{-1} indicates that the R29W mutation in fact compensates significantly for the huge loss in stability that is caused by the removal of the disulfides from G3P*. The effects of the R29W mutation are subsequently discussed further in the light of the structures of the wild type and the mutated protein.

The structure of the most stable variant, 0SS-G3P*-stab

The crystal structure of the N1–N2 fragment of the most strongly stabilized disulfide-free G3P (variant 27 in Table 1; denoted in the following as 0SS-G3P*-stab) was solved at 1.9-Å resolution. Details of its crystallization and structure determination are summarized in Table 5. The asymmetric unit of the crystal consists of two molecules that differ slightly in the orientation of the two domains relative to each other (Fig. 8a). Such a difference was also found for wild-type G3P with intact disulfide bonds.[8] This indicates that the linker region between the domains shows some flexibility. The structures of N1 and N2 of the wild-type protein and the 0SS-G3P*-stab variant were therefore aligned individually. The backbone atoms of the N1 domain superimposed with an RMS of 0.4 Å (Fig. 8b); those of the N2 domain, with an RMS of 0.6 Å (Fig. 8c).

In the structure of 0SS-G3P*-stab, the amino-terminal residues 1–9 are not resolved (Fig. 8b). In 3SS-G3P, they form a short helix that is connected to strand β4 by the C7–C36 disulfide bond. This covalent linkage is disrupted by the C7S/C36I mutations in 0SS-G3P*-stab, and this, apparently, leads to local unfolding.

Table 5. Statistics on diffraction data and structure refinement

	0SS-G3P* max
Total/Unique reflections	197,032/40,287
R_s[a]	0.095 (0.441)
Resolution (Å)	20–1.90 (2.0–1.9)
Completeness (%)	99.5 (96.5)
$(I)/(\sigma I)$	13.8 (3.8)
Model R/R_{free} factor[b] (%)	21.6/26.3
RMSD from ideal geometry	
Bonds (Å)	0.0086
Angles (°)	1.5
Cruickshank dispersion precision indicator for coordinate error	0.218
Ramachandran statistics (%)	
Most favored/Additional/Generously allowed/Disallowed regions	81.9/18.1/0.0/0.0

In the refinement statistics, Friedel mates were merged.
[a] $R_s = \sum_h \sum_i |I_i(h) - \langle I(h)\rangle| / \sum_h \sum_i I_i(h)$, where i is the independent observation of reflection h.
[b] The R_{free} factor was calculated from 5% of the data that were removed at random before the refinement was carried out.

Structural interpretation of the contributions of the individual mutations to stability

In the following, we analyze the consequences of the mutations in the context of the crystal structures

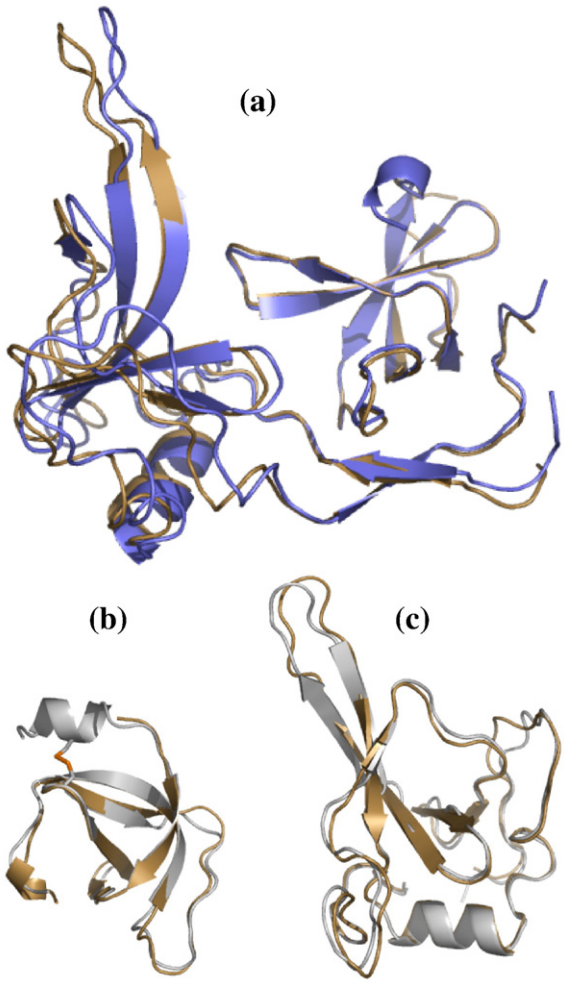


Fig. 8. Three-dimensional structure of 0SS-G3P*-stab. (a) Superposition of the structures of the 0SS-G3P*-stab molecules A (brown) and B (blue) in the asymmetric unit. (b and c) Superpositions of the structures of the (b) N1 and (c) N2 domains of 0SS-G3P*-stab (brown) and wild-type G3P (gray). For this comparison, the 1G3P structure of wild-type G3P was used.[10]

of wild-type G3P and the 0SS-G3P*-stab variant. In this discussion, we also include the four mutations T13I, T101I, Q129H and D209Y, which had been identified in our previous work[17] and are present in the reference variant for this study, 3SS-G3P*.

Position 29

The replacement of Arg29 by Trp is extremely stabilizing and increases ΔG_D by about 25 kJ mol^{-1}. In the wild-type protein (1G3P), Arg29 is located at a peculiar position in the contact region between the N1 and N2 domains and the hinge subdomain (Fig. 1). Its positively charged guanidino group is shielded from the solvent by Tyr102 in the hinge (Fig. 9a). Moreover, it is in close contact with the two sulfur atoms of the C46–C53 disulfide bond as well as with Glu50, thus forming a salt bridge that is buried at the interface between N1 and the hinge region. Glu50 is in contact with Asn138 of the N2 domain (Fig. 9a).

Trp29 in the stabilized 0SS variant is similarly oriented as Arg29 in the wild-type protein (Fig. 9b). It forms van der Waals contacts with the side chains of both Ile46 and Val53, which replace the C46–C53 disulfide bond. Trp29 also forms an edge-on aromatic interaction with Tyr102 in the hinge. The carboxyl group of Glu50, which, in the wild-type protein, formed the buried salt bridge with Arg29, is now exposed at the protein surface. It has changed its orientation such that it occupies the position held by the amide group of Asn138 in the wild-type protein (Fig. 9a). In the stabilized mutant, Asn138 is changed to Gly (Fig. 9b), and thus a clash between the side chains of Glu50 in N1 and Asn138 in N2 is avoided.

We suspect that the buried Arg29–Glu50 salt bridge at the domain interface decreases the stability of the wild-type protein and that the removal of this unfavorable situation by the R29W mutation contributes to the observed very strong stabilization by Trp29. In the wild-type protein, the buried salt bridge is close to the two S_γ atoms of the C46–C53 disulfide bond; thus, the loss in stability might be less severe. This would explain why the R29W mutation stabilizes the wild-type protein less strongly than the 0SS variants. The improvement of the domain interactions by the R29W mutation originates probably from two sources: the removal of the buried salt bridge from the domain interface and the formation of an aromatic interaction between Trp29 and Tyr102 across the domain interface.

The delicate balance of interactions in this part of the domain interface is also reflected by an alternative side-chain arrangement, which was observed in another crystal structure of the wild-type form of the G3P (2G3P).[8] Here, the ion pair between Arg29 and Glu50 is disrupted, Arg29 has moved away from the C46–C53 disulfide and, instead, the side chain of Tyr102 is rotated such that it now contacts this disulfide. Arg29 is still largely buried in this structure, but Glu50 forms an exposed salt bridge with another partner, Lys22.

Position 39

The N39K mutation is also strongly stabilizing and increases ΔG_D by 12.5 kJ mol^{-1} (Table 1). Position 39

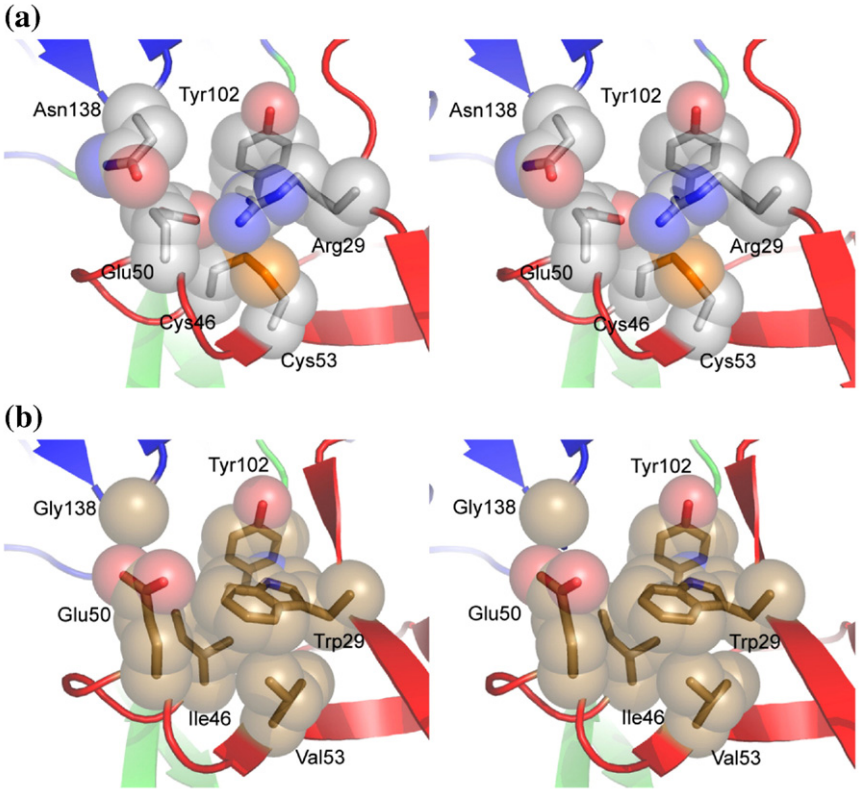


Fig. 9. Stereo representation of the region around residues 29, 46, 53 and 138 in (a) wild-type G3P* and (b) 0SS-G3P*-stab. For this comparison, the 1G3P structure of wild-type G3P was used.[10]

is located at the surface of domain N1 remote from the domain interface. In the wild-type protein, Asn39 is largely buried, but its side-chain amide is solvent accessible. There is no indication for unfavorable interactions of Asn39. In the stabilized variant, Lys39 is probably involved in a coulombic interaction with Asp28 (3.8 Å; Fig. 10a), which connects β-strands 3 and 4 of N1.

Positions 13, 15, 55 and 56

The backbone atoms of these four residues are close to one another in two neighboring antiparallel β-strands at the edge of the N1 domain (Fig. 1). Residues 15 and 55 form two amide hydrogen bonds with each other, but the side chains of the four residues are not in contact. Thr13 is a "boundary" residue that is exposed to the solvent with one face and contacts Phe17 and Trp38 in the protein core with the other face (Fig. 10b). The mutation to Ile improves the interactions with these aromatic residues and leads to an additional interaction with Val215 of the hinge. Probably, this explains the stabilization by the T13I mutation.

The T56I mutation leads to a slight increase in T_M (ΔT_M=1.4 °C), and in combination with other mutations, it is basically neutral (Table 1). Position 56 is at the binding site of N1 for the TolA C-terminal domain, which is the phage receptor at the surface of *E. coli*. T56I might have been selected because it improves the infectivity of the phage.

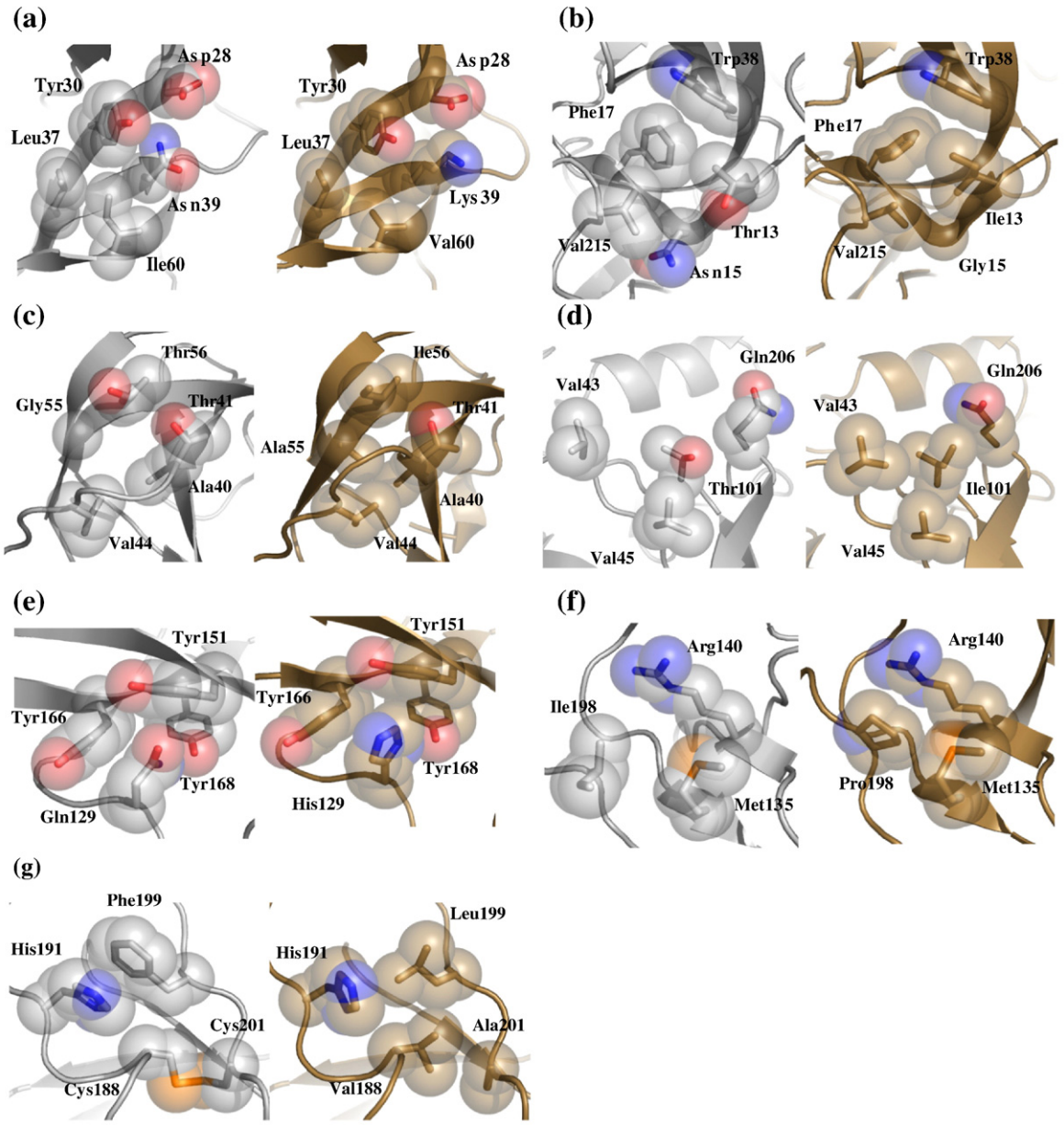


Fig. 10. Comparisons of the local structures of wild-type G3P (left figures; gray) and 0SS-G3P*-stab (right figures; brown) in the vicinities of the stabilizing mutations. For this comparison, the 1G3P structure of wild-type G3P was used.[10]

The N15G and G55A mutations in the N1 domain were both found in the final selection rounds, in the absence of all three disulfide bonds. Asn15 is an exposed residue, and in the structure of the wild-type protein (Fig. 10b), there is no indication for steric strain caused by the side chain of Asn15 that might have been relieved by a mutation to glycine. The dihedral angles φ and ψ at position 15 (φ = −169°, ψ = −176°) are in the allowed region for both Asn and Gly. After the mutation to Gly15, φ changes by about 20° to +168°; otherwise, the backbone angles are unchanged in the stabilized disulfide-free variant.

At position 55, the backbone angles are close to 180° and thus, in principle, more favorable for the wild-type residue Gly than for Ala. The methyl group of Ala55 points towards the interior and apparently fills a cavity in the protein (Fig. 10c). The G55A mutation certainly stabilizes G3P also by reducing the entropy of the unfolded state.

The structure comparison does not provide straightforward explanations for the strong stabilizations observed for both the G55A and N15G mutations (Table 1). We suspect that the mutations at positions 13, 15 and 55 improve the antiparallel β-sheet of the N1 domain. In the wild-type protein, this structure is stabilized by the C7–C36 disulfide, which links the protein N-terminus to the rest of the molecule. In the 0SS variant, this link is absent and, as a consequence, the N-terminus is frayed and invisible in structure up to residue 9 (Fig. 8b).

Position 60

The Ile-to-Val mutation at position 60 provides a small contribution of about 3 kJ mol^{-1} to stability, independent of other mutations (Table 1). The side chain of Val60 points to the surface of the N1 domain, and its two terminal methyl groups are partially exposed (Fig. 10a). They also engage in hydrophobic contacts with other side chains, in particular, with the C^{γ} and C^{δ} methylene groups of Lys39, another stabilizing residue from a previous selection round. Possibly, Val60 helps orient Lys39 for an optimal interaction with its salt-bridge partner, Asp28.

Position 101

The mutation T101I is one of the four stabilizing mutations that were found in our initial selection[17] and was present in the reference protein for this study (3SS-G3P*). It is situated at the end of β-strand 6 in the hinge subdomain (Fig. 1). In the stabilized variant, the side chain of Ile101 is partially exposed, and it makes hydrophobic contacts with Val43 and Val45 in the N1 domain (Fig. 10d). These interactions are further improved because, in the mutant, the side chain of Val43 is moved towards Ile101. Gln206 also changed positions such that its C^{β} and C^{γ} methylene groups contact Ile101 as well. The T101I mutation provides a further example for the observation that Thr-to-Ile mutations near protein surfaces are stabilizing.[19,20] The generic stabilizing effect of this replacement is related with the fact that Thr, but not Ile, side chains must be desolvated upon folding and that Thr positions usually accommodate other amino acids with β-branched side chains (e.g., Ile) well.

Position 129

Q129H is also one of the four stabilizing mutations that were present in the reference protein 3SS-G3P*.[17] The crystal structure of the 0SS-G3P*-stab variant indicates that this mutation stabilizes G3P because His129 interacts favorably with the three surrounding Tyr residues, Tyr151, Tyr166 and Tyr168 (Fig. 10e).

Position 138

N138G is the most strongly stabilizing mutation in the N2 domain and increases the stability of 0SS-G3P* by 9.7 kJ mol^{-1} (Table 1). Residues 136–139 form a type I′ β-turn in G3P, and Asn138 is at the third position of this turn. Asn residues are unfavorable at this position and occur in only 4 of 127 examined type I′ turns in folded proteins.[21] Gly, in contrast, is the optimal residue. In fact, more than 70% of the type I′ turns show a Gly at position 3. When introduced into the isolated N2 domain, N138G stabilizes by 3.0 kJ mol^{-1} (Table 2), which indicates that the turn optimization contributes about a third to the total stabilization by this mutation.

The remaining 6 kJ mol^{-1} originates apparently from changes in the interactions between the domains. As outlined above, in the stabilized variant (Fig. 9b), Glu50 has lost its salt-bridge partner, Arg29, and, as a consequence, the Glu50 side chain is rotated such that its γ-carboxylate is transferred to a solvent-accessible position, which, in the wild-type protein, is occupied by the side chain of Asn138 (Fig. 9a). The N138G mutation relieves this steric clash and thus improves the domain interactions.

Position 198

Position 198 is occupied by different residues in different phage G3Ps. In our phage strain, the wild-type residue is Leu, whereas in the two published crystal structures, position 198 is occupied by Ile (1G3P8) or Pro (2G3P9). In the *Proside* selections, the substitutions L198V and L198P were found.[3] Both are stabilizing (Table 1), with L198P being more efficient than L198V. The local structure around Pro198 in the stabilized disulfide-free protein (Fig. 10f) matches the corresponding region in the 2G3P structure, which also contains a Pro198. The backbone dihedral angles at position 198 are optimal for proline. The side chain of Pro198 is packed against the side chains of Arg140 and Met135, unlike the side chain of Ile198 (in the 1G3P structure), which is solvent exposed and thus presumably unfavorable for stability (Fig. 10f).

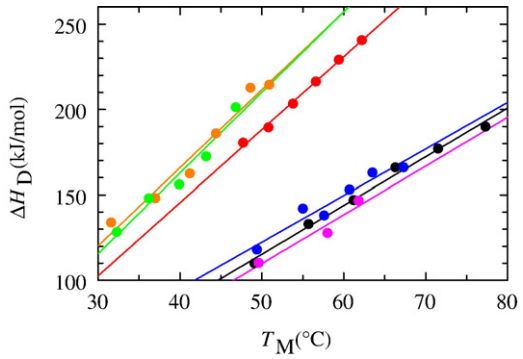


Fig. 11. Heat-induced unfolding of variants of the isolated N1 domain with two, one or zero disulfide bonds at different GdmCl concentrations. The correlations between the enthalpy of unfolding ΔH_D and the transition midpoint T_M are shown for 2SS-wild-type N1 (filled circle in red), 1SS-N1* C46G/C53I (filled circle in green), 1SS-N1* C46A/C53V (filled circle in orange), 0SS-N1* R29W/N39K (filled circle in pink), 0SS-N1* R29W/N39K/G55A/I60V (filled circle in blue) and 0SS-N1* R29W/N39K/N15G/G55A (filled circle in black). All N1* variants were derived from G3P* and contain the Q129H mutation. The protein concentration was 3 μM in 100 mM potassium phosphate, pH 7.0. The CD was measured at 230 nm (10-mm pathlength). The data were derived from two-state analyses of the unfolding transitions.

Pro198 stabilizes probably in a dual way: by decreasing the entropy of the unfolded state and by interacting favorably with Arg140 and Met135 in the folded state. Position 198 is remote from the domain interface, and, in fact, the stabilization observed for the L198P mutation in the stabilized 0SS-two-domain protein (Table 1) is similar to the stabilization by the same mutation in the isolated N2 domain (Table 2).

Position 199

The F199L mutation stabilizes 0SS-G3P* by about 5 kJ mol^{-1} (Table 1). In this protein, the disulfide-bonded residues Cys188 and Cys201 are replaced by the residue pair Val188 and Ala201. Leu199 is in direct contact with Val188 (Fig. 10g), and this suggests that Leu199 stabilizes the disulfide-free protein by favorable interactions with the residue that replaced Cys188. The stabilization by the F199L mutation is indeed specific for the disulfide-free protein. The N2 domain with the C188–C201 disulfide bond is slightly destabilized by the F199L mutation (Table 2).

Decreased cooperativity of unfolding of N1 in the absence of the C7–C36 disulfide bond

The thermal unfolding transitions of the 0SS variants of the N1 domain are less cooperative than those of the disulfide-bonded wild-type protein.[3] To investigate this further, we measured thermal unfolding transitions in the presence of guanidinium chloride (GdmCl) at increasing concentrations for wild-type N1 with both disulfide bonds, for two variants with only the C7–C36 disulfide and for three variants without disulfides. Figure 11 shows the van't Hoff enthalpies of unfolding of these variants as a function of the T_M value. Wild-type N1 with both disulfide bonds and the variants with only the C7–C36 disulfide show similar enthalpies of unfolding (ΔH_D) and an increase in heat capacity ΔC_p of about 4500 J mol^{-1} K^{-1}. For all 0SS variants, ΔH_D is lower and ΔC_p is reduced to about 2800 J mol^{-1} K^{-1} (Fig. 11). These indicate that the cooperativity of unfolding is decreased when the C7–C36 disulfide bond is opened. The ΔC_p values from Fig. 11 are slightly larger than the ΔC_p values for unfolding because the enthalpy of denaturant binding contributes to the apparent enthalpies of unfolding.[22] However, this affects all variants in an analogous fashion and cannot explain the observed differences.

We also measured GdmCl-induced unfolding transitions for several of the 0SS variants by Trp fluorescence, which is a selective probe for the unfolding of the N1 domain (Fig. 12). The cooperativities of these transitions are lowered as well, and the m values are decreased from 6.5 kJ mol^{-1} M^{-1} (for 2SS-N1) to about 5.3 kJ mol^{-1} M^{-1} (for the 0SS variants).

The crystal structure of 0SS-G3P*-stab provides an explanation for this decrease in the cooperativity of unfolding. In the absence of the C7–C36 disulfide bond, the amino-terminal helix of N1 is unfolded; thus, it does not contribute to the cooperative unit of unfolding.

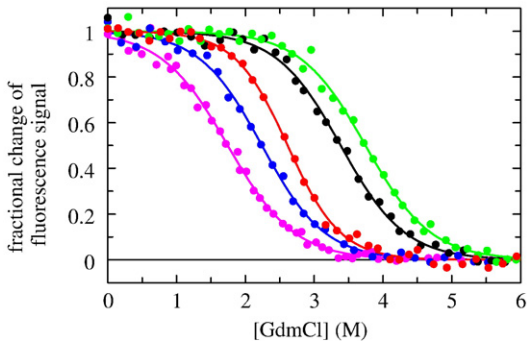


Fig. 12. GdmCl-induced unfolding transitions of 2SS-wild-type G3P* (filled circle in red; m = 6.5 kJ mol^{-1} M^{-1}) and the 0SS-G3P* variants R29W/N39K (filled circle in pink; m = 5.1 kJ mol^{-1} M^{-1}), R29W/N39K/G55A/I60V (filled circle in blue; m = 5.3 kJ mol^{-1} M^{-1}), R29W/N39K/N15G/G55A (filled circle in black; m = 5.4 kJ mol^{-1} M^{-1}) and N15G/R29W/N39K/G55A/T56I/I60V (filled circle in green; m = 5.1 kJ mol^{-1} M^{-1}). The protein concentration was 3 μM in 100 mM potassium phosphate, pH 7.0. Unfolding at 25 °C was followed by tryptophan fluorescence at 360 nm (after excitation at 280 nm), which is a selective probe for the N1 domain. The fractional change of the fluorescence signal as obtained after a two-state analysis is shown as a function of GdmCl concentration.

Conclusions

Disulfide bonds provide the major contribution to the stability of many extracellular proteins. For the phage G3P, this contribution amounts to roughly 40–60 kJ mol^{-1} (Fig. 7); thus, it is much larger than the net Gibbs free energy of stabilization of this protein. This implies that the non-covalent interactions alone would be much too weak to keep it in the folded state. Optimal replacements of cystines by pairs of other amino acids caused only modest increases in conformational stability, but several second-site mutations, as identified in an *in vitro* selection, led to variants that are much more stable than the disulfide-bonded wild-type protein.[3]

The stabilities of two-domain proteins are determined by the stabilities of the two individual domains and by the strength of the inter-domain interactions. How these three factors cooperate to establish the overall stability is revealed by the analysis of the G3P variants. In the wild-type protein, domain N1 (two disulfide bonds) is stable in isolation, but N2 (one disulfide bond) is maintained in the folded form only by the domain interactions and unfolds when they are broken in the course of thermal unfolding.

In the protein without the disulfides (0SS-G3P*), the individual domains are unstable and kept in the folded state only by the domain interactions. Both unfold upon domain disassembly, and unfolding is thus a single cooperative reaction. The introduction of multiple stabilizing mutations into the N1 domain re-stabilized this domain and restored the unfolding mechanism of the wild-type protein, in which N1 unfolds in a separate transition at high temperature.

Having both the crystal structure of the optimized protein and stability data for many single and multiple mutants allowed us to evaluate the molecular basis of stabilization. Only two mutations (R29W and F199L) stabilized 0SS-G3P by optimizing the structure around the residues that replaced the disulfide bonds. Most of the other mutations, however, provided independent contributions to ΔG_D at sites remote from the disulfide positions.

The R29W mutation alone increased the T_M value of the highly cooperative unfolding transition by 14 °C. It stabilized 0SS-G3P in three ways: (i) by removing a charged side chain that forms a buried salt bridge in the disulfide-containing wild-type protein, (ii) by optimizing the local packing with the residues that replace the C46–C53 disulfide and (iii) by improving the domain interactions. These highlight that particular positions in proteins can play key roles for the stability and that such positions can be identified by *in vitro* selection.

One mutation led to a new salt bridge at the protein surface and another one optimized a turn. In most other cases, the stabilizing effects seem to originate from improved side-chain packing. We are aware that this analysis is incomplete because it does not account for potential changes in the unfolded protein.[23–26] It agrees well, however, with previous structural analyses of stabilized protein variants that emerged from *in vitro* selections,[27,28] and it supports models that assume that improved side-chain packing is an important factor for the stability of proteins[29] and, in particular, of those from thermophilic organisms.[30,31] Such differences in packing are presumably local effects, because, overall, proteins from mesophilic and those from thermophilic organisms do not differ in density.[32]

Libraries of proteins that are produced by error-prone PCR and presented on a filamentous phage represent only tiny fractions of the sequence spaces of these proteins. Finding many stabilizing variants with mutations at or near the protein surface in such libraries is encouraging. It indicates that there are many solutions to the problem of stabilizing proteins by *in vitro* evolution or by computational design. It also validates the classic concept of protein stability from the times prior to directed mutagenesis and protein engineering. It assumed that the conformational stability of a protein is determined by a balance between large numbers of stabilizing and destabilizing contributions.[33] Results from protein engineering and protein *in vitro* evolution[15,17,19,20,27,28,34–41] now indicate that this balance can be strongly shifted towards increased stability by a few amino acid replacements, in particular, at the protein surface.

Materials and Methods

Expression and purification of the G3P variants

G3P* variants without disulfide bonds were amplified as fragments coding for residues 1–217 of the mature G3P, followed by ProSerGly$(His)_6$, from the corresponding double-stranded phage DNA with the primers 5′-GGAATTCCATATGGCTGAAACTGTTGAAAGT-3′ and 5′-CGGGATCCTTAATGATGATGATGATGATGCCCG-GACGGAGCATTGACAGGAGG-3′ and cloned into the expression plasmid pET11a (Novagen, Madison, Wisconsin, USA) *via* their NdeI and BamHI restriction sites. The proteins were overexpressed in *E. coli* BL21 (DE3) pLysS (Stratagene, La Jolla, CA, USA), isolated in soluble form from the cytosol and purified as described for refolded disulfide-containing G3P*.[17]

For the expression of the isolated N1 domain [residues 1–67 of mature G3P, followed by Ala$(His)_6$], the corresponding gene fragments without signal sequence were amplified with the primers 5′-GGAATTCCATAT-GAAAAAATTAT-3′ and 5′-CGGGATCCTTAATGATGAT-GATGATGATGAGCATTTTCAGGGATAGCAAGCCC-3′ and cloned into pET11a. The proteins were overexpressed in *E. coli* BL21 (DE3) pLysS and purified as described for the N1–N2 fragment.[17] Variants of the isolated N2 domain [residues 102–205 of mature G3P, extended by $(His)_6$] were expressed and purified as described previously.[42] Site-directed mutagenesis of N2 was performed by QuikChange (Stratagene).

Protein crystallization and structure determination

0SS-G3P*-stab was crystallized by vapor diffusion using the hanging drop setup at 16 °C. The reservoir solution (500 μl) contained 25%–30% PEG (polyethylene glycol)

3350, 0.05 M $CaCl_2$ and 0.2 M NH_4Cl in 0.1 M Tris buffer, pH 8.5. Two drops were set up per reservoir by mixing 2 μl of protein solution (20 and 30 mg/ml in 20 mM Tris–HCl, pH 8.0, respectively) with 2 μl of reservoir solution. Crystals were measured on an in-house X-ray facility consisting of a rotating Cu anode (Nonius FR 591, Bruker AXS, Karlsruhe, Germany) coupled to an image plate detector (mar345dtb, Marresearch, Hamburg, Germany). The data set was processed and scaled using XDS.[43] The crystals belong to space group $P2_12_12$, with cell dimensions $a = 48.9$ Å, $b = 92.2$ Å and $c = 94.4$ Å, and contain two molecules in the asymmetric unit.

The structure was determined using Patterson search techniques. The search model was the chain of molecule A of the crystal structure of G3P from the filamentous phage fd (Protein Data Bank accession code 2G3P), and the best replacement solution was obtained with the program AMoRe.[44] Model building was done with MAIN,[45] and structure refinement was performed with Crystallography & NMR System.[46] The refined model displayed good stereochemical values (Table 5). The loop region between amino acids Gln157 and Lys163 was ill-defined in the electron density and modeled in analogy to the phage fd G3P structure.

Thermal unfolding transitions

The heat-induced unfolding transitions were measured by CD (JASCO J600 A CD spectrometer equipped with a PTC 348 WI Peltier element) at a protein concentration of 4 μM in 100 mM potassium phosphate, pH 7.0, and a rate of 60 K h^{-1}. The transitions were monitored by the decrease of the CD signal at 230 nm with 1-nm bandwidth and 10-mm pathlength. Several unfolding transitions were also followed by the changes in absorbance at 286 and 296 nm (Hewlett–Packard HP 8452A diode array spectrometer, 10-mm pathlength) at a heating rate of 60 K h^{-1}. The experimental data were analyzed on the basis of a three-state approximation for the transitions at 230 nm, with a fixed heat capacity change ΔC_p of 10,000 J mol^{-1} K^{-1} for the first transition (domain N2) and a ΔC_p of 1000 J mol^{-1} K^{-1} for the second transition (domain N1).[47]

GdmCl-induced unfolding transitions

Samples of G3P* (0.5 μM) were incubated for 2 h 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 with a Hitachi F-4010 fluorescence spectrometer in 10-mm cells at 360 nm after excitation at 295 nm. The bandwidths were 5 nm for excitation and 10 nm for emission. 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.[48]

DSC measurements

The DSC measurements were performed with a VP-DSC instrument (MicroCal, Northampton, MA, USA) at protein concentrations between 50 and 183 μM in 100 mM potassium phosphate, pH 7.0. The scan rate was 1.5 K min^{-1}, and the cell volume was 0.523 ml. The measured excess molar heat capacity $C_p(T)$ was analyzed using 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.

Protein Data Bank accession code

The coordinates and structure factor amplitudes are deposited in the Protein Data Bank with accession code 3DGS.

Acknowledgements

This study was supported by grants from the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie. We thank Linn Carstensen and Martin Pfeiffer for their experimental help and the members of our group for many discussions of this work.

References

1. Liu, Y., Breslauer, K. & Anderson, S. (1997). "Designing out" disulfide bonds: thermodynamic properties of 30–51 cystine substitution mutants of bovine pancreatic trypsin inhibitor. *Biochemistry*, **36**, 5323–5335.
2. Hagihara, Y., Shiraki, K., Nakamura, T., Uegaki, K., Takagi, M., Imanaka, T. & Yumoto, N. (2002). Screening for stable mutants with amino acid pairs substituted for the disulfide bond between residues 14 and 38 of bovine pancreatic trypsin inhibitor (BPTI). *J. Biol. Chem.* **277**, 51043–51048.
3. Kather, I., Bippes, C. A. & Schmid, F. X. (2005). A stable disulfide-free gene-3-protein of phage fd generated by *in vitro* evolution. *J. Mol. Biol.* **354**, 666–678.
4. Scott, J. K. & Smith, G. P. (1990). Searching for peptide ligands with an epitope library. *Science*, **249**, 386–390.
5. Marvin, D. A. (1998). Filamentous phage structure, infection and assembly. *Curr. Opin. Struct. Biol.* **8**, 150–158.
6. Lubkowski, J., Hennecke, F., Pluckthun, 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. 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.
8. 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.
9. Koradi, R., Billeter, M. & Wuthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics*, **14**, 51–55.
10. Lubkowski, J., Hennecke, F., Pluckthun, A. & Wlodawer, A. (1998). The structural basis of phage display elucidated by the crystal structure of the N-terminal domains of G3P. *Nat. Struct. Biol.* **5**, 140–147.
11. 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.

12. 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.
13. Click, E. M. & Webster, R. E. (1997). Filamentous phage infection: required interactions with the TolA protein. *J. Bacteriol.* **179**, 6464–6471.
14. Sieber, V., Pluckthun, A. & Schmid, F. X. (1998). Selecting proteins with improved stability by a phage-based method. *Nat. Biotechnol.* **16**, 955–960.
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. Martin, A., Kather, I. & Schmid, F. X. (2002). Origins of the high stability of an *in vitro*-selected cold-shock protein. *J. Mol. Biol.* **318**, 1341–1349.
17. 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.
18. Eckert, B. & Schmid, F. X. (2007). A conformational unfolding reaction activates phage fd for the infection of *Escherichia coli*. *J. Mol. Biol.* **373**, 452–461.
19. Malakauskas, S. M. & Mayo, S. L. (1998). Design, structure and stability of a hyperthermophilic protein variant. *Nat. Struct. Biol.* **5**, 470–475.
20. Wunderlich, M., Martin, A., Staab, C. A. & Schmid, F. X. (2005). Evolutionary protein stabilization in comparison with computational design. *J. Mol. Biol.* **351**, 1160–1168.
21. Hutchinson, E. G. & Thornton, J. M. (1994). A revised set of potentials for beta-turn formation in proteins. *Protein Sci.* **3**, 2207–2216.
22. Makhatadze, G. I. & Privalov, P. L. (1992). Protein interactions with urea and guanidinium chloride—a calorimetric study. *J. Mol. Biol.* **226**, 491–505.
23. 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.
24. Kuhlman, B., Luisi, D. L., Young, P. & Raleigh, D. P. (1999). pK_a values and the pH dependent stability of the N-terminal domain of L9 as probes of electrostatic interactions in the denatured state. *Biochemistry*, **38**, 4896–4903.
25. Cho, J. H., Sato, S., Horng, J. C., Anil, B. & Raleigh, D. P. (2008). Electrostatic interactions in the denatured state ensemble: their effect upon protein folding and protein stability. *Arch. Biochem. Biophys.* **469**, 20–28.
26. Bowler, B. E. (2007). Thermodynamics of protein denatured states. *Mol. Biosyst.* **3**, 88–99.
27. Max, K. E., Wunderlich, M., Roske, Y., Schmid, F. X. & Heinemann, U. (2007). Optimized variants of the cold shock protein from *in vitro* selection: structural basis of their high thermostability. *J. Mol. Biol.* **369**, 1087–1097.
28. Wunderlich, M., Max, K. E., Roske, Y., Mueller, U., Heinemann, U. & Schmid, F. X. (2007). Optimization of the G-beta1 domain by computational design and by *in vitro* evolution: structural and energetic basis of stabilization. *J. Mol. Biol.* **373**, 775–784.
29. Otzen, D. E., Rheinnecker, M. & Fersht, A. R. (1995). Structural factors contributing to the hydrophobic effect: the partly exposed hydrophobic minicore in chymotrypsin inhibitor 2. *Biochemistry*, **34**, 13051–13058.
30. Jaenicke, R. & Böhm, G. (1998). The stability of proteins in extreme environments. *Curr. Opin. Struct. Biol.* **8**, 738–748.
31. Chen, J. M. & Stites, W. E. (2001). Packing is a key selection factor in the evolution of protein hydrophobic cores. *Biochemistry*, **40**, 15280–15289.
32. Karshikoff, A. & Ladenstein, R. (1998). Proteins from thermophilic and mesophilic organisms essentially do not differ in packing. *Protein Eng.* **11**, 867–872.
33. Jaenicke, R. (1987). Folding and association of proteins. *Prog. Biophys. Mol. Biol.* **49**, 117–237.
34. Waldburger, C. D., Schildbach, J. F. & Sauer, R. T. (1995). Are buried salt bridges important for protein stability and conformational specificity? *Nat. Struct. Biol.* **2**, 122–128.
35. Declerck, N., Machius, M., Joyet, P., Wiegand, G., Huber, R. & Gaillardin, C. (2003). Hyperthermostabilization of *Bacillus licheniformis* alpha-amylase and modulation of its stability over a 50 degrees C temperature range. *Protein Eng.* **16**, 287–293.
36. Bai, Y. W. & Feng, H. (2004). Selection of stably folded proteins by phage-display with proteolysis. *Eur. J. Biochem.* **271**, 1609–1614.
37. Eijsink, V. G. H., Bjork, A., Gaseidnes, S., Sirevag, R., Synstad, B., van den Burg, B. & Vriend, G. (2004). Rational engineering of enzyme stability. *J. Biotechnol.* **113**, 105–120.
38. Mosavi, L. K., Minor, D. L., Jr. & Peng, Z. Y. (2002). Consensus-derived structural determinants of the ankyrin repeat motif. *Proc. Natl Acad. Sci. U. S. A.* **99**, 16029–16034.
39. Merz, T., Wetzel, S. K., Firbank, S., Pluckthun, A., Grutter, M. G. & Mittl, P. R. E. (2008). Stabilizing ionic interactions in a full-consensus ankyrin repeat protein. *J. Mol. Biol.* **376**, 232–240.
40. Makhatadze, G. I., Loladze, V. V., Gribenko, A. V. & Lopez, M. M. (2004). Mechanism of thermostabilization in a designed cold shock protein with optimized surface electrostatic interactions. *J. Mol. Biol.* **336**, 929–942.
41. Wunderlich, M., Martin, A. & Schmid, F. X. (2005). Stabilization of the cold shock protein CspB from *Bacillus subtilis* by evolutionary optimization of Coulombic interactions. *J. Mol. Biol.* **347**, 1063–1076.
42. Jakob, R. & Schmid, F. X. (2008). Energetic coupling between native-state prolyl isomerization and conformational protein folding. *J. Mol. Biol.* **377**, 1560–1575.
43. Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800.
44. Navaza, J. (2001). Implementation of molecular replacement in AMoRe. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **57**, 1367–1372.
45. Turk, D. (1992). Weiterentwicklung eines Programms fuer Molekuelgraphik und Elektrondichte-Manipulation und seine Anwendung auf verschiedene Protein Strukturaufklaerungen. Technische Universitat Munchen.
46. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography & NMR System: a new software suite for macromolecular structure determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **54**, 905–921.
47. 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.
48. 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.

Teilarbeit B:

Kather I., **Jakob R.P.**, Dobbek, H. and Schmid, F.X. (2008).
Increased folding stability of TEM-1 beta-lactamase by in vitro selection.
J Mol Biol. **383**, 238-51.

doi:10.1016/j.jmb.2008.07.082 J. Mol. Biol. (2008) 383, 238–251

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Increased Folding Stability of TEM-1 β-Lactamase by *In Vitro* Selection

Insa Kather[1], Roman P. Jakob[1], Holger Dobbek[2] and Franz X. Schmid[1]*

[1]*Laboratorium für Biochemie und Bayreuther Zentrum für Molekulare Biowissenschaften, Universität Bayreuth, 95440 Bayreuth, Germany*

[2]*Laboratorium Proteinkristallographie, Universität Bayreuth, 95440 Bayreuth, Germany*

Received 5 June 2008; received in revised form 16 July 2008; accepted 29 July 2008
Available online 3 August 2008

In vitro selections of stabilized proteins lead to more robust enzymes and, at the same time, yield novel insights into the principles of protein stability. We employed Proside, a method of *in vitro* selection, to find stabilized variants of TEM-1 β-lactamase from *Escherichia coli*. Proside links the increased protease resistance of stabilized proteins to the infectivity of a filamentous phage. Several libraries of TEM-1 β-lactamase variants were generated by error-prone PCR, and variants with increased protease resistance were obtained by raising temperature or guanidinium chloride concentration during proteolytic selections. Despite the small size of phage libraries, several strongly stabilizing mutations could be obtained, and a manual combination of the best shifted the profiles for thermal unfolding and temperature-dependent inactivation of β-lactamase by almost 20 °C to a higher temperature. The wild-type protein unfolds in two stages: from the native state *via* an intermediate of the molten-globule type to the unfolded form. In the course of the selections, the native protein was stabilized by 27 kJ mol^{-1} relative to the intermediate and the cooperativity of unfolding was strongly increased. Three of our stabilizing replacements (M182T, A224V, and R275L) had been identified independently in naturally occurring β-lactamase variants with extended substrate spectrum. In these variants, they acted as global suppressors of destabilizations caused by the mutations in the active site. The comparison between the crystal structure of our best variant and the crystal structure of the wild-type protein indicates that most of the selected mutations optimize helices and their packing. The stabilization by the E147G substitution is remarkable. It removes steric strain that originates from an overly tight packing of two helices in the wild-type protein. Such unfavorable van der Waals repulsions are not easily identified in crystal structures or by computational approaches, but they strongly reduce the conformational stability of a protein.

Keywords: protein stabilization; evolutionary protein design; phage display; protein crystallography; protein folding

Edited by C. R. Matthews

Introduction

During the natural evolution of a protein, only minimal stability is maintained—just high enough to warrant its function in a particular organism. Most spontaneous mutations are, in fact, unfavorable for protein stability; as a consequence, when stabilizing mutations occur, they are not preserved in the absence of evolutionary pressure.

It should be possible to uncover such stabilizing mutations in protein libraries by applying an artificial selection pressure for increased stability. We developed such a selection method for protein stabilization *in vitro*. This technique,[1,2] termed Proside, links the increased protease resistance of stabilized protein variants to the infectivity of the filamentous phage fd. Infection of *Escherichia coli* by this phage is mediated by gene-3-protein (G3P), which consists of three domains N1, N2, and CT. The

**Corresponding author.* E-mail address: fx.schmid@uni-bayreuth.de.

Abbreviations used: β-lactamase, TEM-1 β-lactamase from *E. coli*; G3P, gene-3-protein; GdmCl, guanidinium chloride; EDTA, ethylenediaminetetraacetic acid; T_M, midpoint of a thermal unfolding transition; $[urea]_{1/2}$, midpoint of a urea-induced unfolding transition; ΔG_D, Gibbs free energy of unfolding; m, dependence of ΔG_D on the urea concentration.

CT domain anchors G3P in the phage coat, and N1 and N2 are used for binding the phage first to an F pilus and then to the TolA protein at the surface of the infected cell.[3] The linkage between the domains thus must remain intact for the phage to be infective.[4] In Proside, a repertoire of sequences coding for the protein to be stabilized is inserted between the N2 and CT domains of G3P. In the presence of a protease, these two domains remain covalently linked only when the inserted guest protein is stably folded and not cleaved by the protease. Variants with increased conformational 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.

This was demonstrated in the past with small single-domain proteins such as the cold shock protein CspB from *Bacillus subtilis*[5–8] and the β domain of streptococcal G protein.[7,9] We now used the Proside selection method to stabilize a larger protein, TEM-1 β-lactamase from *E. coli*.

β-Lactamases are monomeric enzymes that convey resistance to β-lactam antibiotics.[10–13] TEM-1 β-lactamase consists of 263 residues and contains a single disulfide bond between Cys77 and Cys123. The stability and folding kinetics of β-lactamase have been analyzed,[14–17] and several crystal structures for the wild-type protein and for variants,[11,18–20] including a structure with a very high resolution of 0.85 Å, are available.[21] These data will be invaluable for understanding the molecular origins of stabilizing mutations resulting from the Proside *in-vitro* selections. TEM-1 β-lactamase has been used as a target to develop methods for optimizing protein stability[22] and function[23–25] by *in-vitro* evolution.

Here we created several gene libraries of β-lactamase by error-prone PCR and selected stabilized variants from them using Proside. The most stabilizing substitutions were combined, and the crystal structure of the corresponding variant was solved. In combination with thermodynamic analysis, it revealed that most of the selected replacements increased conformational stability by improving side-chain packing and, in particular, by optimizing helices at their amino termini.

Results and Discussion

The selection strategy

Phage can lose the inserted guest protein by genetic recombination. Such recombined phage become rapidly enriched during Proside selections because they are protease-resistant and more infectious than phage hosting guest proteins. The frequency of these recombination events increases with the length of the guest protein. In Proside experiments with β-lactamase, the proteolytic selection step was followed by infection of *E. coli* in the presence of ampicillin. Thus, only phage that carried a functional β-lactamase insert could be propagated, and recombined phage without a guest protein were eliminated. Phage propagation in the presence of ampicillin also ensured that variants with inactivating mutations in the active site of β-lactamase were removed from the library.

To stabilize TEM-1 β-lactamase, we performed several Proside selections. In the first set of the experiments, the wild-type form of β-lactamase with intact Cys77-Cys123 disulfide bond was employed as starting material. The gene for this form served as the template for libraries that were created by error-prone PCR. The range for further stabilization was then extended by using the gene for the C77A/C123A variant of β-lactamase as template. It is destabilized because it cannot form the disulfide bond. In the course of the selections, the temperature was raised to increase selection pressure. The libraries based on the C77A/C123A template were also employed for selections in which the concentration of the denaturant guanidinium chloride (GdmCl) was gradually increased. In addition, a library was employed in which two of the positions with strongly stabilizing mutations (positions 182 and 224) were randomized by saturation mutagenesis to identify the optimal residues for these positions. Finally, the best stabilizing mutations from the individual selection experiments were combined manually to yield a single strongly stabilized variant of β-lactamase.

Random mutagenesis of wild-type β-lactamase and thermal selection

Starting with the gene for wild-type β-lactamase, we created in parallel four libraries by error-prone PCR of the entire gene, as described in Materials and Methods. They were then used for 10–15 rounds of Proside selections in the presence of 0.25 μM chymotrypsin at 40–45 °C. After the selections, single clones were picked, and their β-lactamase genes were sequenced. Three of the four clones from library 1 contained the mutation M182T, and all five clones from library 2 showed the mutation A224V (Table 1). These two dominant mutations were, in some variants, accompanied by mutations at positions 27, 208, and 223. Five sequenced clones from libraries 3 and 4 contained the M182T mutation, and the clones from library 4 also showed a silent mutation at position 158. Four of the five variants were expressed and purified, and their thermal stabilities were determined. In addition, variants with single mutations were constructed and analyzed as well. Figure 1a shows two representative thermal unfolding transitions in comparison to wild-type β-lactamase. The substitutions M182T and A224V are indeed strongly stabilizing. Individually, they increased the midpoint of the thermal unfolding transition T_M of β-lactamase by 5.7° and 3.1°, respectively (variants 14 and 15; Table 2). Under the assumption that thermal unfolding of β-lactamase is a two-state reaction, this is equivalent to increases of 9.0 and 6.0 kJ mol^{-1} in the Gibbs free energy of unfolding ΔG_D. The mutations that accompanied M182T and A224V in several variants were either slightly destabilizing (P27S, variant 21;

Table 1. Stabilized variants of β-lactamase from the Proside selections

		Variant											
Position	Wild type	A1	A2	A3	A4	A5	B1	B2	B3	B4	C1	D1	D2
27	Pro		Ser										
62	Pro								Ser				
80	Val								Ile	Ile			
118	Thr									Ile			
146	Lys						Arg				Arg		
147	Glu											Gly	
182	Met	Thr		Thr			Thr	Thr	Thr	Thr	Thr		
201	Leu								Pro	Pro			Pro
208	Ile					Met							
223	Ser			Trp									
224	Ala				Val	Val		Thr					
240	Glu												His
247	Ile								Val				
275	Arg										Leu		

Variants A1–A5 were selected after random mutagenesis of wild-type β-lactamase. Four independent libraries were created and subjected to several rounds of Proside selection at increasing temperature. Variant A1 (selected eight times) occurred in all four libraries, variant A2 (selected twice) occurred in libraries 1 and 2, variant A3 (selected once) occurred in library 1, variant A4 (selected once) occurred in library 2, and variant A5 (selected once) occurred in library 2. Variants B1–B4 were selected after random mutagenesis of β-lactamase C77A/C123A (without disulfide bond). Two independent libraries were created and subjected to several rounds of Proside selection at increasing temperature. Variant B1 (selected three times) and variant B2 (selected once) are from library 1, whereas variant B3 (selected three times) and variant B4 (selected once) are from library 2. Variant C1 was obtained after random mutagenesis of β-lactamase C77A/C123A, which contained, in addition, the mutations K146R and M182T and a Proside selection at increasing concentrations of GdmCl. Variant C1 was selected twice. Variants D1 and D2 were obtained from libraries of β-lactamase C77A/C123A, followed by Proside selection at increasing GdmCl concentration. Variant D1 was selected five times, and variant D2 was selected six times.

S223W, variant 22) or slightly stabilizing (I208M, variant 13; Table 2).

The mutations M182T and A224V had been discovered before. M182T is a global suppressor mutation that appeared repeatedly in clinical isolates of extended-spectrum β-lactamase with broadened substrate specificities.[26–29] One hundred sixty-one extended-spectrum or inhibitor-resistant β-lactamase variants are known†. The M182T mutation occurs in 16 variants, and the A224V mutation occurs in one of these variants. By its globally stabilizing effect, M182T probably compensates for losses in stability caused by mutations at the active site that were necessary to change the substrate specificities of these variants.[28]

In addition to these natural occurrences, the M182T mutation was also identified in the course of artificial *in-vitro* selections for altered substrate specificities of TEM-1 β-lactamase[24,25] and in an *in-vivo* selection that searched for mutations compensating for the loss of stability caused by a truncation of β-lactamase.[22] In this selection, the A224V mutation was identified as well. Finding two stabilizing mutations that had been discovered before was disappointing but, at the same time, reassuring because it showed that Proside indeed identified strongly stabilizing mutations.

Selection of stabilized β-lactamase variants in the absence of its disulfide bond

The temperature range for Proside selections is limited by the stabilities of the protease (chymotrypsin) and the phage G3P, both of which are stable up to approximately 60 °C. The single M182T substitution increased the transition temperature of β-lactamase to 56.5 °C; thus, the limit of the selection system is almost reached when this mutation has occurred. In order to select further stabilizing mutations, we used a variant of β-lactamase as a template, in which Cys77 and Cys123 (which form the single disulfide bond of β-lactamase) were substituted by Ala. These substitutions destabilize β-lactamase towards thermal unfolding (Table 2).

Two libraries of C77A/C123A β-lactamase were created in parallel by error-prone PCR, as described in Materials and Methods, and used for Proside selections. In the course of the 11 selection rounds, the temperature was gradually increased from 40 to 50 °C. At the end, four different variants were found (B1–B4; Table 1). Again, all of them contained the M182T mutation, in combination with other mutations. For three of these variants, the selected mutations were introduced into the disulfide-containing wild-type protein, and the corresponding proteins were overexpressed and purified. They all were stabilized relative to wild-type β-lactamase (Fig. 1b); however, in two of them (B1, variant 6; B4, variant 4), the additional mutations did not improve the stability relative to the variant with the single M182T replacement (Table 2). Only variant B3 (variant 5; Fig. 1b), which contained four additional substitutions, was significantly more stable than M182T β-lactamase.

The selected variants not only are significantly more stable than the wild-type protein but also showed catalytic activities similar to those of wild-type β-lactamase when assayed with the artificial substrate nitrocefin[30] (Table 2).

† See complete list at http://www.lahey.org/Studies/

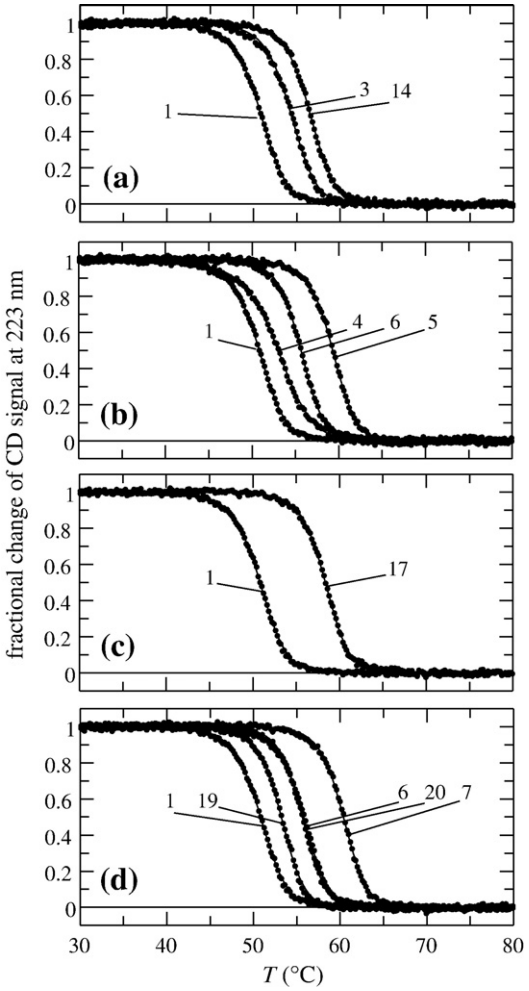


Fig. 1. Thermal stabilities of the selected β-lactamase variants. (a) Variants obtained after random mutagenesis of wild-type β-lactamase: (1) wild-type β-lactamase (in all panels); (3) I208M/A224V; (14) M182T. (b) Variants obtained after random mutagenesis of β-lactamase without the disulfide bond: (4) V80I/T118I/M182T/L201P; (6) K146R/M182T; (5) P62S/V80I/M182T/L201P/I247V. (c) Variant (17) M182T/A224V as obtained after saturation mutagenesis at positions 182 and 224. (d) Variants obtained after random mutagenesis of β-lactamase without the disulfide bond and selection at increasing concentrations of GdmCl: (19) E147G; (6) K146R/M182T; (20) R275L; (7) K146R/M182T/R275L. The transitions were measured by CD at 223 nm. The fractional changes in the CD signal, as obtained after a two-state analysis of the data, are shown as a function of temperature. The transitions were measured at a protein concentration of 4 μM in 50 mM sodium phosphate (pH 7.0) at a path length of 10 mm. The T_M values are given in Table 1. The numbering is as in Table 1.

Determination of an optimal combination for positions 182 and 224

With ΔG_D increases of 9.0 and 6.0 kJ mol^{-1}, the substitutions M182T and A224V, respectively, provide the major contributions to stability (Table 2). Error-prone PCR leads to single-base changes and, therefore, accesses only a small fraction of sequence space at a particular position. To examine whether even more stabilizing mutations can be found for positions 182 and 224, we randomized the corresponding codons in the gene for C177A/C123A β-lactamase by saturation mutagenesis and used this library for a 10-round Proside selection at 40–50 °C. All sequenced variants showed the combination M182T/A224V. The corresponding variant of wild-type β-lactamase showed a T_M of 58.4 °C (Fig. 1c). This suggests that, indeed, Thr and Val are the optimal residues for positions 182 and 224, respectively.

Urea-induced unfolding of stabilized variants

Unlike the monophasic thermal transitions of β-lactamase, those induced by denaturants are complex.[15,17,31] Figure 2 shows urea-induced unfolding transitions, as determined by CD and fluorescence for the wild-type protein and the M182T variant.

It is assumed that native wild-type β-lactamase (N) first unfolds into a molten-globule-type intermediate (I), which is then converted into the unfolded form (U). The successive transitions N↔I and I↔U are both accompanied by changes in CD at 220 nm, and the measured unfolding curve becomes broad (Fig. 2a, curve 1). Trp fluorescence largely follows the N↔I reaction; therefore, the fluorescence-detected transition of the wild-type protein occurs at a lower urea concentration and shows apparent cooperativity higher than the curve measured by CD (Fig. 2).

The M182T replacement preferentially stabilizes N relative to I with little effect on the I↔U transition. This leads to an apparent increase in the cooperativity of the CD-detected transition (Fig. 2a) and to a strong shift of the fluorescence-detected N↔I transition (Fig. 2b). The analysis indicates that the M182T substitution increases the stability of N (relative to I) by 15.2 kJ mol^{-1} (variant 14; Table 3).

Urea-induced unfolding transitions were also determined for several other variants. The corresponding data are shown in Table 3. For the CD-detected transitions, only the midpoints and apparent m (dependence of ΔG_D on urea concentration) values are given. They are correlated, indicating that, with increasing stability, apparent cooperativity also increases, presumably because the selected mutations primarily stabilize the folded state N relative to the intermediate I. The fluorescence-detected N↔I transitions were analyzed as two-state reactions, and the corresponding changes in ΔG_D are given as well. Most of these data correlate well with those derived from the thermal transitions (Table 2). Stronger stabilizations are, however, calculated from the urea transitions for variants 14 and 17.

Selection in the presence of GdmCl

To find further stabilizing mutations in β-lactamase, we used the library without the disulfide

Table 2. Stabilities and activities of the wild-type and mutated forms of β-lactamase

Selection	Number	β-Lactamase variant	T_M (°C)	$\Delta G_{D(50\ °C)}$ (kJ mol^{-1})	Activity (A_{486} s^{-1})
	1	Wild type	50.8	1.4	0.00125
	2	C77A C123A	48.5	−1.9	0.00023
A5	3	I208M/A224V	54.3	7.0	0.00246
B4	4	V80I/T118I/M182T/L201P	53.0	3.6	0.00052
B3	5	P62S/V80I/M182T/L201P/I247V	59.3	14.7	0.00082
B1	6	K146R/M182T	55.6	10.1	0.00137
C1	7	K146R/M182T/R275L	60.4	16.2	0.00114
	8	L201P	52.2	3.3	0.00105
	9	P62S	51.8	3.2	0.00115
	10	I247V	51.2	2.0	0.00207
	11	V80I	51.6	2. 7	0.00148
	12	A224T	50.3	0.4	0.00172
	13	I208M	51.8	3.1	0.00190
A1	14	M182T	56.5	10.4	0.00209
A4	15	A224V	53.9	6.4	0.00114
	16	K146R	50.5	0.8	0.00230
	17	M182T/A224V	58.4	13.3	0.00309
	18	E240H	51.1	1.9	0.00105
D1	19	E147G	53.2	6.0	0.00204
	20	R275L	55.8	9.3	0.00178
A2	21	P27S	48.8	−2.0	n.d.
	22	S223W	48.4	−2.6	n.d.
	23	P62S/V80I/M182T/L201P/ I247V E147G/I208M/A224V/R275L	69.2	28.6	0.00705

n.d., not detected.

For all variants, the melting temperature T_M and the Gibbs free energy of denaturation at 50 °C, $\Delta G_{D(50\ °C)}$, are given. Thermodynamic parameters were derived from thermal unfolding transitions measured by CD at 223 nm. The data were analyzed according to a two-state model with a fixed heat capacity change ΔC_p of 30 kJ mol^{-1} K^{-1}. The accuracy of the T_M values is about ±0.2°, whereas the accuracy of $\Delta G_{D(50\ °C)}$ is correlated with T_M. For variants with T_M values between 45 and 55 °C, it is ±0.3 kJ mol^{-1}; for variants with T_M values >55 °C, it is ±0.5 kJ mol^{-1}. The activities were determined at 25 °C with 2.5 nM protein and 20 μM nitrocefin in 100 mM potassium phosphate and 1 mM EDTA (pH 7.0). The activity of the combined stabilized variant (variant 23) was measured in the presence of 100 μM nitrocefin. The initial reaction rates are given. The first column gives the denotation from the selections (cf. Table 1), whereas the second column provides identification numbers for variants that are also used in the figures. Variants without entries in column 1 were created by directed mutagenesis for analytical purposes.

bond and performed 10 additional Proside rounds. In the course of this selection, the temperature was kept constant (25 °C), and the concentration of GdmCl was increased in steps from 0.75 to 1.75 M. Two new variants (D1 and D2; Table 1) with three mutations were identified. We introduced them individually into wild-type β-lactamase to determine their contributions to stability. All three of them are stabilizing and increase T_M by 0.3 °C (E240H), 2.4 °C (L201P), and 3.4 °C (E147G).

Surprisingly, the M182T mutation was not found, although it had been present in all variants from the thermal selections with this library. To examine this further, we used the K146R/M182T variant, which originated from the thermal selection (B1; Table 1), as template for an additional library of variants. It was created by error-prone PCR and used for an analogous selection as described above in the presence of GdmCl at increasing concentration. At the end, a single variant could be identified (C1; Table 1). It still carried the two mutations from the starting library and, in addition, the mutation R275L. Its T_M value of 60.4 °C (Table 2) was the highest of all selected variants. As a single substitution, R275L is indeed strongly stabilizing (Table 2) and increases ΔG_D by 7.9 kJ mol^{-1}. As in all other variants, those from the selections in the presence of GdmCl were enzymatically active (Table 2).

Several of the stabilizing substitutions involved changes in the distribution of charged residues. To examine whether changes in coulombic interactions contribute to the effects of these substitutions, we measured the thermal unfolding transitions of the wild-type protein and variants 5, 7, 19, and 20 as a function of NaCl concentration. For all these proteins, ΔG_D increased in the same fashion with salt concentration (data not shown), indicating that changes in charge interactions do not contribute significantly to the observed stabilizations.

Combination of stabilizing mutations

With the information gathered from the selections and the analysis of the single variants, we identified the nine best mutations and introduced them together into wild-type β-lactamase with the disulfide bond. The T_M value of this protein (variant 23; Table 1) is increased by 18.4 °C (T_M = 69.2°C; Fig. 3a), and the thermal stability is increased by 27.2 kJ mol^{-1} (Table 2). To our knowledge, this is currently the most stable variant of TEM-1 β-lactamase.[22,28,32]

This variant could not be unfolded completely by urea; therefore, we used the stronger denaturant GdmCl. Figure 3b compares the transitions of variant 23 and of the wild-type protein, as followed by circular dichroism at 220 nm. The GdmCl-

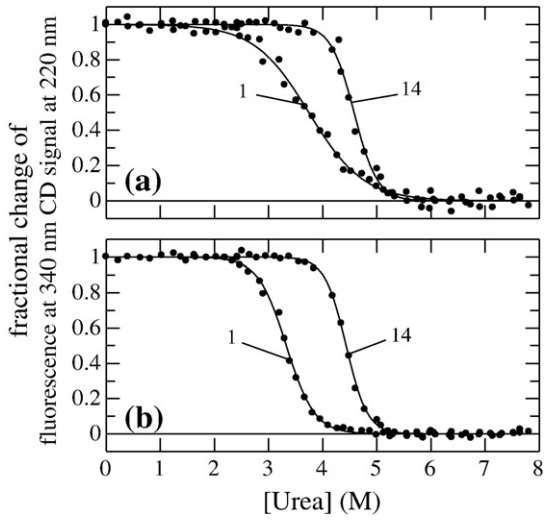


Fig. 2. Urea-induced unfolding transitions of (1) wild-type β-lactamase and (variant 14) the M182T variant. The transitions were measured (a) by CD at 220 nm and (b) by fluorescence at 340 nm. The fractional changes in CD or fluorescence, as obtained after a two-state analysis of the data, are shown as a function of urea concentration. The transitions were measured at a protein concentration of 0.35 μM in 50 mM sodium phosphate (pH 7.0) at a path length of 10 mm at 25 °C.

induced unfolding of wild-type β-lactamase is biphasic, and the consecutive N↔I and I↔U transitions show midpoints of 1.1 and 2.2 M GdmCl, respectively. The sum of the corresponding *m* values of 22.6 and 5.8 kJ mol^{-1} M^{-1} agrees surprisingly well with a value of 26.7 kJ mol^{-1} M^{-1}, which was calculated by the procedure of Myers *et al.* for the cooperative unfolding of a protein with 263 residues and a single disulfide bond as β-lactamase.[33] The change in Trp fluorescence upon unfolding essentially follows the first transition to the folding intermediate (Fig. 3c).

The stabilized variant 23 differs from this pattern. It shows an apparently cooperative transition when followed either by circular dichroism or by Trp fluorescence ($[GdmCl]_{1/2}$ = 2.7 M). The apparent *m* value of 12.7 kJ mol^{-1} M^{-1} is, however, much smaller than the value of 26.7 kJ mol^{-1} M^{-1}, as expected for a fully cooperative N↔U two-state transition. This indicates that the apparently cooperative transitions for variant 23 in Fig. 3b and c in fact consist of two closely spaced transitions. An analysis, therefore, is very difficult. Under the assumption, that the first part of the unfolding transition in Fig. 3b is shifted by 1.1 M GdmCl and that the m value for the N↔I transition is similar for the wild-type protein (variant 1) and the stabilized variant (variant 23) are the same (21.6 kJ mol^{-1} M^{-1}), an increase in stability of N relative to I of about 24 kJ mol^{-1} can be estimated. This is in fair agreement with the value of 27 kJ mol^{-1} (at 50 °C) from the analysis of the thermal transitions (Table 2).

The *in vitro* selection thus led to an apparent change in the unfolding mechanism of β-lactamase. The mutations stabilized the folded form so strongly (relative to the intermediate) that it remained folded up to 2 M GdmCl (Fig. 3). Its transition midpoint (2.7 M GdmCl) is higher than the midpoint of the second transition of the wild-type protein (2.2 M GdmCl), in which the intermediate is converted into the unfolded state. Therefore, for the stabilized variant, unfolding of the intermediate becomes coupled with its formation from the native state, or, in other words, overall unfolding becomes more cooperative and resembles a two-state N↔U process.

The stabilized variant 23 shows an activity higher than that of the wild-type enzyme in the assay with nitrocefin (Fig. 4a). As expected from the strongly increased thermal stability, this variant lost its activity at a temperature that was almost 20 °C higher than the corresponding value for the wild-type enzyme. The assay is not performed at a saturating substrate concentration; therefore, the activity profiles in Fig. 4a are determined by the temperature dependences of the conformational stability, the K_m value, and the k_{cat} value. Variant 23 also shows a decreased rate of irreversible thermal inactivation at 60 °C (Fig. 4b).

Table 3. Stability data for β-lactamase variants derived from urea-induced unfolding transitions

		CD at 220 nm		Fluorescence at 340 nm		
Number	Variant	$[Urea]_{1/2}$ (M)	*m* (kJ mol^{-1} M^{-1})	$[Urea]_{1/2}$ (M)	*m* (kJ mol^{-1} M^{-1})	$\Delta\Delta G_D$[a] (kJ mol^{-1})
2	C77A/C123A	2.7	7.4	2.6	9.7	−6.8
1	Wild-type β-lactamase	3.7	4.9	3.3	9.5	0
4	V80I/T118I/M182T/L201P	3.7	8.4	3.7	7.0	2.8
6	K146R/M182T	4.4	10.5	4.3	10.1	10.1
5	P62S/V80I/M182T/L201P/I247V	4.5	10.8	4.6	10.6	13.8
14	M182T	4.6	11.6	4.5	12.7	15.2
17	M182T/A224V	4.8	12.0	4.8	11.8	17.7

The accuracies of *m* (dependence of ΔG_D on urea concentration) and $[urea]_{1/2}$ (urea concentration at the midpoint of transition) are ≈1 kJ mol^{-1} M^{-1} and ≈0.2 M, respectively. The parameters were derived from a two-state analysis of urea-induced unfolding transitions measured by CD at 220 nm and by fluorescence at 340 nm at a protein concentration of 0.35 μM in 50 mM sodium phosphate (pH 7.0) and at a path length of 10 mm at 25 °C, as shown in Fig. 2. The variants are numbered as in Table 2.

[a] $\Delta\Delta G_D$ is the difference in Gibbs free energy of unfolding relative to the wild-type protein (variant 1) in the absence of urea.

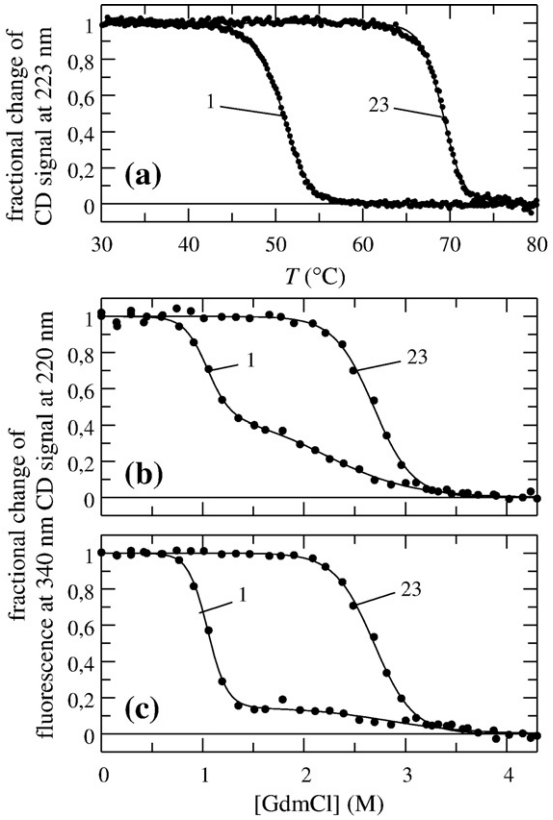


Fig. 3. Stability of the most stable β-lactamase variant (23) in comparison to the wild-type protein (1). (a) Thermal unfolding as measured by CD at 223 nm. The fractional changes in the CD signal, obtained after a two-state analysis of the data, are shown as a function of temperature. The T_M values are given in Table 1. The GdmCl-induced unfolding at 25 °C as measured (b) by CD at 220 nm and (c) by fluorescence at 340 nm. The fractional changes in CD or fluorescence, as obtained after two-state or three-state analysis, are shown as a function of the GdmCl concentration. The protein concentrations were 4 μM (thermal unfolding) and 0.35 μM (GdmCl-induced unfolding) in 50 mM sodium phosphate (pH 7.0). A three-state analysis of curve 1 in (b) gives midpoints at 1.1 and 2.2 M GdmCl and apparent m values of 21.6 and 5.8 kJ M^{-1} mol^{-1}, respectively. A two-state analysis of curve 23 in (b) gives a midpoint of 2.7 M GdmCl and an apparent m value of 12.7 kJ M^{-1} mol^{-1}. The corresponding analyses in (c) give midpoints of 1.1 and 2.8 M GdmCl and apparent m values of 26 and 6 kJ M^{-1} mol^{-1} for variant 1, and a midpoint of 2.7 M GdmCl and an apparent m value of 12.8 kJ M^{-1} mol^{-1} for variant 23.

The structure of the best β-lactamase variant

The crystal structure of the most stable variant (variant 23; Table 2) was solved at 2.0 Å resolution. Details of the crystallization and structure determination are summarized in Table 4. The protein crystallized in $P2_12_12$ symmetry (as the wild-type form of the protein), and the backbone atoms of the two proteins superimposed with an rmsd of 0.5 Å (Fig. 5). For the comparisons in Fig. 6, the high-resolution structure 1M40 was used as reference.[21] This structure contains already the M182T replacement; therefore, to interpret the consequences of this substitution, we used the structure of wild-type TEM-1 β-lactamase (1BTL) for comparison.[18]

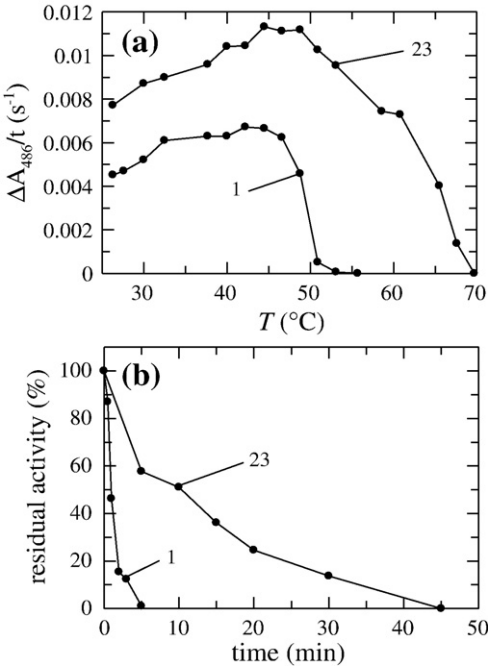


Fig. 4. Enzymatic activity of the most stable β-lactamase variant (23) in comparison to the wild-type protein (1). (a) Temperature dependence of activities. (b) Residual activities measured at 25 °C after incubation at 60 °C for different time periods. The activity was measured by absorbance at 486 nm at a protein concentration of 2.5 nM in 100 mM potassium phosphate and 1 mM EDTA (pH 7.0) in the presence of 100 μM nitrocefin.

Structural interpretation of the contributions of individual mutations to stability

The conservative replacements V80I and I247V add or remove a methyl group from a β-branched

Table 4. Statistics on diffraction data and structure refinement

Data set	Variant 23
Total/unique reflections	79,907/19,290
R_s [a]	0.146 (0.348)
Resolution (Å)	19.4–2.0 (2.1–2.0)
Completeness (%)	98.5 (97.6)
$(I)/(\sigma I)$	8.3 (3.6)
Model R/R_{free} factor (%)[b]	21.8/27.2
rmsd from ideal geometry	
Bonds (Å)	0.012
Angles (°)	1.4
Ramachandran statistics (%)	
Most favored/additionally allowed/ generously allowed/disallowed regions	93.0/6.1/0.9/0.0

Friedel mates were merged in the refinement statistics.
[a] $R_s = \sum_h \sum_i I_i(h) - \langle I(h) \rangle / \sum_h \sum_i I_i(h)$, where i represents independent observations of reflection h.
[b] R_{free} factor was calculated from 5% of the data, which were removed at random before the refinement was carried out.

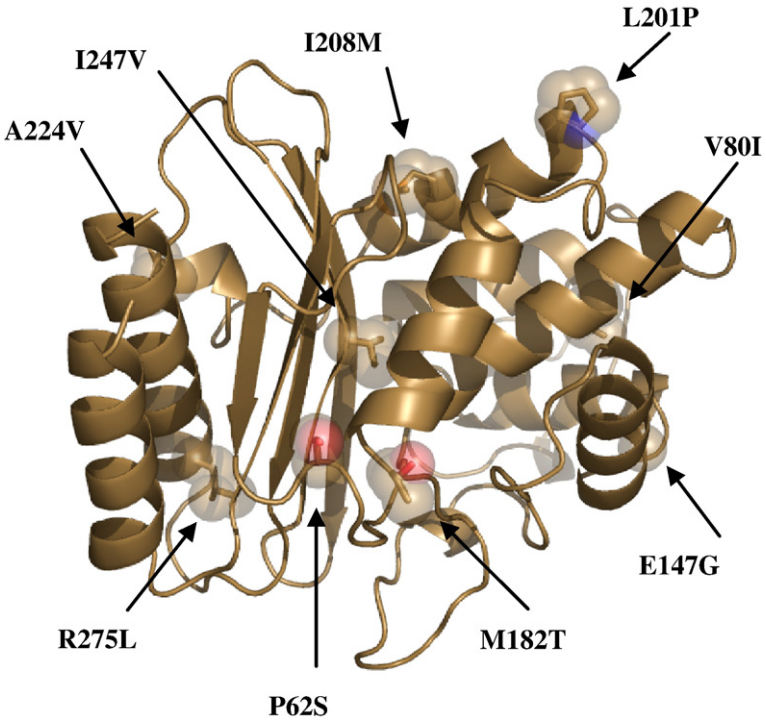


Fig. 5. Crystal structure of the most stable β-lactamase variant (variant 23). The side chains of the selected mutations are shown in space-filling representation.

amino acid. Both are slightly stabilizing by 1.3 and 0.6 kJ mol^{-1}. These values are too small to warrant a structure-based interpretation.

I208M

Ile208 is in the middle region of the 201–212 helix (Fig. 5), and the I208M substitution adds 1.7 kJ mol^{-1} to the stability of β-lactamase. Ile208 is at the bottom of a groove that is lined by Leu194 on one side and by Pro257 and Phe230 on the other (Fig. 6a). In the wild-type protein, Ile208 contacts only Leu194, whereas in the stabilized variant, Met208, with its larger side chain, contacts Pro257 and Phe230 as well. We assume that the observed stabilization originates from this improvement in molecular packing. Experiments with T4 lysozyme showed that multiple mutations to Met were well suited to repack the core of this protein.[34]

L201P

The L201P substitution stabilizes β-lactamase by 1.9 kJ mol^{-1}. Leu201 occupies the first position of the 201–212 helix (Fig. 5) and shows backbone angles of $\Phi = -55°$ and $\Psi = -47°$. This position and its backbone conformation are ideally suited for Pro. In fact, Pro residues are especially good for the first position of α-helices.[35] The side chain of Leu201 in the reference protein is solvent-accessible and shows two alternative conformations. Thus, the side chain of Pro201 is well accommodated (Fig. 6b). The L201P replacement should also be intrinsically stabilizing because it reduces the entropy of the unfolded state.

A224V

Ala224 is in a loop near the protein surface (Fig. 5), and its methyl side chain points to a pocket that is lined by the side chains of Ala280, Ala284, and Ile287 of the 272–289 helix (Fig. 6c). The side chain of Val224 in the stabilized variant fills the pocket between these residues much better than the alanine in the wild-type protein (Fig. 6d). This improved packing leads to a 6.0-kJ mol^{-1} gain in stability (Table 2). The A224V substitution in β-lactamase provides further example of stabilization by improved side-chain packing, as observed previously in other *in vitro* selections of stabilized protein variants.[9,36]

R275L

The R275L substitution is strongly stabilizing and increases ΔG_D by 7.9 kJ mol^{-1}. The aliphatic part of the Arg275 side chain is completely buried in the wild-type protein; of the guanidino head group, only 30% is accessible to water. Arg275 is close to the N-terminus of the 272–289 helix; thus, it might interact unfavorably with the dipole of this helix. Adjacent to the C^{γ} methylene group of Arg275, there is little extra space (Fig. 6d) and, as a consequence, the C^{γ}, C^{δ}, and N^{ε} atoms of Arg272 show alternative conformations in the 1M40 reference structure.[21] Leu275, with its branched side, apparently fills this space better. We suggest that removal of a partially buried charged guanidino group, which, moreover, is located near the N-terminus of a helix, is largely responsible for the strong stabilization afforded by the R275L replacement.

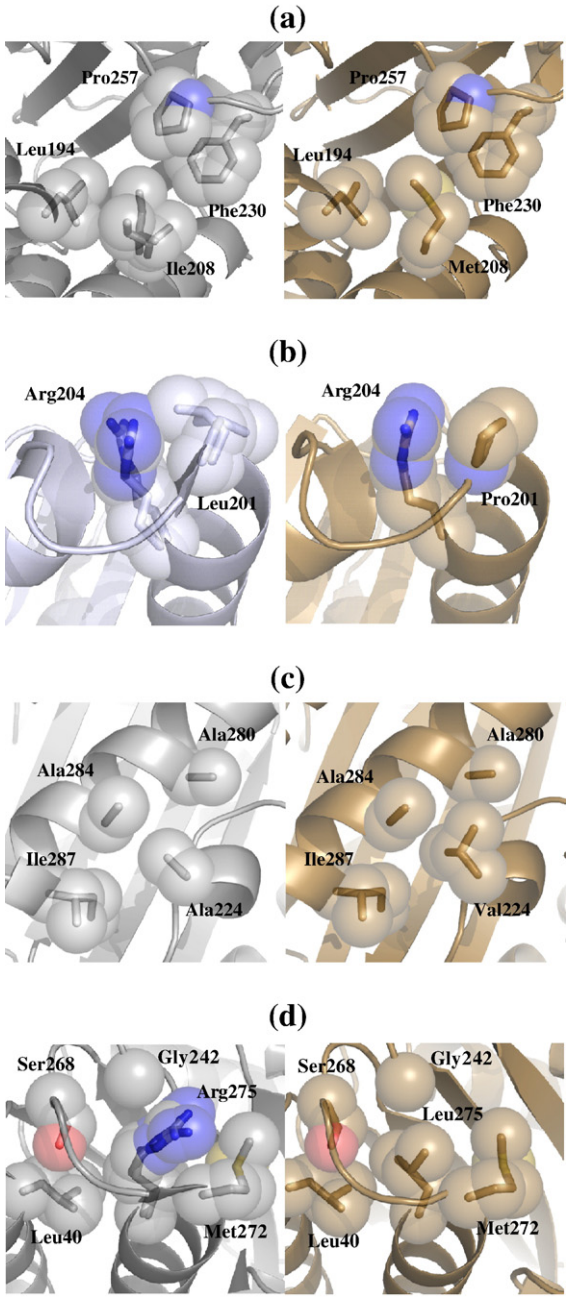


Fig. 6. Comparisons of the local structure of wild-type β-lactamase (left) and the local structure of the most stable β-lactamase variant (variant 23; right) in the vicinity of the stabilizing mutations (a) I208M, (b) L201P, (c) A224V, and (d) R275L. For comparison, the 1M40 structure of β-lactamase[21] was used.

E147G

The E147G substitution stabilizes β-lactamase by 4.6 kJ mol^{-1} (Table 2)—a puzzling result because it apparently violates several basic rules of protein stability. Glu147 is located at the third position of the 145–155 helix, and its side chain is exposed to the solvent. The carboxyl group of Glu147 should interact favorably with the dipole of this helix and thus be stabilizing. The substitute Gly147 is a helix-breaking residue and, moreover, favors the unfolded state entropically.

The comparison between the structure of the stabilized variant and the structure of the reference protein (1M40) provides an explanation and suggests that the side chain of the original residue Glu147 is in a strained conformation (Fig. 7a). The two consecutive helices 133–142 and 145–155 are linked in an unusual manner by a stretch of two glycine residues with positive Φ angles. Gly143 forms the C-cap of the 133–142 helix, and Gly144 is at the N-cap position of the 145–155 helix. This tight connection between the two helices leads to a close contact between the carbonyl oxygen of the last residue of the first helix (Ile142) and the β-CH_2 of Glu147 at the third position of the second helix (Fig. 7a). The corresponding distances are 3.0 Å in molecule A of the reference structure and 3.3 Å in molecule B of the reference structure (1M40). In both molecules, the side chain of Glu147 shows an unusual conformation with χ_1 and χ_2 angles of 172° and −46°, respectively. Together, this suggests that the close approach between the two helices might lead to a van der Waals repulsion and forces the Glu147 side chain into a high-energy conformation.

In fact, the removal of the side chain of Glu147 by the E147G substitution allows the C^{α} of Gly147 to move 1.4 Å closer to the carbonyl oxygen of Ile142. As a consequence, the ends of the two helices in total move towards each other by more than 1 Å. The

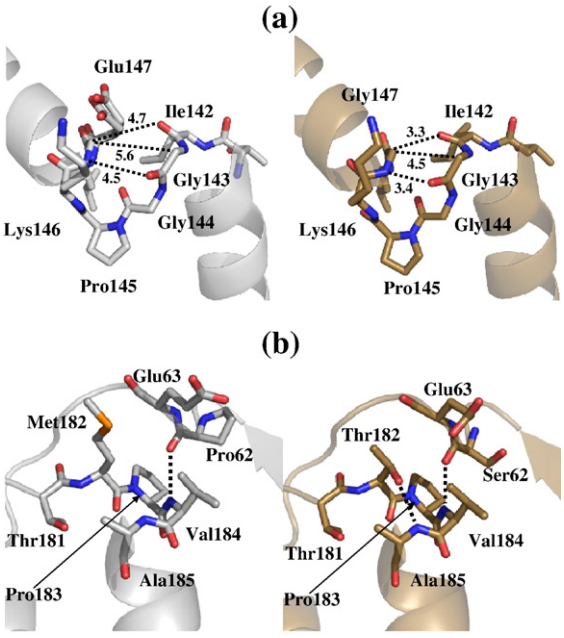


Fig. 7. Comparisons of the local structure of wild-type β-lactamase (left) and the local structure of the most stable β-lactamase variant (variant 23; right) in the vicinity of the stabilizing mutations (a) E147G and (b) M182T. The dotted lines and the numbers in (a) give interatom distances (Å), whereas the dotted lines in (b) represent H bonding in the helix N-cap region. For the comparison in (a), the 0.85-Å 1M40 structure of β-lactamase[21] was used; for the comparison in (b), the 1.8-Å 1BTL structure[18] was used.

steric strain, as present in the reference protein, is thus relaxed, and both packing and H bonding are optimized in this region of the protein. Apparently, these improvements are so strong that they overcompensate for the loss of the interaction of Glu147 with the helix dipole, as well as for the helix-breaking effect and the unfavorable entropic contribution of Gly147 to the mutant.

The overly close contact between the two helices in the reference protein was not identified as unfavorable by PROCHECK.[37] A stabilizing substitution such as E147G that apparently violates the basic rules of helix stability would certainly be very hard to predict by rational design. To understand the stabilizing effects of such mutations, it is essential to combine thermodynamic analyses with high-resolution crystal structures for both the wild-type protein and the mutated protein.

M182T

M182T is an intriguing substitution. It stabilizes β-lactamase by 9.0 kJ mol^{-1}, and it has been found in naturally occurring variants,[38,39] as well as in variants from *in-vitro* evolution experiments.[22,24,25,40] It is also present as a single replacement in our reference protein for which the 0.85-Å structure (1M40) was determined.[21] This structure and that of our stabilized variant (variant 23; Table 1) superimpose very well in the vicinity of Thr182. They reveal that, unlike Met182 of the wild-type protein, Thr182 functions as the N-cap residue for the 183–195 helix (Fig. 7b). The O^{γ} of Thr182 engages in a hydrogen bond with the NH of Ala185.

Moreover, in the presence of Thr182, the neighboring residue (Pro62 in the wild-type protein and Ser62 in the stabilized variant) is able to move closer to the N-terminus of this helix. The distance between the NH of Val184 and the CO of residue 62 is shortened from 3.2 Å (wild-type protein) to 3.0 Å in the stabilized variant, indicating that the corresponding H bond is strengthened.

After the M182T mutation, the amino-terminal region of the 183–195 helix is thus almost perfect. Thr182 is an excellent N-cap residue, the H bond between the NH of residue 2 (Val184) and the CO of residue 62 is strengthened, and helix position 1 is occupied by a proline (Pro183), which is the preferred residue for the first position of a helix. Thus, no unpaired NH is left in the first turn of this helix.

The interpretation of the structural basis of the M182T substitution has a long history. Shortly after its discovery in a clinical isolate,[38] it was suggested from molecular modeling that Thr182 is oriented such that its hydroxyl group engages in a stabilizing H-bond with Glu64 across the domain interface of β-lactamase.[41] Orencia *et al.* followed this suggestion when they interpreted a 2.4-Å-resolution structure of a β-lactamase variant with the same mutation,[40] which had originated from a gene-shuffling experiment.[24] The structure of our stabilized variant, as well as the high-resolution structure of Minasov *et al.*, shows that the side chain of Thr182 is oriented such that the γ-methyl group points towards the domain interface and that the Thr182 hydroxyl group engages in an N-cap H bond with Ala185.[21] The same arrangement was found in SHV-2 β-lactamase, which differs from the TEM-1 enzyme at 98 positions.[42] Minasov *et al.* and Nukaga *et al.* focused on the active sites of these β-lactamases and did not discuss the structural implications of the M182T substitution.[21,42]

Conclusions

In previous work, we had used the Proside selection method to stabilize small and monomeric single-domain proteins with less than 100 residues.[5–7,9,43] With 263 residues, β-lactamase is much larger, and it is the first enzyme to have been stabilized by Proside. By combining the *in vitro* proteolytic selection step with phage propagation in the presence of ampicillin, we could prevent the loss of the β-lactamase insert from the phage and eliminate β-lactamase variants without enzymatic activity from the libraries.

Only a tiny fraction of the sequence space could be explored by using libraries that were created by error-prone PCR. Nevertheless, several strongly stabilizing amino acid substitutions—which, in combination, increased the T_M value by 18.4°—could be found. Not all of the selected mutations were, in fact, stabilizing. Several neutral or slightly destabilizing mutations were retained because they accompanied other mutations with a very high stabilizing potential, such as M182T. Therefore, it was mandatory to analyze the contributions of the individual mutations by directed mutagenesis.

The β-lactamases provide the unique opportunity to compare the results from evolution *in vivo* and *in vitro*. The use of new β-lactam antibiotics "forced" pathogenic bacteria to change the specificities of their β-lactamases and to counterbalance the corresponding destabilizing mutations in the active site by compensatory stabilizing mutations at remote positions. M182T is one of these stabilizing "suppressor" mutations.[38,39,44] It occurred in 16 out of 161 mutants of TEM-1 β-lactamase. The R275L mutation occurred three times, whereas the A224V mutation occurred once in this large set of mutants. Three of our four strongly stabilizing mutations have thus been used as suppressors in the adaptation of pathogenic bacteria to new antibiotics.

The wild-type form of β-lactamase unfolds in two steps.[17,31] In the first step (N↔I), the native protein (N) is converted into a molten-globule-like unfolding intermediate (I), which is then fully unfolded in the subsequent I↔U transition. The overall stability of native β-lactamase is thus determined by the N↔I transition. All the selected stabilizing mutations in fact stabilize N relative to I and, thus, shift the midpoint of the first transition to higher denaturant concentrations. For several mutants, this shift is so strong that the N↔I transition couples with the subsequent I↔U transition. In other words,

the stabilizing mutations increase the cooperativity of unfolding. In principle, there are two major routes for stabilizing a protein: an increase in the transition midpoint and an increase in the cooperativity. Our *in vitro* selections apparently follow both principles.

The four most strongly stabilizing mutations share common properties. In all cases, the wild-type residues show well-ordered structures with low crystallographic *B* values and are not conserved in homologues. Three of these mutations occur at or near the amino-terminal ends of helices and stabilize by optimizing these helices. This may provide guidelines for finding candidate positions for protein stabilization strategies. Consensus analysis of homologous proteins can identify positions in which the protein deviates from its homologs, and the analysis of the local structure can give evidence for unfavorable interactions. Saturation mutagenesis at such positions, followed by Proside selections, would then be suitable for finding the optimal residues for these positions. Regarding the fact that only 43 out of 263 residues are critical for the function of β-lactamase,[12] there should be ample room for further stabilization by *in vitro* evolution.

Materials and Methods

Creation of the phage libraries

For the selection of stabilized variants of β-lactamase, we used a phage derived from fCKCBS,[45] containing a G3P with the mutations T13I, T101I Q129H, and D209Y,[46] and wild-type β-lactamase as the guest protein between the N2 and CT domains. To introduce random mutations, error-prone PCR with a reduced concentration of dNTPs (40–200 μM), 2.5–5 mM $MgCl_2$, and 0.25 mM $MnCl_2$ was performed using flanking primers and *Taq* polymerase. The gene fragments coding for β-lactamase and containing the random mutations were introduced into the phage genome as described.[2]

Libraries with positions 182 and 224 randomized were created by PCR with the primer 5′-GAGCGTGACACCACG(AGCT)(AGCT)(AGCT)CCTGTAGCAATGGCA-3′ at randomized position 182, and with the primer 5′-CCAGCCAGCCGGAAG(AGCT)(AGCT)(AGCT)CGAGCGCAGAAGTGG-3′ at randomized position 224. The gene fragments obtained from PCR experiments were introduced into the phage genome as described.[1,2]

Proside selections

Phage were isolated from *E. coli* TG1 culture medium by polyethylene glycol precipitation and stored in phosphate-buffered saline containing 0.01% (wt/vol) NaN_3. The Proside selections were performed with about 10^8 phage in 50 μl of 100 mM potassium phosphate and 100 μM $CaCl_2$ (pH 8.0). After incubation of the phage solution for 5 min at the respective temperature, proteolysis was performed with 0.25 μM chymotrypsin for 15 min at temperatures between 40 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* TG1 with $A_{600} = 1$. Phage were propagated at 37 °C for 8 h in dYT medium[47] in the presence of ampicillin (β-lactamase with disulfide bond: 300 μg ml^{-1} ampicillin; β-lactamase without disulfide bond: 20 μg ml^{-1} ampicillin). To determine the relative infectivities of the phage of every selection cycle, two 25-μl samples with about 10^8 phage each in 100 mM potassium phosphate and 100 μM $CaCl_2$ (pH 8.0) were incubated for 15 min at the respective temperature: one in the absence of 0.25 μM chymotrypsin and one in the presence of 0.25 μM chymotrypsin. Five-microliter aliquots of these samples were added to 495 μl of a culture of *E. coli* TG1 ($A_{600} \approx 1$) and incubated at 37 °C for 30 min for infection. The number of infected cells and, thus, the number of infectious phage after proteolysis were determined by plating serial dilutions on $dYT^{cam/amp}$ agar. Phage suspensions were inactivated by autoclaving. The completeness of inactivation was confirmed by infection assays.

Expression and purification of the β-lactamase variants

For the expression of the β-lactamase variants with the disulfide bond, the gene fragments coding for the β-lactamase, including the signal sequence, were PCR-amplified from the corresponding double-stranded phage DNA with the primers 5′-GCTCTAGAATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCT-3′ and 5′-AAAAGGCCTTTACCAATGCTTAATCAGTGAGGC-3′. The fragments were cloned into the expression plasmid pTAC11[48] *via* introduced XbaI and StuI restriction sites. The mutations from the selections without disulfide bond were introduced into β-lactamase with the disulfide bond by blunt-end PCR, using pTAC11 with the gene for wild-type β-lactamase as template. The proteins were overexpressed in *E. coli* RB791 in LB medium in the presence of 300 μg ml^{-1} ampicillin. Overexpression was induced by addition of 0.5 mM IPTG. Cells were incubated overnight at 25 °C, then harvested by centrifugation, suspended in lysis buffer [50 mM Tris–HCl, 50 mM NaCl, and 10 mM EDTA, pH 8.0], and frozen at −80 °C. The periplasmic fraction was prepared by treatment with lysozyme and DNaseI (Roche Diagnostics, Basel, Switzerland) for 60 min. After centrifugation, the protein-containing supernatant was treated with polyethyleneimine (0.4%) and centrifuged again to remove DNA. The supernatant was dialyzed against 20 mM Tris–HCl (pH 7.7), diluted threefold with 20 mM Tris–HCl buffer (pH 7.7), and loaded onto an anion-exchange column (Fractogel TMAE; Merck KGaE, Darmstadt, Germany) equilibrated with the same buffer. The protein eluted at 100 mM NaCl in a linear 0–0.2 M NaCl gradient. The fractions with β-lactamase activity were pooled and dialyzed against 20 mM 4-morpholineethanesulfonic acid buffer (pH 6.0). Anion-exchange chromatography was repeated under the same conditions. The β-lactamase-containing fractions were pooled and purified by gel filtration on a Superdex 75 prep-grade column (GE Healthcare, Uppsala, Sweden) with 20 mM Tris–HCl (pH 7.7). The fractions containing β-lactamase were pooled and stored at −20 °C.

Thermal unfolding transitions

The heat-induced unfolding transitions were measured by CD (JASCO J600 CD spectrometer equipped with a PTC 348 WI Peltier element) at a protein concentration of 4 μM in 50 mM sodium phosphate (pH 7.0) at a rate of 60 K h^{-1}. The transitions were monitored by the increase in CD signal at 223 nm (1 nm bandwidth; 10 mm path length).

The experimental data were analyzed according to a two-state model using nonlinear regression[49] and the program Grafit (Erithacus Software, Staines, UK). The heat capacity change ΔC_p was held constant at 30 kJ mol^{-1} K^{-1}.

Urea-induced and GdmCl-induced unfolding transitions

Samples of β-lactamase (0.35 μM) were incubated at 25 °C in the presence of 50 mM sodium phosphate (pH 7.0) and varying concentrations of urea or GdmCl. Protein unfolding was followed by fluorescence at 340 nm after excitation at 280 nm in 10-mm cells in a JASCO FP 600 fluorimeter. The experimental data were analyzed according to a two-state model by assuming linear dependence of fluorescence emission on the urea or GdmCl concentration. A nonlinear 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 denaturant concentration.[50]

CD measurements of urea-induced or GdmCl-induced unfolding were performed with 0.35 μM protein in 50 mM sodium phosphate (pH 7.0). The CD signal was measured at 220 nm (1 nm bandwidth; 10 mm path length) at 25 °C. The experimental data were analyzed on the basis of a two-state approximation for the urea-induced unfolding transitions[50] or a three-state approximation for the GdmCl-induced unfolding transitions.[51]

Activity measurements

β-Lactamase activity was determined using the chromogenic cephalosporin derivative nitrocefin. Activity was followed by the change in absorbance at 486 nm, which was determined with a Hewlett Packard HP8452 spectrophotometer. Measurements were performed with 2.5 nM protein in 100 mM potassium phosphate, 1 mM EDTA (pH 7.0), and addition of 20–100 μM nitrocefin at varying temperatures from 25 to 70 °C. The protein was preequilibrated in the buffer at the corresponding temperature for 5 min, and the cleavage reaction was then started by the addition of nitrocefin, which was incubated at the same temperature. The initial reaction rate was determined for each temperature. As measurements were not performed under substrate saturation, only relative activities could be determined.

The kinetics of irreversible thermal inactivation were determined by incubating the β-lactamase variants for variable time intervals at 60 °C in 100 mM potassium phosphate and 1 mM EDTA (pH 7.0). The remaining activity was then measured at 25 °C under the same conditions in the presence of 100 μM nitrocefin.

Protein crystallization and structure determination

The most stable variant (variant 23; Table 2) was crystallized using the sitting-drop vapor-diffusion method. The optimized condition contained 1.6–1.8 M ammonium citrate and 5% polyethylene glycol 400 in 0.1 M Tris buffer (pH 7.5) and a protein/reservoir ratio of 1:1. Crystals were directly frozen in cold airstream (100 K) after supplementation with 10% polyethylene glycol 400. Diffraction data were collected at −180 °C on a rotating anode X-ray generator (Nonius FR591; Bruker AXS, Karlsruhe, Germany) equipped with an image plate detector (mar345dtb; MarResearch GmbH, Hamburg, Germany).

The data set was processed and scaled using XDS.[52] The crystals belong to space group $P2_12_12$ with cell dimensions $a=39.8$ Å, $b=79.7$ Å, and $c=87.1$ Å, and contain one molecule per asymmetric unit.

Structure determination was performed using Patterson search techniques. The search model was the crystal structure of TEM-1 (Protein Data Bank accession code 1BTL),[18] and the best replacement solution was obtained with the program AMoRe.[53] Model building and structure refinement were performed with Coot[54] and REFMAC5.[55]

Data bank accession codes

The coordinates and structure factor amplitudes have been deposited in the Protein Data Bank with ID code 3DTM.

Acknowledgements

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

References

1. Sieber, V., Plückthun, A. & Schmid, F. X. (1998). Selecting proteins with improved stability by a phage-based method. *Nat. Biotechnol.* **16**, 955–960.
2. Martin, A., Schmid, F. X. & Sieber, V. (2003). Proside: a phage based method for selecting thermostable proteins. *Methods Mol. Biol.* **230**, 57–70.
3. Click, E. M. & Webster, R. E. (1997). Filamentous phage infection: required interactions with the TolA protein. *J. Bacteriol.* **179**, 6464–6471.
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. 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.
6. Martin, A., Kather, I. & Schmid, F. X. (2002). Origins of the high stability of an *in vitro*-selected cold-shock protein. *J. Mol. Biol.* **318**, 1341–1349.
7. Wunderlich, M., Martin, A., Staab, C. A. & Schmid, F. X. (2005). Evolutionary protein stabilization in comparison with computational design. *J. Mol. Biol.* **351**, 1160–1168.
8. Max, K. E., Wunderlich, M., Roske, Y., Schmid, F. X. & Heinemann, U. (2007). Optimized variants of the cold shock protein from *in vitro* selection: structural basis of their high thermostability. *J. Mol. Biol.* **369**, 1087–1097.
9. Wunderlich, M., Max, K. E., Roske, Y., Mueller, U., Heinemann, U. & Schmid, F. X. (2007). Optimization of the G-beta1 domain by computational design and by *in vitro* evolution: structural and energetic basis of stabilization. *J. Mol. Biol.* **373**, 775–784.
10. Joris, B., Ghuysen, J. M., Dive, G., Renard, A., Dideberg, O., Charlier, P. *et al.* (1988). The active-site-serine penicillin-recognizing enzymes as members of the *Streptomyces* R61 DD-peptidase family. *Biochem. J.* **250**, 313–324.

11. Herzberg, O. (1991). Refined crystal structure of beta-lactamase from *Staphylococcus-aureus* PC1 at 2.0-A resolution. *J. Mol. Biol.* **217**, 701–719.
12. Huang, W., Petrosino, J., Hirsch, M., Shenkin, P. S. & Palzkill, T. (1996). Amino acid sequence determinants of beta-lactamase structure and activity. *J. Mol. Biol.* **258**, 688–703.
13. Frere, J. M., Dubus, A., Galleni, M., Matagne, A. & Amicosante, G. (1999). Mechanistic diversity of beta-lactamases. *Biochem. Soc. Trans.* **27**, 58–63.
14. Laminet, A. A. & Plückthun, A. (1989). The precursor of beta-lactamase: purification, properties and folding kinetics. *EMBO J.* **8**, 1469–1477.
15. Vanhove, M., Raquet, X. & Frere, J.-M. (1995). Investigation of the folding pathway of the TEM-1 β-lactamase. *Proteins: Struct. Funct. Genet.* **22**, 110–118.
16. Vanhove, M., Guillaume, G., Ledent, P., Richards, J. H., Pain, R. H. & Frere, J. M. (1997). Kinetic and thermodynamic consequences of the removal of the Cys-77-Cys-123 disulphide bond for the folding of TEM-1 beta-lactamase. *Biochem. J.* **321**, 413–417.
17. Vanhove, M., Lejeune, A. & Pain, R. H. (1998). Beta-lactamases as models for protein-folding studies. *Cell. Mol. Life Sci.* **54**, 372–377.
18. Jelsch, C., Mourey, L., Masson, J. M. & Samama, J. P. (1993). Crystal structure of *Escherichia coli* TEM1 beta-lactamase at 1.8 A resolution. *Proteins: Struct. Funct. Genet.* **16**, 364–383.
19. Strynadka, N. C., Martin, R., Jensen, S. E., Gold, M. & Jones, J. B. (1996). Structure-based design of a potent transition state analogue for TEM-1 beta-lactamase. *Nat. Struct. Biol.* **3**, 688–695.
20. Dideberg, O., Charlier, P., Wery, J. P., Dehottay, P., Dusart, J., Erpicum, T. *et al.* (1987). The crystal structure of the beta-lactamase of *Streptomyces albus* G at 0.3 nm resolution. *Biochem. J.* **245**, 911–913.
21. Minasov, G., Wang, X. & Shoichet, B. K. (2002). An ultrahigh resolution structure of TEM-1 beta-lactamase suggests a role for Glu166 as the general base in acylation. *J. Am. Chem. Soc.* **124**, 5333–5340.
22. Hecky, J. & Müller, K. M. (2005). Structural perturbation and compensation by directed evolution at physiological temperature leads to thermostabilization of beta-lactamase. *Biochemistry*, **44**, 12640–12654.
23. Stemmer, W. P. (1994). DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. *Proc. Natl Acad. Sci. USA*, **91**, 10747–10751.
24. Stemmer, W. P. (1994). Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature*, **370**, 389–391.
25. Zaccolo, M. & Gherardi, E. (1999). The effect of high-frequency random mutagenesis on *in vitro* protein evolution: a study on TEM-1 beta-lactamase. *J. Mol. Biol.* **285**, 775–783.
26. Huang, W. & Palzkill, T. (1997). A natural polymorphism in beta-lactamase is a global suppressor. *Proc. Natl Acad. Sci. USA*, **94**, 8801–8806.
27. Sideraki, V., Huang, W., Palzkill, T. & Gilbert, H. F. (2001). A secondary drug resistance mutation of TEM-1 beta-lactamase that suppresses misfolding and aggregation. *Proc. Natl Acad. Sci. USA*, **98**, 283–288.
28. Wang, X., Minasov, G. & Shoichet, B. K. (2002). Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *J. Mol. Biol.* **320**, 85–95.
29. Petrosino, J., Cantu, C., III & Palzkill, T. (1998). Beta-lactamases: protein evolution in real time. *Trends Microbiol.* **6**, 323–327.
30. O'Callaghan, C. H., Morris, A., Kirby, S. M. & Shingler, A. H. (1972). Novel method for detection of beta-lactamases by using a chromogenic cephalosporin substrate. *Antimicrob. Agents Chemother.* **1**, 283–288.
31. Vanhove, M., Lejeune, A., Guillaume, G., Virden, R., Pain, R. H., Schmid, F. X. & Frere, J. M. (1998). A collapsed intermediate with nonnative packing of hydrophobic residues in the folding of TEM-1 beta-lactamase. *Biochemistry*, **37**, 1941–1950.
32. Zaccolo, M. & Gherardi, E. (1999). The effect of high-frequency random mutagenesis on *in vitro* protein evolution: a study on TEM-1 beta-lactamase. *J. Mol. Biol.* **285**, 775–783.
33. 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.
34. Gassner, N. C., Baase, W. A. & Matthews, B. W. (1996). A test of the "jigsaw puzzle" model for protein folding by multiple methionine substitutions within the core of T4 lysozyme. *Proc. Natl Acad. Sci. USA*, **93**, 12155–12158.
35. Richardson, J. S. & Richardson, D. C. (1988). Amino acid preferences for specific locations at the ends of α-helices. *Science*, **240**, 1648–1652.
36. Kather, I., Jakob, R., Dobbek, H. & Schmid, F. X. (2008). Changing the determinants of protein stability from covalent to non-covalent interactions by *in-vitro* evolution: a structural and energetic analysis. *J. Mol. Biol.* **381**, 1040–1054.
37. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
38. Blazquez, J., Baquero, M. R., Canton, R., Alos, I. & Baquero, F. (1993). Characterization of a new TEM-type beta-lactamase resistant to clavulanate, sulbactam, and tazobactam in a clinical isolate of *Escherichia coli*. *Antimicrob. Agents Chemother.* **37**, 2059–2063.
39. Huang, W. & Palzkill, T. (1997). A natural polymorphism in beta-lactamase is a global suppressor. *Proc. Natl Acad. Sci. USA*, **94**, 8801–8806.
40. Orencia, M. C., Yoon, J. S., Ness, J. E., Stemmer, W. P. & Stevens, R. C. (2001). Predicting the emergence of antibiotic resistance by directed evolution and structural analysis. *Nat. Struct. Biol.* **8**, 238–242.
41. Farzaneh, S., Chaibi, E. B., Peduzzi, J., Barthelemy, M., Labia, R., Blazquez, J. & Baquero, F. (1996). Implication of Ile-69 and Thr-182 residues in kinetic characteristics of IRT-3 (TEM-32) beta-lactamase. *Antimicrob. Agents Chemother.* **40**, 2434–2436.
42. Nukaga, M., Mayama, K., Hujer, A. M., Bonomo, R. A. & Knox, J. R. (2003). Ultrahigh resolution structure of a class A beta-lactamase: on the mechanism and specificity of the extended-spectrum SHV-2 enzyme. *J. Mol. Biol.* **328**, 289–301.
43. Wunderlich, M., Martin, A. & Schmid, F. X. (2005). Stabilization of the cold shock protein CspB from *Bacillus subtilis* by evolutionary optimization of coulombic interactions. *J. Mol. Biol.* **347**, 1063–1076.
44. Sideraki, V., Huang, W., Palzkill, T. & Gilbert, H. F. (2001). A secondary drug resistance mutation of TEM-1 beta-lactamase that suppresses misfolding and aggregation. *Proc. Natl Acad. Sci. USA*, **98**, 283–288.
45. 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.

46. 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.
47. Miller, J. H. (1972). *Experiments in Molecular Genetics.* Cold Spring Harbor Laboratories, Cold Spring Harbor, NY.
48. Amann, E., Brosius, J. & Ptashne, M. (1983). Vectors bearing a hybrid trp-lac promoter useful for regulated expression of cloned genes in *Escherichia coli*. *Gene*, **25**, 167–178.
49. 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.
50. 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.
51. 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.
52. Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800.
53. Navaza, J. (2001). Implementation of molecular replacement in AMoRe. *Acta Crystallogr. Sect. D*, **57**, 1367–1372.
54. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D*, **60**, 2126–2132.
55. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D*, **53**, 240–255.

Teilarbeit C:

Jakob, R.P. and Schmid, F.X. (2008).
Energetic Coupling between Native-State Prolyl Isomerization and Conformational Protein Folding.
J. Mol. Biol. **377**, 1560-75.

doi:10.1016/j.jmb.2008.02.010 J. Mol. Biol. (2008) 377, 1560–1575

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Energetic Coupling Between Native-State Prolyl Isomerization and Conformational Protein Folding

Roman P. Jakob and Franz X. Schmid*

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

Received 30 November 2007; received in revised form 2 February 2008; accepted 6 February 2008
Available online 13 February 2008

In folded proteins, prolyl peptide bonds are usually thought to be either *trans* or *cis* because only one of the isomers can be accommodated in the native folded protein. For the N-terminal domain of the gene-3 protein of the filamentous phage fd (N2 domain), Pro161 resides at the tip of a β hairpin and was found to be *cis* in the crystal structure of this protein. Here we show that Pro161 exists in both the *cis* and the *trans* conformations in the folded form of the N2 domain. We investigated how conformational folding and prolyl isomerization are coupled in the unfolding and refolding of N2 domain. A combination of single-mixing and double-mixing unfolding and refolding experiments showed that, in unfolded N2 domain, 7% of the molecules contain a *cis*-Pro161 and 93% of the molecules contain a *trans*-Pro161. During refolding, the fraction of molecules with a *cis*-Pro161 increases to 85%. This implies that 10.3 kJ mol^{-1} of the folding free energy was used to drive this 75-fold change in the Pro161 *cis*/*trans* equilibrium constant during folding. The stabilities of the forms with the *cis* and the *trans* isomers of Pro161 and their folding kinetics could be determined separately because their conformational folding is much faster than the prolyl isomerization reactions in the native and the unfolded proteins. The energetic coupling between conformational folding and Pro161 isomerization is already fully established in the transition state of folding, and the two isomeric forms are thus truly native forms. The folding kinetics are well described by a four-species box model, in which the N2 molecules with either isomer of Pro161 can fold to the native state and in which *cis*/*trans* isomerization occurs in both the unfolded and the folded proteins.

Keywords: protein folding kinetics; Φ value; gene-3 protein; phage fd; protein stability

Edited by C. R. Matthews

**Corresponding author.* E-mail address: fx.schmid@uni-bayreuth.de.

Abbreviations used: G3P, gene-3 protein of the filamentous phage fd; N2 domain, N-terminal domain of G3P; N2′, N2 containing the stabilizing mutation Q129H; $[Urea]_M$, midpoint of a urea-induced unfolding transition; T_M, midpoint of a thermal unfolding transition; ΔH_D, van't Hoff enthalpy of denaturation at T_M; ΔG_D, Gibbs free energy of denaturation; *m*, cooperativity value of a denaturant-induced equilibrium unfolding transition; *A*, amplitude of a folding reaction; λ, apparent rate constant of a folding reaction; τ (=λ^{-1}), time constant of a folding reaction; k_{NU}, microscopic rate constant; m_{NU} (=$\delta \ln k_{NU}/\delta$[urea]), kinetic *m*-value of unfolding; β_T, Tanford β value; U_{cis}, unfolded form of N2′ with a *cis*-Pro161; U_{trans}, unfolded form of N2′ with a *trans*-Pro161; N_{cis}, native form of N2′ with a *cis*-Pro161; N_{trans}, native form of N2′ with a *trans*-Pro161; EDTA, ethylenediaminetetraacetic acid.

Introduction

The peptide bonds that precede proline (prolyl bonds) show peculiar properties. When structural constraints are absent, as in short peptides and unfolded proteins, they occur as an equilibrium mixture of the *cis* form and the *trans* form, in which *trans* is preferred to *cis*.[1–4] The conformations of folded proteins are usually compatible with only one of the isomers; therefore, only the molecules with the correct prolyl isomers can fold to the native conformation—the others must undergo prolyl isomerization in the course of refolding to reach the fully folded state.

The kinetics of the *cis*/*trans* interconversion is intrinsically slow because it involves the rotation about a partial double bond; therefore, prolyl iso-

merizations can retard protein folding by 1000-fold or more, in particular when proteins contain *cis* prolines in the folded state. Most of the work on prolyl isomerization has thus focused on its role as a rate-determining reaction in folding.[3–8] The biological importance of prolyl isomerization extends, however, beyond protein folding. Prolyl *cis*/*trans* isomerizations have been discovered in folded proteins[9–19] and assumed to be important for the function of such protein (e.g., in regulation).[20–22]

Energy is required to change the *cis*/*trans* equilibrium at prolyl bonds, and this energy must be diverted from the Gibbs free energy of stabilization that becomes available in the course of conformational folding. Protein folding and prolyl isomerization are thus thermodynamically linked. The extent of linkage (i.e., the shift in the prolyl *cis*/*trans* equilibrium constant upon folding) is determined by the amount of conformational energy that is transferred to a particular proline during folding.

Here we asked how conformational folding and prolyl isomerization are coupled in the unfolding and the refolding of the N-terminal domain of the gene-3 protein of the filamentous phage fd (G3P; N2 domain).[23,24] G3P mediates the infection of *Escherichia coli* by the phage,[25,26] and its N2 domain initiates this process by binding to an F pilus.[27] Pro161 of the N2 domain is at the tip of a β hairpin; in the crystal structure of G3P, it was found to be *cis* in molecule A of the asymmetric unit (Fig. 1). In molecule B, the region between Gly158 and Val162 was not resolved.[24] We investigated how Pro161 *cis*/*trans* isomerization is interrelated with conformational folding and found that 10.3 kJ mol^{-1} of conformational free energy is used to increase the *cis*/*trans* ratio at Pro161 during folding until it reaches a value of 85/15. The folding kinetics are well described by a box model in which the N2 molecules with either isomer of Pro161 can fold to the native state and in which *cis*/*trans* isomerization can occur in both the unfolded and the folded proteins, albeit with different rates.

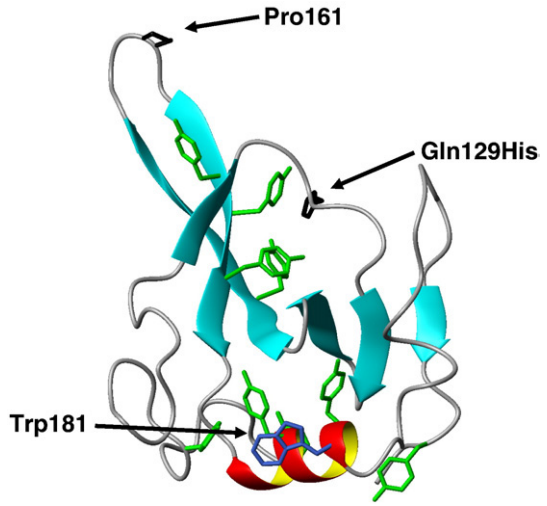


Fig. 1. Tertiary structure of the N2 domain of G3P (residues 102–205; coordinates from Holliger *et al.*[24]). The side chains of His129 and Pro161 (black), the single Trp181 (blue), and the nine Tyr residues (green) are shown in stick representation. The mutation Gln129His was modeled using the Swiss-PdbViewer[28] based on the structure of the wild-type protein. The figure was prepared using MOLMOL.[29]

Results

Stability of the N2 domain

The N2 domain comprises residues 102–205 of the phage G3P (Fig. 1). In intact G3P, the N2 domain interacts with the amino-terminal (N1) domain of this protein, and these interactions stabilize the folded state of N2 domain.[30,31] In isolation, N2 domain is marginally stable and prone to aggregation. N2 domain is stabilized by the Q129H mutation, which resulted from an *in vitro* selection for variants of the fd phage with increased stability. This mutation is located at the surface of N2 domain and increases the Gibbs free energy of denaturation (ΔG_D) of G3P by about 8 kJ mol^{-1}.[30] Here we used N2 domain with the Q129H mutation as our pseudo-wild-type protein and denoted it as N2′.

N2′ is a folded protein, and its CD spectrum shows a maximum near 195 nm, as expected for a protein with a high content of antiparallel β-sheet structure (data not shown). It elutes at the position of a monomer from a Superdex 75 gel-filtration column (data not shown). The thermal unfolding transition measured by CD is cooperative and shows a midpoint at 38.8 °C (pH 7.0) and an unfolding enthalpy of 320 kJ mol^{-1} (Table 1). It is independent of the protein concentration and is about 50% reversible when the unfolded protein is kept in the thermally unfolded form for 10 min. In the presence of the N1 domain, the T_M (midpoint of a thermal unfolding transition) value of N2′ increased to 50.8 °C, and the unfolding enthalpy increased to 580 kJ mol^{-1} (Fig. 2a).

The fluorescence of N2′ decreases by 40% between 0 M and 5 M urea, and the corresponding unfolding transition (Fig. 2b) shows a midpoint at 2.7 M urea and an *m* value of 6.4 kJ mol^{-1} M^{-1}. Similar values (midpoint at 2.6 M urea and $m = 6.7$ kJ mol^{-1} M^{-1}) were found when unfolding was followed by CD (Fig. 2c). The ΔG_D values extrapolated from the thermal unfolding transition and the two urea-induced unfolding transitions are very similar (Table 1), suggesting that the equilibrium unfolding of N2′ is adequately described as a two-state reaction. To elucidate the role of Pro161 in the folding mechanism, the P161A variant was also studied. It destabilized N2′ by 2.6 kJ mol^{-1} (Fig. 2c).

Unfolding and refolding kinetics of N2′

Unfolding of N2′ by 5.8 M urea is a monoexponential reaction with a time constant of 252 ms, and

Table 1. Stability data for N2′ derived from thermal and urea-induced unfolding transitions

	Thermal transition			Urea transition		
	T_M (°C)	ΔH_D (kJ mol^{-1})	$\Delta G_D^{15\ °C}$ (kJ mol^{-1})	$[Urea]_M$ (M)	m (kJ mol^{-1} M^{-1})	$\Delta G_D^{15\ °C}$ (kJ mol^{-1})
N2′ in G3P	50.8±0.08	580±19	43.4±1.4			
N2′	38.8±0.14	320±13	18.6±0.6	2.6±0.07[a]	6.7±0.2	17.9±0.5
				2.7±0.07[b]	6.4±0.2	17.1±0.6
N2′-P161A	35.4±0.13	285±9	14.3±0.5	2.3±0.07[b]	6.4±0.2	14.5±0.5

The stability parameters result from two-state analyses of the thermal unfolding transitions (columns 2–4) and urea-induced unfolding transitions (columns 5–7). For all proteins, the melting temperatures (T_M), the van't Hoff enthalpy of denaturation at T_M (ΔH_D), the midpoint of a urea-induced unfolding transition ($[urea]_M$), the cooperativity values (m), and the Gibbs free energies of denaturation at 15 °C ($\Delta G_D^{15\ °C}$) are given. N2′ in G3P refers to stability data for N2′ linked covalently with N1 in fragments 1–218 of the gene-3 protein. All measurements were performed in 100 mM potassium phosphate (pH 7.0).

[a] Urea-induced unfolding transition measured by CD at 222 nm.

[b] Urea-induced unfolding transition measured by the change in fluorescence at 340 nm after excitation at 280 nm.

its amplitude accounts for the entire difference in fluorescence between the native state and the unfolded state (Fig. 3a). Refolding at 0.8 M urea occurs in three phases. Two of them—the minor fast phase ($\tau = 100$ ms, $A = 0.09$ V) and the major intermediate phase ($\tau = 930$ ms, $A = 0.70$ V)—were measured after stopped-flow mixing (Fig. 3b). They account for about 80% of the entire change in fluorescence. The third phase, slow reaction, which was measured separately after manual mixing (Fig. 3c), showed a time constant of 90 s and accounted for 18% of the change in fluorescence, as observed in the corresponding equilibrium unfolding transition (cf. Fig. 2b).

Figure 4a shows the dependence of the measured rate constants of unfolding and refolding (the chevron plots) on urea concentration, and Fig. 4b shows the initial fluorescence values and the values observed after the completion of the fast and intermediate refolding phases, as obtained from the stopped-flow experiments. The amplitudes of these two phases are shown in Fig. 4c; those for the slow refolding and unfolding reactions (as obtained after manual mixing) are shown in Fig. 4d. They were measured under the same conditions as the equilibrium unfolding transition (Fig. 2b).

The fast and intermediate reactions follow two chevron plots with a common unfolding limb. In refolding, the two reactions differ in rate by about 30-fold (at 1–2 M urea), but their dependences on the urea concentration (the m values of refolding) are similar. The chevron plot for the fast reaction shows an apparent downward curvature below 1.5 M urea. The amplitude of the fast reaction is, however, very small (Fig. 4c), and the significance of this curvature remains unclear (Fig. 4a).

N2′ could also be unfolded by incubation at pH 2.0 in the absence of urea. The refolding rates at 0 M urea obtained from the refolding of the acid-unfolded protein agree well with those from the urea dilution experiments, and they indicate that the chevron plot for the major intermediate phase is linear between 0 M and 1 M urea (Fig. 4a).

In the transition region, identical rates were obtained from the unfolding and the refolding kinetics, suggesting that the folding of N2′ is reversible. It should be noted that neither chevron plot follows the equilibrium unfolding transition of N2′. The midpoints of 3.8 M urea for the fast reaction and 1.8 M urea for the intermediate reaction (Fig. 4a) do not coincide with the midpoint of the equilibrium unfolding transition, which is at 2.7 M urea (Fig. 2c).

The intermediate reaction dominated refolding at low urea concentration, but its amplitude decreased strongly with increasing urea concentration and vanished above 2.8 M urea (Fig. 4c). In fact, between 2 M and 4 M urea, only a minor fraction of the entire change in fluorescence occurred during the fast and the intermediate reactions (Fig. 4b). As a consequence, the final fluorescence values reached in the stopped-flow refolding and unfolding experiments showed pronounced hysteresis and deviated strongly from the equilibrium transition, which is shown as a red dashed line in Fig. 4b).

This strong deviation is caused by the slow folding reaction, which was measured after manual mixing. It showed a small amplitude at low urea concentration, which increased with the final urea concentration and reached a maximum in the transition region (Fig. 4d). At the transition midpoint (2.7 M urea), it accounted for ≥ 90% of the change in fluorescence, as observed in the equilibrium unfolding transition (cf. Fig. 2b). The slow reaction could also be monitored in unfolding experiments that ended in the transition region. Its amplitude shows a maximum near 3.5 M urea, and it vanishes above 5 M urea (Fig. 4d).

The rate of the slow reaction decreases slightly with urea concentration, passes through a minimum near 2.7 M urea, and increases again to reach a constant value above 4.2 M urea (Fig. 4a). The corresponding time constants (100 s at 0–1 M urea; 500 s at 2.7 M urea; 50 s at 4.2–4.8 M urea) are in the range expected for prolyl isomerization. In fact, the slow reaction is strongly accelerated by the prolyl isomerases SlyD[32] and cyclophilin-18[33] (Fig. 3c), suggesting that it is limited in rate by prolyl isomerization.

The amplitude of the slow refolding reaction becomes very small when the fluorescence is excited at 295 nm, instead of 280 nm. This indicates that it originates predominantly from the energy transfer from Tyr residues to the single Trp181.

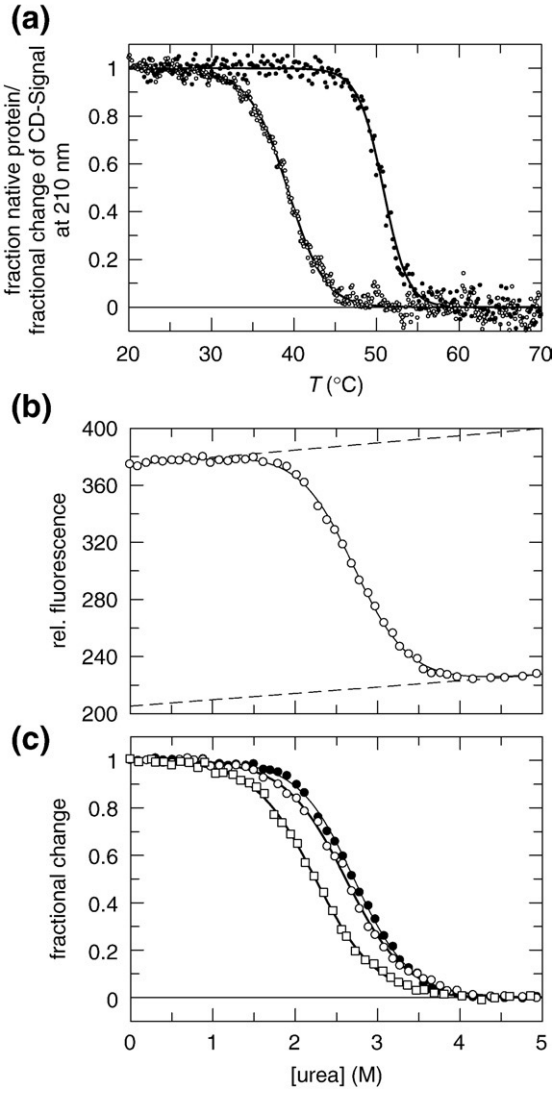


Fig. 2. (a) Thermal stabilities of the isolated N2′ domain (○) and the N2′ domain in G3P (●). Thermal transitions of the proteins were measured by CD at 210 nm (G3P) and 222 nm (N2′). The fractional changes as a function of temperature were obtained from two-state analyses of the data. The transitions were measured with 1 μM (G3P) and 4 μM (N2′) in 100 mM K phosphate (pH 7.0) at a path length of 1 cm. (b) Urea-induced equilibrium unfolding transition of N2′ at 15 °C measured by protein fluorescence at 340 nm after excitation at 280 nm. Broken lines represent the baselines for the native and the unfolded proteins. (c) The fractional changes in fluorescence at 340 nm (○) and in CD at 222 nm (●) of N2′, and the fractional change in fluorescence of N2′-P161A (□) derived from two-state analyses of the data are shown as a function of the urea concentration. Continuous lines represent the analyses of the data based on a two-state model. The stability data are given in Table 1. The protein concentration was 1.0 μM in 100 mM K phosphate (pH 7.0).

To examine whether this isomerization originates from Pro161 (the single *cis* proline of N2′), we measured the unfolding and refolding kinetics of the P161A variant, again after stopped-flow mixing and manual mixing. The corresponding rate constants are shown in Fig. 4e. The unfolding kinetics are virtually unchanged, but the refolding kinetics are strongly simplified by the P161A mutation. In the stopped-flow experiments, only a single fast refolding reaction could be detected, which was about twofold faster than the intermediate phase of the wild-type protein. The minimum of the chevron plot for N2′-P161A (2.3 M urea) coincided with the midpoint of its equilibrium unfolding transition (cf. Fig. 2c). The final values of the stopped-flow unfolding

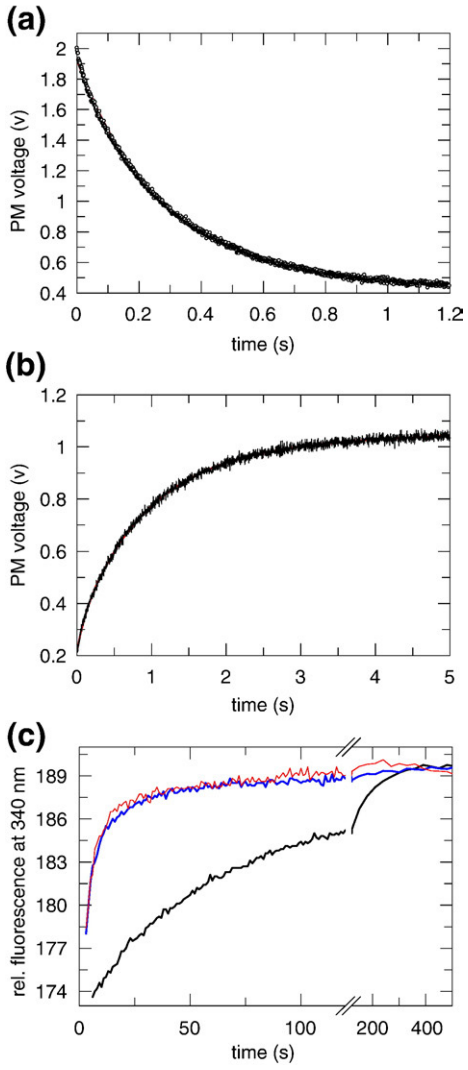


Fig. 3. Unfolding and refolding kinetics of N2′. (a) Monoexponential unfolding kinetics ($\tau = 252$ ms) at 5.8 M urea. (b) Biexponential refolding kinetics at 0.8 M urea ($\tau_1 = 100$ ms, $\tau_2 = 930$ ms) measured after stopped-flow mixing by the change in fluorescence above 320 nm after excitation at 280 nm. Continuous lines show fitted curves for single-exponential and double-exponential kinetics; the derived time constants (τ) are given above in parentheses. (c) Slow refolding (black) in 0.8 M urea ($\tau = 94$ s) in the absence of prolyl isomerases and in the presence of 1 μM SlyD from *E. coli* (blue; $\tau = 5.0$ s) or 1 μM human cyclophilin-18 (red; $\tau = 5.4$ s), followed by the change in fluorescence at 340 nm after excitation at 280 nm. Kinetics were measured at 15 °C in 0.1 M K phosphate (pH 7.0).

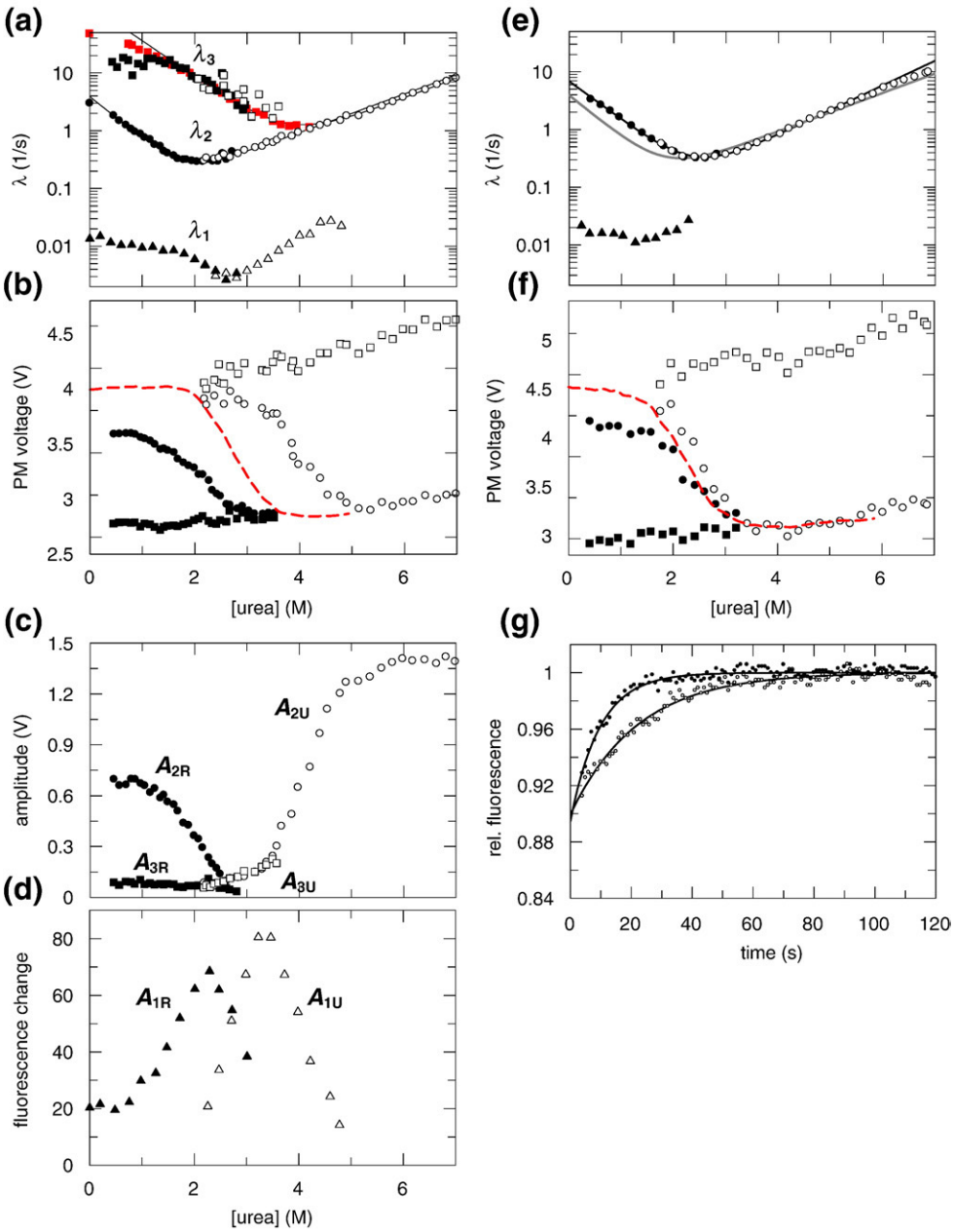


Fig. 4. Refolding kinetics (filled symbols) and unfolding kinetics (open symbols) of (a) N2′ and (e) N2′-P161A. The apparent rate constants λ and their amplitudes *A* are shown as a function of the urea concentration. The folding kinetics were measured after stopped-flow single mixing (●,■,○,□) or after manual 10-fold dilution of the protein (▲,△). Red squares represent refolding kinetics measured after a 2.3-s unfolding pulse. The λ values at 0 M urea were obtained after refolding of N2′ denatured at pH 2.0. The continuous lines in (a) and (e) represent fits to the unfolding and refolding rates based on a linear two-state model. The kinetic parameters derived from the analysis are given in Table 2. The gray line in (e) represents the fit for the intermediate reaction of N2′, taken from (a). Initial (■,□) and final (●,○) values of the unfolding (open symbols) and the refolding (closed symbols) kinetics of (b) N2′ and (f) N2′-P161A as a function of the urea concentration. Red dashed lines indicate the course of the urea-induced equilibrium unfolding transition. (c) Amplitudes of the fast (■,□) and intermediate (●,○) phases of the refolding and unfolding of N2′ as a function of the concentration of urea. The indices R and U refer to refolding and unfolding, respectively. (d) Amplitudes of the slow refolding (▲) and unfolding (△) phases of N2′ as a function of the concentration of urea, measured after manual mixing using the same fluorimeter settings as in equilibrium unfolding (cf. Fig. 2b). (g) Refolding of N2′-P161A in the absence (○) and in the presence (●) of 1 μM SlyD*. The folding kinetics were measured at 15 °C in 0.1 M K phosphate and 0.6 M urea (pH 7.0) by the change in fluorescence at 340 nm after excitation at 280 nm. The final concentration of protein for all measurements was 1.0 μM in 100 mM K phosphate (pH 7.0).

and refolding kinetics of the P161A variant also deviate slightly from the equilibrium unfolding transition (Fig. 4f), but the noncoincidence is much smaller than in the case of the wild-type protein (Fig. 4b). It is caused by a minor slow reaction that could be followed after manual mixing. Its time constant is

about 60 s, independent of the urea concentration (Fig. 4e). Its amplitude accounts for 11–16% of the total change in fluorescence, and it is accelerated by the prolyl isomerase SlyD (Fig. 4g). This reaction possibly reflects the contributions of one or several of the eight *trans* prolines to the folding kinetics. They could not be detected for the wild-type protein because its slow folding is dominated by the strong contribution of Pro161 isomerization.

Pro161 isomerization in the unfolded protein

The *cis*/*trans* isomerization of Pro161 in the unfolded protein was measured by double-mixing experiments.[5,34] At 5.0 M urea, N2′ unfolds in a fast monoexponential reaction with $\tau = 0.5$ s, which implies that after 2.3 s, conformational unfolding is 99% complete, whereas the isomeric state of Pro161 should still be the same as in the folded protein. In a stopped-flow double-mixing experiment, a sample of N2′ was unfolded at 5.0 M urea for 2.3 s, and then refolded at 0.8 M urea. The corresponding refolding reaction (Fig. 5a) consisted of two phases, with $\tau =$ 34 ms and $\tau = 830$ ms, respectively. These values agreed with those measured for the fast and the intermediate phases in the single refolding experiments (Fig. 4a). The amplitudes were, however, strongly different, and the fast refolding reaction dominated, accounting for 88% of the entire amplitude. When the unfolding step was prolonged from 2.3 s to 200 s, the amplitude of the very fast refolding reaction decreased and reached a value of 12%, as observed for the same reaction in the single refolding experiments. In a reciprocal fashion, the amplitude of the fast reaction increased to 88% (Fig. 5a).

Next, we varied the time of unfolding between 20 ms and 200 s. The variation of the amplitude of the fast refolding reaction as a function of the time of unfolding (Fig. 5b) shows that the fast refolding molecules are produced within the first 2 s by the fluorescence-monitored conformational unfolding reaction. They reached a maximal population after 2.5 s and then decreased again because they were converted into the molecules that refold via the intermediate reaction. The decrease in the fast folding molecules and the increase in the amplitude of the intermediate refolding reaction are reciprocal reactions, and both show the time constant of the proline-limited unfolding reaction (Fig. 5b). These double-mixing experiments demonstrate that the fast refolding reaction originates from the molecules with a *cis*-Pro161 and that the intermediate refolding reaction originates from molecules with a *trans*-Pro161.

The formation of the molecules that refold on the proline-limited track (Fig. 5b) should show a lag because they are formed by a sequential reaction, in which prolyl isomerization in the unfolded protein is preceded by conformational unfolding. This lag should lead to a kinetic phase with the time constant of the conformational unfolding reaction ($\tau = 500$ ms) and a negative amplitude. A reaction with this time constant is indeed observed, but its amplitude is positive (Fig. 5b). This suggests that a small fraction of molecules with an incorrect (*trans*) prolyl isomer preexists in the folded protein. Their conformational unfolding thus leads directly to molecules that refold on the proline-limited track.

Figure 5c shows the fractional amplitudes of the fast and the intermediate refolding reactions as a function of the duration of unfolding. This alternative representation reveals the two different processes that took place during unfolding. Within the first 2 s, conformational unfolding occurred and produced, in parallel and with identical rates, two species: a major species with correct prolyl isomers that refolds rapidly on the fast direct track and a minor species with incorrect prolyl isomers that refolds slowly on the proline-limited track. The ratio between the two amplitudes remained virtually constant during the first 2 s of unfolding because the proline-limited interconversion between the two species is slow. It could not be influenced by preincubating N2′ with stabilizing cosolutes such as glycerol or ethylene glycol or by adding urea at nondenaturing concentrations (data not shown). In the subsequent 200 s, most of the unfolded protein molecules underwent *cis*/*trans* isomerization at Pro161; finally, an equilibrium mixture was reached, in which only 7% of the unfolded molecules still contained a *cis*-Pro161 and thus could refold on the fast direct track.

It was difficult to characterize the minor refolding reaction of these molecules in the single-mixing refolding experiments (cf. Fig. 4a). To improve the quality of these data, we repeated them with samples that had been unfolded for 2.3 s only, as in Fig. 5a. Thus, >80% of all unfolded molecules still contained a *cis*-Pro161, and their fast refolding could be measured very well. The resulting rate constants are included in Fig. 4a. They agree with those measured initially, but show a much higher accuracy.

At pH 2.0, conformational unfolding of N2′ occurs with a time constant of 2 ms (data not shown). It is thus about 10^4-fold faster than prolyl isomerization and is completed within the shortest delay time accessible in the double-mixing experiments (20 ms). The refolding traces obtained at pH 7 after 20–2000 ms of unfolding at pH 2 were all biphasic and accounted for the full change in fluorescence as expected for completely unfolded N2′ molecules. The ratio of the amplitudes of the fast and the intermediate refolding reactions (90:10) was approximately constant during the first second of unfolding (Fig. 5d). This validates our conclusion that a mixture of species with correct and incorrect prolines preexists in folded N2′. Interestingly, the fraction of fast folding molecules present after long-term unfolding at pH 2.0 (16%) was about 2-fold higher than the corresponding fraction observed after unfolding by 5 M urea at pH 7.0 (7%). This indicates that the protonation of Asp160 (at pH 2) shifts the *cis*/*trans* equilibrium at the Asp160-Pro161 bond slightly towards the *cis* form.

The unfolding of the P161A variant was probed by double-jump refolding assays as well (Fig. 5e). It was initiated by a dilution to 5.0 M urea; after va-

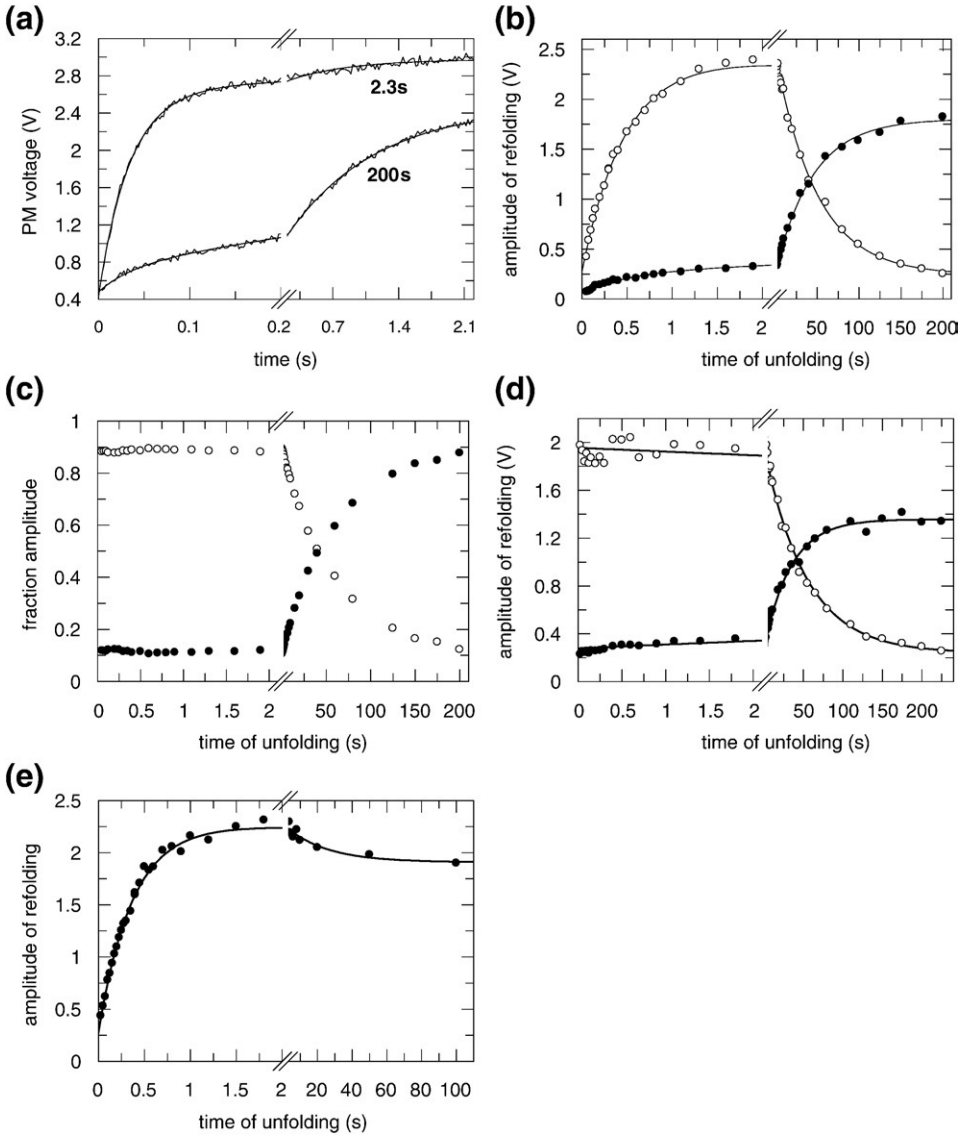


Fig. 5. Stopped-flow double-mixing experiments to probe Pro161 isomerization in unfolded N2′. (a) Refolding kinetics of N2′ after 2.3 s (upper curve; $\tau_1 = 33.6$ ms, $\tau_2 = 830$ ms) and 200 s (lower curve; $\tau_1 = 34.7$ ms, $\tau_2 = 950$ ms) of unfolding by 5.0 M urea in 100 mM K phosphate (pH 7.0). (b) Amplitudes of the fast (○; $\tau_1 = 470$ ms, $\tau_2 = 47.6$ s) and the intermediate (●; $\tau_1 = 535$ ms, $\tau_2 = 45.5$ s) refolding reactions of N2′ as a function of the duration of unfolding in 5.0 M urea and 100 mM K phosphate (pH 7.0). (c) Relative amplitudes of the fast (○) and the intermediate (●) refolding reactions as a function of the duration of unfolding. (d) Amplitudes of the fast (○; $\tau = 52.6$ s) and the intermediate (●; $\tau = 55.5$ s) refolding reactions of N2′ as a function of the duration of unfolding in 100 mM glycine (pH 2.0). (e) Amplitude of the fast refolding reaction of N2′-P161A ($\tau_1 = 370$ ms, $\tau_2 = 24$ s) as a function of the duration of unfolding. In all stopped-flow double-mixing experiments, unfolding was initiated by an 11-fold dilution of native protein. Refolding was induced by a further 6-fold dilution to final folding conditions of 1 μM protein and 100 mM K phosphate (pH 7.0), followed by the increase in protein fluorescence above 320 nm. Kinetics were measured at 15 °C. Continuous lines show the curve fits for monoexponential or biexponential functions; the derived time constants (τ) are given in parentheses.

riable times, samples were withdrawn and refolded in 0.8 M urea. Unlike in the case of the pseudo-wild-type protein, a transient population of a very fast folding species was not observed. The molecules that refold with a time constant of 0.5 s are formed directly in the conformational unfolding reaction. It shows the same time constant ($\tau = 0.37$ s) as the unfolding reaction that was monitored directly by the decrease in protein fluorescence ($\tau = 0.44$ s) in the single-mixing unfolding experiments (Fig. 4e).

The intermediate refolding reaction of the pseudo-wild-type protein originated from the molecules with a *trans*-Pro161, which were formed slowly by the spectroscopically silent *cis*-to-*trans* isomerization at this proline. The absence of such a reaction in the unfolding of the P161A mutant suggests that the Asp160-Ala161 bond in this variant is *trans* in both the native and the unfolded forms, as expected for a nonprolyl peptide bond. When, for the P161A variant, the duration of unfolding was extended to

100 s, the amplitude of the 0.5-s refolding phase decreased by about 15% and, in parallel, the slow refolding reaction ($\tau = 24$ s) appeared. This confirms that this minor reaction in the unfolded protein is, in fact, caused by the slow *cis*/*trans* equilibration at prolines that were *trans* in the folded mutant protein.

Pro161 isomerization in the folded protein

The kinetic analysis in Fig. 4 and the double-mixing experiments in Fig. 5 suggest that the refolding of the N2′ domain proceeds on two parallel pathways. Seven percent of the unfolded molecules contain correct prolines and refold very fast on the direct track ($\tau = 30$ ms at 1 M urea). Ninety-three percent contain at least a single incorrect prolyl isomer, in most cases a *trans*-Pro161. They first react to an intermediate ($\tau = 1$ s at 1 M urea) and then, in the final step ($\tau = 110$ s at 1 M urea), undergo *trans*-to-*cis* isomerization at Pro161 to reach the final native state.

Previously, we developed a generic two-step assay for following the formation of fully folded protein molecules during a refolding reaction.[35] This assay is based on the fact that those 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. However, partially folded molecules, which have not yet passed the final transition state, unfold rapidly.

In the two-step assay for native molecules, the unfolded protein is first mixed with refolding buffer to initiate refolding. Then, after variable time intervals, samples are withdrawn and 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 folded N2′ molecules during refolding at 1.0 M urea. The unfolding assays were performed in 5.0 M urea at 15 °C—conditions under which native N2′ unfolds in a monoexponential reaction ($\tau = 0.5$ s; cf. Fig. 4a). The time course of formation of the molecules that unfold with $\tau = 0.5$ s is shown in Fig. 6a. It consists of three phases. A 4% increase during the dead time of the double-mixing experiment (20 ms) originates from the fast refolding molecules. The major formation of slowly unfolding molecules occurs in the intermediate phase ($\tau = 1.1$ s) and accounts for 78% of the entire amplitude of the unfolding assay. An additional phase with 18% amplitude is created by the slow refolding reaction ($\tau = 46$ s). It reflects the further increase in fluorescence that accompanies the slow proline-limited refolding reaction.

These results show that the molecules with Pro161 that are still in the incorrect *trans* conformation, which are formed in the intermediate phase of refolding, are native-like, as judged by their slow unfolding. In other words, they have already crossed the major kinetic barrier of refolding and thus, kinetically, belong to the ensemble of native species. The slow proline-limited reaction is also monitored in the assay for native molecules because the slow conversion of the intermediate form to the native form is accompanied by a small increase in protein fluorescence under these conditions (1.0 M urea), which is reversed in the unfolding assays.

The refolding of the P161A variant was also probed by the assay for fully folded molecules. No hidden kinetic phases were detected for this species. At 0.5 M urea, 72% of the molecules reached the folded state with a time constant of 0.4 s, and 28% of the molecules reached the folded state with a time constant of 30 s (Fig. 6b). These values are similar to those measured directly in the single-mixing folding kinetics (Fig. 4e).

Stabilities of the N2′ forms with *cis*-Pro161 or *trans*-Pro161

The *cis* and *trans* forms of folded N2′, N_{cis} and N_{trans}, are kinetically isolated because they differ by about 30-fold in the rate of refolding and because Pro161 isomerization in both the unfolded and the folded proteins is two orders of magnitude slower than the conformational unfolding and refolding reactions of N_{cis} and N_{trans} (Fig. 4a). The chevron plots and the final values of the fast and the

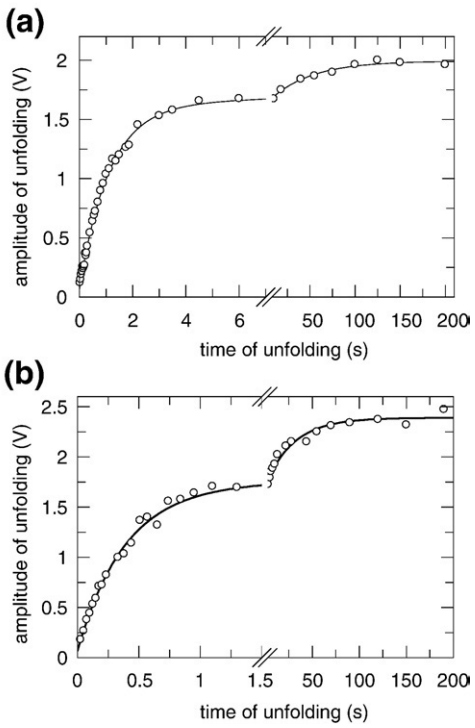


Fig. 6. Time course of the formation of native molecules during the refolding of (a) N2′ and (b) P161A-N2′. The kinetics were measured by stopped-flow interrupted refolding experiments. The amplitudes of unfolding by 5.0 M urea in 100 mM K phosphate are shown as a function of the duration of refolding. The fitted curves (continuous lines) give time constants (τ) of 1.12 s and 47.6 s for N2′, and 0.39 s and 30 s for N2′-P161A.

intermediate folding phases in Fig. 4a and b can therefore be used to determine the stability of N_{cis} relative to U_{cis}, and the stability of N_{trans} relative to U_{trans} (Table 2). The analysis of the chevron plots according to a linear two-state model[37] gave midpoints of 3.8 M urea for $N_{cis} \leftrightarrow U_{cis}$ and 1.8 M for $N_{trans} \leftrightarrow U_{trans}$, and the corresponding ΔG_D values are 20.7 kJ mol^{-1} and 10.9 kJ mol^{-1}.

Very similar values were obtained from the analysis of the amplitude data in Fig. 4b. The fluorescence values reached after the fast unfolding reaction trace the $N_{cis} \leftrightarrow U_{cis}$ reaction, and those reached after the intermediate refolding reaction trace the $N_{trans} \leftrightarrow U_{trans}$ reaction (Fig. 4b). An analysis of these curves gave ΔG_D values of 22.8 kJ mol^{-1} and 11.1 kJ mol^{-1} for the two unfolding reactions (Table 2), indicating that the conformational stabilities of N_{trans} and N_{cis} differ by 10.9 kJ mol^{-1}.

The corresponding analyses for the P161A variant gave similar *m* and β_T values as for the *cis* and *trans* isomers of the pseudo-wild-type protein. The P161A variant is about 7 kJ mol^{-1} less stable than the *cis*—but 3 kJ mol^{-1} more stable than the *trans*—isomer of the pseudo-wild-type protein (Table 2).

Kinetic mechanism for the coupling between conformational folding and prolyl isomerization

In combination, the single-mixing and double-mixing experiments (Figs. 4–6) indicate that species with *cis* and *trans* isomers of Pro161 coexist in the folded N2′ domain and that unfolding and refolding are accompanied by strong shifts in the *cis*/*trans* ratio at this proline. The four-species box model in Fig. 7 provides the simplest kinetic mechanism for the coupling between conformational folding and prolyl isomerization in the folded and the unfolded states of a protein. A similar box model has been employed in the analysis of prolyl isomerism in the folded and the unfolded forms of staphylococcal nuclease by NMR spectroscopy.[38,39]

The box model consists of four reactions with eight rate constants. The closed nature of the box provides one relation between these rate constants and reduces the number of independent variables to seven. However, only six experimental parameters (three λ values and three amplitudes) are available and, therefore, the system is underdetermined.[40] For N2′, the slow folding reaction had to be measured separately by manual mixing, which complicated the calculation of relative amplitudes. As a consequence, we cannot rigorously use the box model in Fig. 7 to derive the microscopic rate constants from the measured rates and amplitudes, and to test the validity of the model. Rather, we will use a hybrid approach to evaluate whether the folding mechanism of N2′ is adequately described by the four-species box model.

The double-mixing experiments in Fig. 5 provided particularly important information. After the ultrafast conformational unfolding at pH 2 (Fig. 5d), a major fast refolding form and a minor slow refolding form were present immediately, which demonstrated that the folded state is heterogeneous. The percentages of molecules with *cis*-Pro161 and *trans*-Pro161 (15% N_{cis} and 85% N_{trans}, respectively) could be derived from the relative amplitudes of the fast and the intermediate refolding reactions after very short unfolding times, by taking into account that the N_{cis} and N_{trans} species differ slightly in their fluorescence intensity, as revealed in the unfolding assays of Fig. 6a. They had shown that, in the folding of the U_{trans} molecules, 81% of the fluorescence changes during the $U_{trans} \rightarrow N_{trans}$ reaction ($\tau = 1.1$ s) and 19% of the fluorescence changes during the subsequent $N_{trans} \rightarrow N_{cis}$ reaction ($\tau = 48$ s).

The equilibrium distribution of U_{cis} and U_{trans} molecules in unfolded N2′ (7% U_{cis} and 93% U_{trans}) could be derived from the relative amplitudes for fast and intermediate refoldings, as observed after long times of unfolding (Fig. 5) when the *cis*/*trans* reaction in the unfolded protein had reached its equilibrium. In combination, the double-mixing unfolding and refolding experiments thus gave the rates and the equilibrium constants for the Pro161 *cis* ↔ *trans* reactions in both the N and the U states

Table 2. Unfolding and refolding kinetics of N2′ and N2′-P161A

Analysis	k_{UN} (s^{-1})	k_{NU} (s^{-1})	m_{UN} (kJ mol^{-1} M^{-1})	m_{NU} (kJ mol^{-1} M^{-1})	β_T	m (kJ mol^{-1} M^{-1})	$[Urea]_M$ (M)	$\Delta G_D^{15\ °C}$ (kJ mol^{-1})
N2′								
Chevron (λ_3)	36.1±0.05	1.69±0.05	−3.4±0.06	1.9±0.04	0.64	5.3±0.1	3.8±0.1	20.7±0.4
Amplitude (λ_3)						5.8±0.7	3.9±0.4	22.8±3.0
Chevron (λ_2)	0.70±0.05	1.99±0.05	−4.1±0.03	1.8±0.05	0.68	6.0±0.1	1.8±0.1	10.9±0.7
Amplitude (λ_2)						5.6±0.4	2.0±0.1	11.1±0.7
N2′-P161A								
Chevron	1.47±0.03	2.18±0.16	−3.7±0.03	2.4±0.03	0.61	6.1±0.1	2.4±0.1	14.6±0.3
Amplitude						5.4±0.3	2.6±0.1	13.8±0.6

The kinetic parameters were determined at 15 °C (pH 7.0) from the kinetic data shown in Fig. 4. The microscopic rate constants of refolding k_{UN} at 1.0 M and of unfolding k_{NU} at 5.0 M urea are given. m_{UN} and m_{NU} are the kinetic *m* values for refolding and unfolding, respectively. The transition midpoint $[urea]_M$ is calculated as: $[urea]_M = \ln(k_{NU}/k_{UN})/(m_{UN} - m_{NU})$. The cooperativity values (*m*) are calculated from the kinetic *m* values ($m = m_{NU} - m_{UN}$). The Gibbs free energies of denaturation at 0 M urea and 15 °C ($\Delta G_D^{15\ °C}$) were derived from the ratio of the rate constants k_{NU}/k_{UN} from the chevron plots in Fig. 4a and e and from a two-state analysis[36] of the final values of the fast unfolding reaction (λ_3) or the intermediate refolding reaction (λ_2) in Fig. 4b. The data for P161A-N2′ are from the corresponding analyses of the data in Fig. 4e and f.

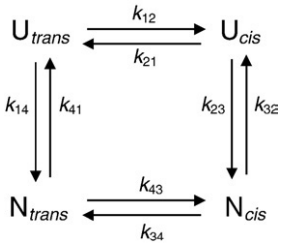


Fig. 7. Kinetic model for the folding and prolyl isomerization of N2′.

(the $U_{trans} \leftrightarrow U_{cis}$ and $N_{trans} \leftrightarrow N_{cis}$ reactions in the box model of Fig. 7). For the calculations, we assume that these isomerizations are urea-independent.

To calculate the microscopic rate constants for the two conformational folding reactions ($U_{trans} \leftrightarrow N_{trans}$ and $U_{cis} \leftrightarrow N_{cis}$), we assumed that under refolding conditions, λ_3 (Fig. 4a) reflects the $U_{cis} \leftrightarrow N_{cis}$ reaction and λ_2 reflects the $U_{trans} \leftrightarrow N_{trans}$ reaction. This simplification seems justified because λ_2 and λ_3 differ by >30-fold. Also, we assumed that under unfolding conditions, U_{cis} and U_{trans} unfold with identical rates (λ_2) and that the logarithms of all rate constants depend linearly on the urea concentration.

We used these rate constants—fluorescence values of 0 for U_{cis} and U_{trans}, 0.81 for N_{trans}, and 1.0 for N_{cis}—and the differential equations for the individual species in Fig. 7 to calculate the kinetics of the fluorescence changes during unfolding and refolding as a function of the urea concentration. The calculated kinetic curves were then analyzed as the sum of exponential functions, and the obtained rates and amplitudes were compared with the experimental values (Fig. 8).

Above 1.5 M urea, the agreement between the experimental and the calculated values for λ_2 and λ_3 is perfect, as expected, because the experimental values served as the input for the calculation and because λ_2 and λ_3 were well separated. At low urea concentrations, the experimental log λ_3 values deviate slightly from linearity, possibly because intermediates start to form on the $U_{cis} \rightarrow N_{cis}$ path. The agreement for λ_1 is good in the baseline regions, but a small deviation is noted in the transition region near 2.6 M urea. This suggests that the rates of prolyl isomerization in the folded protein are slightly urea-dependent. In fact, when the rate of the Pro161 *trans*→*cis* reaction during refolding is assumed to decrease with urea concentration in the same fashion (with the same *m* value) as λ_1 in Fig. 4a, the agreement between the experimental and the calculated data is improved (see the dashed line for λ_1 in Fig. 8a).

Figure 8b and c compares the experimental amplitudes with those calculated based on the box model in Fig. 7. Since the amplitudes for the fast and the intermediate phases, on the one hand, and the

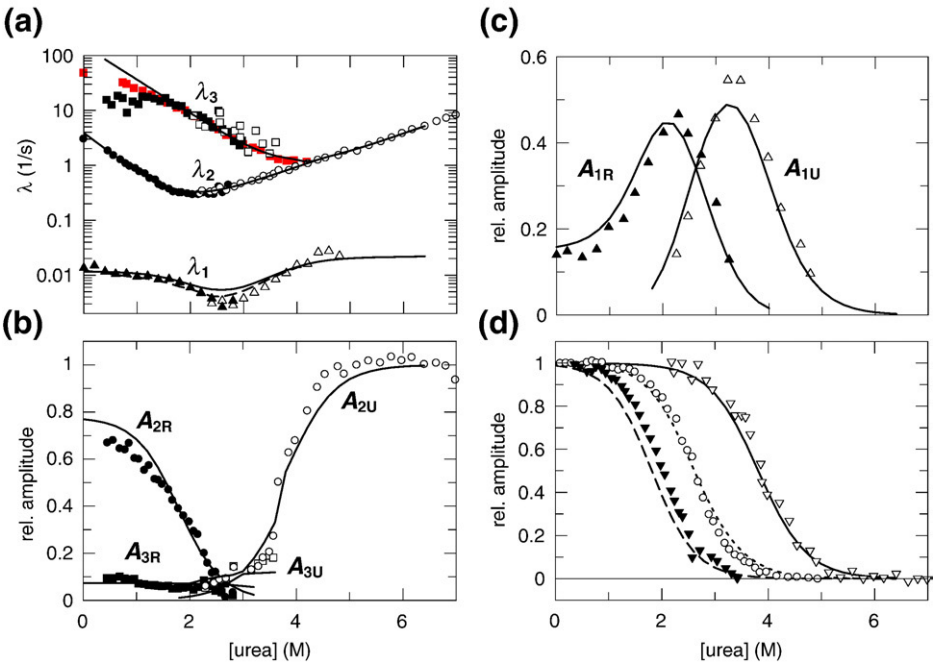


Fig. 8. Comparison of experimental and simulated data. (a) Apparent rate constants. (b) Normalized fluorescence amplitudes of the fast (■,□) and the intermediate (●,○) folding phases. (c) Normalized fluorescence amplitudes of the slow phase (▲,△). Filled symbols refer to refolding; open symbols refer to unfolding. In (a)–(c), the experimental data are taken from Fig. 4a, c, and d. Continuous lines represent the data simulated based on the box model in Fig. 7; the dashed line in panel (a) represents the simulated data for λ_1 including slight urea dependence. (d) Normalized unfolding transitions calculated based on the box model for $N_{cis} \leftrightarrow U_{cis}$ (—), $N_{trans} \leftrightarrow U_{trans}$ (– –), and $(N_{cis}+N_{trans}) \leftrightarrow (U_{cis}+U_{trans})$ (—). These transitions are compared with the fluorescence values reached in unfolding after the fast $N_{cis} \rightarrow U_{cis}$ unfolding reaction (▽) and after the intermediate refolding reaction (▼), and with the equilibrium unfolding transition of N2′ (○). The amplitude data were taken from Fig. 4b and normalized; the equilibrium data were taken from Fig. 2c. The indices R and U refer to refolding and unfolding, respectively.

amplitudes for the slow phase, on the other hand, originated from separate experiments, we compared them individually with the calculated profiles. The amplitude of the slow reaction (A_1) is given relative to the fluorescence change measured in the urea-induced equilibrium transition in Fig. 2b. The amplitudes A_2 and A_3 are given relative to the difference between the two baselines for the unfolded and the native proteins as defined by the endpoints of the stopped-flow kinetics in Fig. 4b. The calculated amplitude profiles closely follow the experimental amplitudes. In particular, the maxima of the profiles for A_1 in refolding and unfolding are well reproduced by the calculations (Fig. 8c).

The calculated rate constants were also used to derive the separate equilibrium constants for the unfolding of N_{cis} and N_{trans}, as well as the apparent equilibrium constant for the unfolding of $N_{cis}+N_{trans}$ (Fig. 8d). The computed equilibrium constant N_{cis}/U_{cis} agrees well with the final values reached experimentally after the fast $N_{cis}\rightarrow U_{cis}$ unfolding reaction (Fig. 4b), and the computed values for N_{trans}/U_{trans} agree with the endpoints of the intermediate refolding reaction in Fig. 4b. The small deviation between the calculated transition for $N_{trans}\leftrightarrow U_{trans}$ reaction and the final fluorescence after the intermediate refolding reaction is probably related to the poor definition of the experimental baseline at low urea concentration (Fig. 3b). The computed equilibrium constant $(N_{cis}+N_{trans})/(U_{cis}+U_{trans})$ coincides with the measured overall equilibrium unfolding transition in Fig. 2c. The good reproduction of the equilibrium unfolding curve by the calculated rate constants, as well as the agreement between the experimental and the calculated amplitudes, provides good evidence that the four-species model in Fig. 7 describes the coupling between conformational folding and prolyl isomerization of N2′ in an adequate manner.

Figure 9 provides an energy diagram for the folding of the *cis* and the *trans* forms of the N2 domain at 1 M urea. It is based on the equilibrium and kinetic data in Table 2. In the unfolded protein, the *cis* form is 6.2 kJ mol^{-1} less stable than the *trans* form; however, in the folded form, *cis* is 4.2 kJ mol^{-1} more stable than *trans*. This increase in the relative stability of the *cis* form has already occurred when the transition state of refolding was reached and was responsible for the observed 30-fold difference between the microscopic rates of refolding of the U_{cis} and U_{trans} forms. The rates of unfolding are the same for N_{cis} and N_{trans} because the energy differences between the two isomeric forms are the same in the folded state and in the transition state.

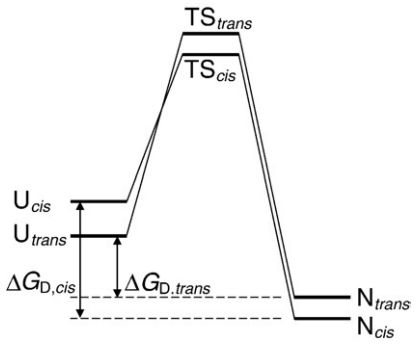


Fig. 9. Energy diagram for the folding of the N2 domain. It is based on the equilibrium and kinetic data in Table 2. $\Delta G_{D,trans}$ and $\Delta G_{D,cis}$ are the ΔG_D values for the $N_{trans}\leftrightarrow U_{trans}$ and $N_{cis}\leftrightarrow U_{cis}$ reactions, respectively.

Discussion

Proline heterogeneity in the folded form of the N2′ domain of the phage gene-3 protein

In folded proteins, prolyl peptide bonds are usually thought to be well ordered, being either *trans* or *cis* in virtually all molecules, because only one of the isomers is compatible with the native fold of the protein. Exceptions have been found in a few cases, where conformational heterogeneity at proline residues could be identified by NMR spectroscopy.[9–18,20,22]

The N2′ domain of the phage gene-3 protein accommodates in its folded form both the *cis* and the *trans* isomers of Pro161. In the crystal structures of the gene-3 protein,[23,24] this proline is *cis* in one molecule and not resolved in the other, and our kinetic analysis of the folding mechanism showed that, in fact, two folded forms, N_{trans} (15%) and N_{cis} (85%), coexist in solution. In the unfolded state, the corresponding fractions are 93% U_{trans} and 7% U_{cis}. The equilibrium constant thus changes 75-fold during refolding, which implies that 10.3 kJ mol^{-1} of the folding free energy is used to drive this shift in the *cis*/*trans* equilibrium in the course of conformational folding.

The box model in Fig. 7 and the energy diagram in Fig. 9 explain the kinetic and thermodynamic relations between prolyl isomerization and chain folding. The conformational folding reactions are much faster than the prolyl isomerization reactions, and this kinetic uncoupling enabled us to determine the folding kinetics and the stabilities of the forms with the *cis* and the *trans* isomers of Pro161 separately. The conformational stability of N_{cis} (relative to U_{cis}) is, in fact, 10 kJ mol^{-1} higher than the stability of N_{trans} (relative to U_{trans}). Thus, there is a coupling energy of 10 kJ mol^{-1} between conformational folding and prolyl isomerization, which implies that the native conformation stabilizes the *cis* isomer of Pro161 by 10 kJ mol^{-1} and, in turn, *cis*-Pro161 increases the stability of the folded conformation of N2′ by the same amount.

The 75-fold increase during the folding of the *cis*/*trans* equilibrium constant at Pro161 is caused by changes in both rate constants. The *trans*→*cis* rate increased 7-fold and the *cis*→*trans* rate decreased 11-fold in the folded state. The native environment thus lowers the energy of the *cis* form and increases the energy of the *trans* form.

Incorrect prolyl isomers as temporary backbone mutations

We view *trans*-Pro161 as a temporary backbone mutation of *cis*-Pro161, the "wild-type residue." The two forms coexist at equilibrium, and their conformational unfolding and refolding reactions could be measured separately because they are ≥50-fold faster than the *cis*↔*trans* equilibrations in N or U. The *cis*→*trans* backbone mutation affects only the refolding kinetics, which is equivalent to a Φ value[41] of 1. In other words, the local environment of the mutated backbone segment at Pro161 is native-like in the folding transition state, and the thermodynamic coupling energy of 10 kJ mol^{-1} between conformational folding and Pro161 *cis*/*trans* equilibration is already fully established in the transition state of folding. Unlike in the case of classical mutant analyses, the analysis performed here is rigorous because the "wild-type" *cis* form and the "mutant" *trans* form are connected by a chemical reaction, prolyl *cis*/*trans* isomerization.

The interpretation of the folding kinetics of the *cis* and the *trans* forms of N2′ is supported by the results obtained with the Pro161Ala variant. The mutation *cis*-Pro→*trans*-Ala destabilized N2′ by about 7 kJ mol^{-1}, and this destabilization decreased the refolding rate by 20-fold without changing the rate of unfolding (i.e., the Φ value for this mutation is also 1). The mutation *trans*-Pro→*trans*-Ala stabilized N2′ by about 3 kJ mol^{-1}, and, again, the Φ value was close to 1. A *trans*-Ala is thus better accommodated at position 161 than a *trans*-Pro.

Mutations are usually regarded as probes of the local structure and energetics in the folding transition state. Pro161 is in an exposed position at the tip of an irregular structure that connects the two β strands of a hairpin (Fig. 1). We consider it unlikely that the conformational free energy that is required to shift the equilibrium towards the *cis* form originates from the local folding right at Pro161. Rather, we assume that this energy must be transmitted from another folding unit to Pro161. Pro161, with its Φ value of 1, thus reports that the cooperativity between the conformational folding in the main body of the molecule and the tip of the hairpin is already fully established in the folding transition state. Because it is linked thermodynamically with a shift in the *cis*/*trans* equilibrium at Pro161, we could determine the magnitude of this energetic interaction.

It remains unclear how N_{cis} and N_{trans} differ structurally. The N_{trans}→N_{cis} reaction is accompanied by a small increase in energy transfer from Tyr residues to the single Trp residue of N2 domain (Trp181), which is remote from Pro161 (Fig. 1). We speculate that the cluster of Tyr residues (Y151, Y165, Y166, and Y168) in the β hairpin that leads into the Pro161-containing loop might be responsible for this effect. Changes in energy transfer can be brought about by changes not only in the distance between donor and acceptor but also in the quantum yield and the relative orientation of the donor and the acceptor.

Prolyl isomerization and protein function

Native-state isomerism at proline residues was first discovered for small proteins such as calbindin[12] and staphylococcal nuclease[9] by NMR spectroscopy. These and subsequent studies focused on the structural consequences of alternative prolyl isomers. Our analysis of the kinetics and thermodynamics of the proline-coupled unfolding and refolding of the N2′ domain allowed us to determine the amount of folding free energy that is available and transmittable to the prolyl bond to shift its *cis*/*trans* equilibrium and to demonstrate that both protein isomers belong to the ensemble of native states.

Although Pro161 is exposed and located at the very tip of an extended hairpin, 10 kJ mol^{-1} of conformational free energy is available to shift its *cis*/*trans* equilibrium. This is equivalent to half of the entire conformational stability of the *cis* form of N2′ (cf. Table 2). We do not know where this energy originates from and how it is transmitted to Pro161.

The analysis of the unfolding kinetics showed that both isomers belong to the ensemble of native states because they unfold with the same rate. Their unfolding transitions show similar *m* values (Table 2), and the ratio between them could not be shifted by stabilizing or destabilizing cosolutes, providing clear evidence that they expose similarly native-like surfaces to the solvent. Species with native-like properties and alternative prolyl isomers have also been identified for other proteins, such as the ribonucleases A[42] and T_1,[43] tendamistat,[19] dihydrofolate reductase,[44] and staphylococcal nuclease.[45–47] For staphylococcal nuclease, the structure and the folding mechanism of the *cis* and the *trans* forms were also investigated in considerable detail, but, in this case, the species with the nonnative *trans*-Pro117 isomer unfolded much more rapidly than the native protein with a *cis*-Pro117 isomer; therefore, we suggest that, in this case, the *trans* form is not a part of the native ensemble.

In the case of the N2′ domain, we do not have evidence that the two isomers have different biological functions. There is, however, little doubt that the importance of prolyl isomerism extends far beyond its well-documented role in protein folding.[21,48–50] Alternative functional states of a protein can, in principle, be locked in by prolyl isomerization because the interconversion between the isomers is intrinsically slow. Pro213, the other *cis* proline of the phage gene-3 protein, provides an illustrative example. Its *cis*→*trans* isomerization is coupled with the unfolding of a subdomain, which, in turn, activates the gene-3 protein for the infection process.[51] The active infectious form is long-lived because the *trans*→*cis* backreaction is very slow. The rate of this reaction is determined by the local sequence around Pro213.[48]

The coupling between conformational folding and prolyl *cis/trans* isomerization provides a simple means for propagating input signals, such as those from ligand binding or covalent modification across a protein, and for converting them into another signal (i.e., an alternative prolyl isomer), which, in turn, could be used for an isomer-specific readout. At the same time, an intrinsically fast signal, such as ligand binding, could be combined with a slow response. The slowness of prolyl isomerization ensures that the signal persists for an extended time. Experimental evidence for the individual components of such proline-mediated signal transduction chains is available. Ligand binding can shift *cis/trans* equilibria,[38,39,52,53] prolyl isomerization can shift conformational equilibria,[21,48,49] and the isomeric state of prolyl bonds can be read out by prolyl isomerases, protein kinases,[54] and phosphatases.[55]

Materials and Methods

Expression and purification of N2′ and N2′-P161A

For the expression of N2′ and N2′-P161A [residues 102–205 of mature G3P, plus $(His)_6$], the gene fragments were PCR-amplified from the corresponding gene-3 of phage fd.[30] The fragments were cloned into the expression plasmid pET11a (Novagen, Madison, WI, USA) via its NdeI and BamHI restriction sites, and the proteins were overproduced in *E. coli* BL21(DE3) (Stratagene, La Jolla, CA, USA).

After lysis of the cells in 50 mM Tris/HCl, 50 mM NaCl, 10 mM ethylenediaminetetraacetic acid (EDTA), and 20 mM DTT (pH 8.0) with a microfluidizer and after centrifugation, the protein was found as inclusion bodies in the pellet. The inclusion bodies were washed three times with 1 mM EDTA, 1 mg ml^{-1} desoxycholate, and 20 mM DTT (pH 8.0); three times with the lysis buffer; and once with 100 mM Tris/HCl, 1 mM EDTA, and 50 mM reduced glutathione (pH 8.5); and then solubilized and reduced by stirring for 1 h at room temperature with 6.0 M GdmCl and 5 mM reduced glutathione in 50 mM Tris/HCl and 5 mM imidazole (pH 8.0). The solubilized proteins were then loaded on a Ni-NTA column (volume, 25 ml) at 4 °C, washed with two column volumes of the same buffer, and refolded on the column by a 15-min incubation with 50 mM Tris/HCl, 0.2 M GdmCl, 5 mM reduced glutathione, and 0.5 mM oxidized glutathione (pH 8.5). The folded protein was then eluted with 250 mM imidazole in 50 mM Tris/HCl (pH 8.0). Finally, the proteins were subjected to size-exclusion chromatography in 100 mM K phosphate (pH 7.0) on a Superdex HiLoad column (GE Healthcare). The protein-containing fractions were pooled and concentrated in Amicon Ultra units (Millipore). Yields were about 4 mg l^{-1}.

Equilibrium unfolding transitions

Samples of N2′ (1.0 μM) were incubated for 2 h at 15 °C in 100 mM K phosphate (pH 7.0) and varying concentrations of urea. The fluorescence of the samples was measured in 1-cm cuvettes at 340 nm (bandwidth, 10 nm) after excitation at 280 nm (bandwidth, 5 nm) in a Hitachi F4010 fluorescence spectrometer. The experimental data were analyzed according to a two-state model by assuming that ΔG_D, as well as the fluorescence emissions of the folded and the unfolded forms, depends linearly on the urea concentration. A nonlinear least-squares fit with proportional weighting of the experimental data was used to obtain ΔG_D as a function of the urea concentration.[36]

For CD measurements of the urea-induced unfolding transition, the protein concentration was 2 μM in 100 mM K phosphate (pH 7.0). The CD signal was measured at 222 nm with 1 nm bandwidth and 10 mm path length in a Jasco J-600A spectropolarimeter.

The heat-induced unfolding transitions were measured in a Jasco J-600 spectropolarimeter equipped with a PTC 348 WI Peltier element at protein concentrations of 4 μM (N2′) and 1 μM (G3P) in 100 mM K phosphate (pH 7.0) at a heating rate of 1 °C min^{-1}. The transitions were monitored for isolated N2′ by the increase in the CD signal at 222 nm, and for N2′ in the N1–N2 two-domain fragment of the gene-3 protein by the decrease in the CD signal at 210 nm with 1 nm bandwidth and 10 mm path length. The experimental data were analyzed based on the two-state approximation,[56] with a heat capacity change ΔC_p of 6400 J mol^{-1} K^{-1} (calculated as described in Privalov and Gill[57]).

Kinetic experiments

All urea-induced unfolding and refolding experiments were performed in 100 mM K phosphate (pH 7.0) at 15 °C at a final protein concentration of 1.0 μM. To measure the kinetics of the slow reaction, native or denatured (in 5.0 M urea) N2′ was manually diluted 20-fold with urea solutions of varying concentrations. The changes in fluorescence (emission at 340 nm after excitation at 280 nm) were monitored in a 10-mm cell at a response time of 0.5 s and at bandwidths of 5 nm and 10 nm for excitation and emission, respectively.

A DX.17MV stopped-flow spectrometer from Applied Photophysics (Leatherhead, UK) was employed in the single-mixing mode to follow the fast folding reactions, and in the sequential mixing mode for the double-mixing experiments. The native or the unfolded (in 4.4 M urea) protein was diluted 11-fold with urea solutions of varying concentrations. The kinetics were followed by the change in fluorescence above 320 nm after excitation at 280 nm (bandwidth, 10 nm) in an observation cell with 2 mm path length. A 0.5-cm cell with acetone was placed between the observation chamber and the photomultiplier to absorb scattered light from the excitation beam. The kinetics were measured at least eight times under identical conditions and averaged to improve the signal-to-noise ratio.

Double-jump (interrupted unfolding) experiments were performed as originally described.[5] In the first step, 66.6 μM native protein was diluted 11-fold with buffer to initiate unfolding at 5 M urea. After delay times of between 20 ms and 250 s, refolding was induced by a further 6-fold dilution to final folding conditions of 1.0 μM protein in 0.8 M urea and 100 mM K phosphate (pH 7.0). To measure the fast refolding reaction of the U_{cis} molecules with improved accuracy, U_{cis} was produced by 2.3 s of unfolding at 5.0 M urea (pH 7.0) or by 100 ms of unfolding in 100 mM glycine buffer (pH 2.0). This unfolding step was followed by refolding in the presence of 0–4.2 M urea in 100 mM K phosphate (pH 7.0).

The formation of native protein during refolding was monitored by stopped-flow interrupted refolding experiments as described.[35] In the first step, 66.7 μM unfolded protein in 5.0 M urea was diluted 11-fold with buffer to

initiate refolding at 1.0 M urea; in the second step, after various times of refolding, it was diluted 6-fold with urea solutions to final conditions of 5.0 M urea in 100 mM K phosphate (pH 7.0). Under these conditions, N2′ unfolds in a monoexponential reaction. The kinetics were followed by the change in fluorescence above 320 nm after excitation at 280 nm (bandwidth, 10 nm) in an observation chamber with 2 mm path length.

Simulations of the unfolding and refolding kinetics

The microscopic rate constants for the two folding reactions, $U_{trans} \leftrightarrow N_{trans}$ (k_{14} and k_{41}) and $U_{cis} \leftrightarrow N_{cis}$ (k_{23} and k_{32}), were derived from independent analyses of the apparent unfolding and refolding rates λ_2 and λ_3 in Fig. 3a based on a linear two-state model.[37] We assumed that, under unfolding conditions, U_{cis} and U_{trans} unfold with the same rate (λ_2), and that the logarithms of all rate constants depend linearly on the urea concentration.

The microscopic rate constants of Pro161 isomerization in the folded protein (k_{34} and k_{43}) and in the unfolded protein (k_{12} and k_{21}) were derived from $\lambda_1 = 0.01\ s^{-1}$ for the $N_{trans} \leftrightarrow N_{cis}$ reaction at 0–1 M urea and from $\lambda_1 = 0.02\ s^{-1}$ for the $U_{trans} \leftrightarrow U_{cis}$ reaction at 4.2–4.8 M urea (Fig. 4a). The corresponding equilibrium constants $[N_{trans}]/[N_{cis}] = 14.5/85.5$ and $[U_{trans}]/[U_{cis}] = 93/7$ were derived from the ratio of the refolding amplitudes after 20 ms and >200 s, respectively, as observed in the slow refolding assays shown in Fig. 5. To convert the measured amplitudes of the refolding assays into fractions of species, we assumed fluorescence values of 0 for U_{cis} and U_{trans}, 0.81 for N_{trans}, and 1.0 for N_{cis}, as evident from Fig. 6a. For the calculations, we assumed that these isomerizations are urea-independent.

To evaluate the four-species box mechanism in Fig. 7, the time courses of the four species in unfolding and refolding were calculated at various urea concentrations by numerical integration of the differential rate equations, Eqs. (1) Eqs. (2) Eqs. (3) Eqs. (4). For these calculations, the program SCIENTIST (Micromath) was used:

$$\frac{dU_{trans}}{dt} = -(k_{12} + k_{14})U_{trans} + k_{21}U_{cis} + k_{41}N_{trans} \quad (1)$$

$$\frac{dU_{cis}}{dt} = -(k_{21} + k_{23})U_{cis} + k_{12}U_{trans} + k_{32}N_{cis} \quad (2)$$

$$\frac{dN_{cis}}{dt} = -(k_{32} + k_{34})N_{cis} + k_{23}U_{cis} + k_{43}N_{trans} \quad (3)$$

$$\frac{dN_{trans}}{dt} = -(k_{41} + k_{43})N_{trans} + k_{14}U_{trans} + k_{34}N_{cis} \quad (4)$$

The corresponding fluorescence changes during unfolding and refolding (ΔF) as a function of time were calculated by assuming fluorescence values of 1.0 for N_{cis}, 0.81 for N_{trans}, and 0 for U_{cis} and U_{trans} (Eq. (5)):

$$\Delta F = N_{cis} + 0.81N_{trans} + 0U_{cis} + 0U_{trans} \quad (5)$$

N_{trans}, N_{cis}, U_{trans}, and U_{cis} in Eqs. (1) Eqs. (2) Eqs. (3) Eqs. (4) Eqs. (5) represent the reduced concentrations of the individual species in Fig. 7. The sum of the reduced concentrations of all species is equal to 1. In the simulations of unfolding, the initial concentrations of N_{cis} and N_{trans} were assumed to be 0.855 and 0.145, respectively; in the refolding simulations, the initial concentrations of U_{cis} and U_{trans} were 0.07 and 0.93, respectively. The calculated kinetic curves were then analyzed as the sum of exponential functions.

Acknowledgements

We thank Barbara Eckert for the expression construct of N2′ and the members of our group for many discussions of this work. This work was supported by grants from the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie.

References

1. Cheng, H. N. & Bovey, F. A. (1977). *Cis–trans* equilibrium and kinetic studies of acetyl-L-proline and glycyl-L-proline. *Biopolymers*, **16**, 1465–1472.
2. Grathwohl, C. & Wüthrich, K. (1981). NMR studies of the rates of proline *cis–trans* isomerization in oligopeptides. *Biopolymers*, **20**, 2623–2633.
3. Balbach, J. & Schmid, F. X. (2000). Prolyl isomerization and its catalysis in protein folding. In *Prolyl isomerization and its catalysis in protein folding* (Pain, R. H., ed), pp. 212–237, Oxford University Press, Oxford.
4. Schmid, F. X. (2005). Prolyl isomerization in protein folding. In *Protein Folding Handbook* (Buchner, J. K. T., ed), pp. 916–945, Wiley-VCH, Weinheim.
5. 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.
6. Schmid, F. X. & Baldwin, R. L. (1978). Acid catalysis of the formation of the slow-folding species of RNase A: evidence that the reaction is proline isomerization. *Proc. Natl Acad. Sci. USA*, **75**, 4764–4768.
7. Nall, B. T. (1985). Proline isomerization and protein folding. *Comments Mol. Cell. Biophys.* **3**, 123–143.
8. Wedemeyer, W. J., Welker, E. & Scheraga, H. A. (2002). Proline *cis–trans* isomerization and protein folding. *Biochemistry*, **41**, 14637–14644.
9. 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.
10. Higgins, K. A., Craik, D. J., Hall, J. G. & Andrews, P. R. (1988). *Cis–trans* isomerization of the proline residue in insulin studied by ^{13}C NMR spectroscopy. *Drug Des. Delivery*, **3**, 159–170.
11. Chazin, W. J., Kördel, J., Drakenberg, T., Thulin, E., Brodin, P., Grundström, T. & Forsén, S. (1989). Proline isomerism leads to multiple folded conformations of calbindin D9k: direct evidence from two-dimensional NMR spectroscopy. *Proc. Natl Acad. Sci. USA*, **86**, 2195–2198.
12. 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.
13. Adjadj, E., Naudat, V., Quiniou, E., Wouters, D., Sautiere, P. & Craescu, C. T. (1997). Solution structure of Lqh-8/6, a toxin-like peptide from a scorpion venom—structural heterogeneity induced by proline *cis/trans* isomerization. *Eur. J. Biochem.* **246**, 218–227.
14. Feng, Y., Hood, W. F., Forgey, R. W., Abegg, A. L., Caparon, M. H., Thiele, B. R. *et al.* (1997). Multiple conformations of a human interleukin-3 variant. *Protein Sci.* **6**, 1777–1782.
15. Yuan, X., Downing, A. K., Knott, V. & Handford, P. A. (1997). Solution structure of the transforming growth factor beta-binding protein-like module, a

domain associated with matrix fibrils. *EMBO J.* **16**, 6659–6666.
16. Loh, S. N., Mcnemar, C. W. & Markley, J. L. (1991). Detection and kinetic characterization of a novel proline isomerism in staphylococcal nuclease by NMR spectroscopy. *Tech. Protein Chem.* **II**, 275–282.
17. Truckses, D. M., Somoza, J. R., Prehoda, K. E., Miller, S. C. & Markley, J. L. (1996). Coupling between *trans*/*cis* proline isomerization and protein stability in staphylococcal nuclease. *Protein Sci.* **5**, 1907–1916.
18. Rosengren, K. J., Wilson, D., Daly, N. L., Alewood, P. F. & Craik, D. J. (2002). Solution structures of the *cis*- and *trans*-Pro30 isomers of a novel 38-residue toxin from the venom of *Hadronyche infensa* sp. that contains a cystine-knot motif within its four disulfide bonds. *Biochemistry*, **41**, 3294–3301.
19. Pappenberger, G., Bachmann, A., Muller, R., Aygun, H., Engels, J. W. & Kiefhaber, T. (2003). Kinetic mechanism and catalysis of a native-state prolyl isomerization reaction. *J. Mol. Biol.* **326**, 235–246.
20. 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. *Nat. Struct. Biol.* **9**, 900–905.
21. Andreotti, A. H. (2003). Native state proline isomerization: an intrinsic molecular switch. *Biochemistry*, **42**, 9515–9524.
22. Sarkar, P., Reichman, C., Saleh, T., Birge, R. B. & Kalodimos, C. G. (2007). Proline *cis–trans* isomerization controls autoinhibition of a signaling protein. *Mol. Cell*, **25**, 413–426.
23. 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. *Nat. Struct. Biol.* **5**, 140–147.
24. 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.
25. Click, E. M. & Webster, R. E. (1997). Filamentous phage infection: required interactions with the TolA protein. *J. Bacteriol.* **179**, 6464–6471.
26. 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.
27. 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.
28. Guex, N. & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, **18**, 2714–2723.
29. Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics*, **14**, 51–55.
30. 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.
31. Martin, A. & Schmid, F. X. (2003). The folding mechanism of a two-domain protein: folding kinetics and domain docking of the gene-3-protein of phage fd. *J. Mol. Biol.* **329**, 599–610.
32. Knappe, T. A., Eckert, B., Schaarschmidt, P., Scholz, C. & Schmid, F. X. (2007). Insertion of a chaperone domain converts FKBP12 into a powerful catalyst of protein folding. *J. Mol. Biol.* **368**, 1458–1468.
33. 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.
34. 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.
35. Schmid, F. X. (1983). Mechanism of folding of ribonuclease A. Slow refolding is a sequential reaction *via* structural intermediates. *Biochemistry*, **22**, 4690–4696.
36. 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.
37. Bieri, O. & Kiefhaber, T. (2000). Kinetic models in protein folding. In *Mechanisms of Protein Folding* (Pain, R. H., ed). pp. 34–64, 2nd edit. Oxford University Press, Oxford.
38. Alexandrescu, A. T., Hinck, A. P. & Markley, J. L. (1990). Coupling between local structure and global stability of a protein: mutants of staphylococcal nuclease. *Biochemistry*, **29**, 4516–4525.
39. Evans, P. A., Kautz, R. A., Fox, R. O. & Dobson, C. M. (1989). A magnetization-transfer nuclear magnetic resonance study of the folding of staphylococcal nuclease. *Biochemistry*, **28**, 362–370.
40. Ikai, A. & Tanford, C. (1973). Kinetics of unfolding and refolding of proteins: I. Mathematical analysis. *J. Mol. Biol.* **73**, 145–163.
41. Fersht, A. R. (1993). Protein folding and stability—the pathway of folding of barnase. *FEBS Lett.* **325**, 5–16.
42. 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.
43. Kiefhaber, T., Schmid, F. X., Willaert, K., Engelborghs, Y. & Chaffotte, A. (1992). Structure of a rapidly formed intermediate in ribonuclease-T_1 folding. *Protein Sci.* **1**, 1162–1172.
44. Jennings, P. A., Finn, B. E., Jones, B. E. & Matthews, C. R. (1993). A reexamination of the folding mechanism of dihydrofolate reductase from *Escherichia coli*—verification and refinement of a 4-channel model. *Biochemistry*, **32**, 3783–3789.
45. Nakano, T., Antonino, L. C., Fox, R. O. & Fink, A. L. (1993). Effect of proline mutations on the stability and kinetics of folding of staphylococcal nuclease. *Biochemistry*, **32**, 2534–2541.
46. Walkenhorst, W. F., Green, S. M. & Roder, H. (1997). Kinetic evidence for folding and unfolding intermediates in staphylococcal nuclease. *Biochemistry*, **36**, 5795–5805.
47. Ikura, T., Tsurupa, G. P. & Kuwajima, K. (1997). Kinetic folding and *cis*/*trans* prolyl isomerization of staphylococcal nuclease. A study by stopped-flow absorption, stopped-flow circular dichroism, and molecular dynamics simulations. *Biochemistry*, **36**, 6529–6538.
48. Eckert, B., Martin, A., Balbach, J. & Schmid, F. X. (2005). Prolyl isomerization as a molecular timer in phage infection. *Nat. Struct. Mol. Biol.* **12**, 619–623.
49. Sarkar, P., Reichman, C., Saleh, T., Birge, R. B. & Kalodimos, C. G. (2007). Proline *cis–trans* isomerization controls autoinhibition of a signaling protein. *Mol. Cell*, **25**, 413–426.
50. Lu, K. P., Finn, G., Lee, T. H. & Nicholson, L. K. (2007). Prolyl *cis–trans* isomerization as a molecular timer. *Nat. Chem. Biol.* **3**, 619–629.

51. Eckert, B. & Schmid, F. X. (2007). A conformational unfolding reaction activates phage fd for the infection of *Escherichia coli*. *J. Mol. Biol.* **373**, 452–461.
52. Breheny, P. J., Laederach, A., Fulton, D. B. & Andreotti, A. H. (2003). Ligand specificity modulated by prolyl imide bond *cis/trans* isomerization in the Itk SH2 domain: a quantitative NMR study. *J. Am. Chem. Soc.* **125**, 15706–15707.
53. Ng, K. K. S. & Weis, W. I. (1998). Coupling of prolyl peptide bond isomerization and Ca^{2+} binding in a C-type mannose-binding protein. *Biochemistry*, **37**, 17977–17989.
54. Weiwad, M., Werner, A., Rucknagel, P., Schierhorn, A., Kullertz, G. & Fischer, G. (2004). Catalysis of proline-directed protein phosphorylation by peptidyl–prolyl *cis/trans* isomerases. *J. Mol. Biol.* **339**, 635–646.
55. Zhou, X. Z., Kops, O., Werner, A., Lu, P. J., Shen, M., Stoller, G. *et al.* (2000). Pin1-dependent prolyl isomerization regulates dephosphorylation of Cdc25C and tau proteins. *Mol. Cell*, **6**, 873–883.
56. Mayr, L. M., Landt, O., Hahn, U. & Schmid, F. X. (1993). Stability and folding kinetics of ribonuclease T_1 are strongly altered by the replacement of *cis*-proline 39 with alanine. *J. Mol. Biol.* **231**, 897–912.
57. Privalov, P. L. & Gill, S. J. (1988). Stability of protein structure and hydrophobic interaction. *Adv. Protein Chem.* **39**, 191–234.

Teilarbeit D:

Jakob, R.P. and Schmid, F.X. (2009).
Molecular determinants of a native-state prolyl isomerization.
J. Mol. Biol. **387**, 1017-31.

doi:10.1016/j.jmb.2009.02.021 J. Mol. Biol. (2009) 387, 1017–1031

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Molecular Determinants of a Native-State Prolyl Isomerization

Roman P. Jakob and Franz X. Schmid*

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

Received 17 December 2008; received in revised form 6 February 2009; accepted 10 February 2009
Available online 20 February 2009

Prolyl *cis*/*trans* isomerizations determine the rates of many protein-folding reactions, and they can serve as molecular switches and timers. The energy required to shift the prolyl *cis*/*trans* equilibrium during these processes originates from conformational reactions that are linked structurally and energetically with prolyl isomerization. We used the N2 domain of the gene-3-protein of phage fd to elucidate how such an energetic linkage develops in the course of folding. The Asp160–Pro161 bond at the tip of a β hairpin of N2 is *cis* in the crystal structure, but in fact, it exists as a mixture of conformers in folded N2. During refolding, about 10 kJ mol^{-1} of conformational energy becomes available for a 75-fold shift of the *cis*/*trans* equilibrium constant at Pro161, from 7/93 in the unfolded to 90/10 in the folded form. We combined single- and double-mixing kinetic experiments with a mutational analysis to identify the structural origin of this proline shift energy and to elucidate the molecular path for the transfer of this energy to Pro161. It originates largely, if not entirely, from the two-stranded β sheet at the base of the Pro161 hairpin. The two strands improve their stabilizing interactions when Pro161 is *cis*, and this stabilization is propagated to Pro161, because the connector peptides between the β strands and Pro161 are native-like folded when Pro161 is *cis*. In the presence of a *trans*-Pro161, the connector peptides are locally unfolded, and thus, Pro161 is structurally and energetically uncoupled from the β sheet. Such interrelations between local folding and prolyl isomerization and the potential modulation by prolyl isomerases might also be used to break and re-establish slow communication pathways in proteins.

Keywords: protein-folding kinetics; protein-folding mechanism; gene-3-protein; proline switch; prolyl isomerase

Edited by C. R. Matthews

**Corresponding author.* E-mail address: fx.schmid@uni-bayreuth.de.

Abbreviations used: N2, N-terminal domain of gene-3-protein of phage fd; N2′, N2 containing the stabilizing mutation Q129H; T_M, midpoint of a thermal unfolding transition; ΔH_D, van't Hoff enthalpy of denaturation at T_M; $[urea]_M$, midpoint of a urea-induced unfolding transition; ΔG_D, Gibbs free energy of denaturation; m, cooperativity value of a denaturant (D)-induced equilibrium unfolding transition; A, λ, amplitude and apparent rate constant of a folding reaction; τ, time constant of a folding reaction; k_{NU}, microscopic rate constant of unfolding; m_{NU}, kinetic m value of unfolding; k_{UN}, microscopic rate constant of refolding; m_{UN}, kinetic m value of refolding; U_c, U_t, N_c, N_t, unfolded and native forms of N2′ with a *cis*- or a *trans*-Pro161, respectively; Φ value, ratio of the equilibrium and the activation Gibbs free energies of refolding.

Introduction

The *cis*/*trans* isomerizations of peptidyl–prolyl bonds in proteins are intrinsically slow reactions, which occur in the time range of several minutes near 25 °C.[1–3] In unfolded or nascent protein chains, the *trans* form is favored over *cis*, and therefore, proteins with *cis* prolyl bonds in the native state must undergo *trans*→*cis* isomerization during their folding.[4–6] Conformational folding and prolyl isomerization are coupled in an intricate fashion. Usually, folding starts with a particular proline still in the incorrect (*trans*) state, but when a certain extent of folding is reached, this *trans*-proline acts as a barrier and blocks further folding.[7–15] In these initial folding steps, conformational energy is accumulated, which becomes available to shift the *cis*/*trans* equilibrium towards the native *cis* state, and as

soon as the proline isomerizes, conformational folding can resume and go to completion rapidly.

In the folded state of such a protein, the *cis* isomer is favored, because more stabilizing interactions can be established in the *cis* form than in the *trans* form. Folding and prolyl isomerization are thus linked energetically. The folded conformation stabilizes the *cis* isomer and, in turn, the *cis* isomer allows the formation of additional stabilizing conformational interactions. During the folding process, this cooperativity is established in a stepwise fashion. Early in folding, conformational energy must be invested to shift the *cis*/*trans* equilibrium; the reciprocal gain in energy upon establishing the interactions of the native state is delayed, because it is tied to the slow *trans*→*cis* isomerization in the rate-limiting step of folding.[16–20]

Conformational transitions that are linked with prolyl isomerization and, thus, retarded into the time range of minutes do not only occur as important rate-limiting steps in protein folding. They can also be used as molecular switches and timers to regulate biological processes,[21–30] and prolyl isomerases that modulate the rates of such proline-limited processes are ubiquitous enzymes.[25,31–34]

Folding intermediates or alternatively folded states with incorrect prolyl isomers are usually elusive species with a low stability. Therefore, it is not known how they accumulate conformational energy before prolyl isomerization, how this energy is used to drive prolyl isomerization, and how the invested energy is returned after the isomerization. In folded proteins, the "native" prolyl isomer is usually strongly favored over the "nonnative" one, and therefore, the species with the nonnative prolyl isomer is hardly populated and cannot be characterized at equilibrium by structural methods. In rare cases, a coexistence of *cis* and *trans* prolyl isomers has been detected by NMR spectroscopy, but detailed structural analyses of the minor species have not been possible.[35–41]

Here, we used the N2 domain of the gene-3-protein of the filamentous phage fd to elucidate the molecular nature of the coupling between conformational folding and prolyl isomerization. N2 provides an excellent system to study this linkage, because the stabilities and the unfolding and refolding kinetics of the *cis* and *trans* forms can be determined simultaneously.[42] The N2 domain is an independently folding unit (Fig. 1a). Pro161, which is located

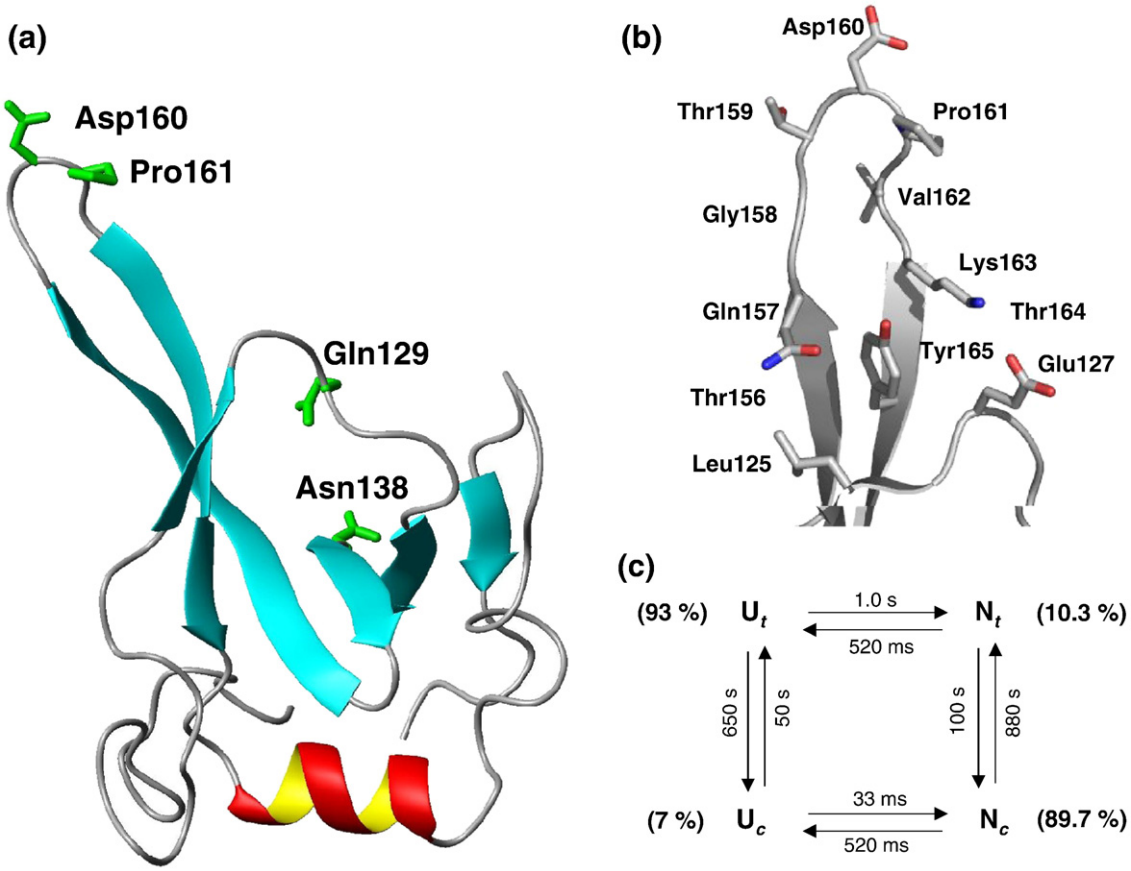


Fig. 1. (a) Tertiary structure of the N2 domain of G3P (coordinates from Ref. 43). The side chains of Gln129, Asn138, Asp160, and Pro161 are shown in stick representation. (b) Enlarged view of the loop region from Thr156 to Tyr165. The figure was prepared by using MOLMOL.[44] (c) Kinetic model for the folding and prolyl isomerization of N2′. The *cis* and *trans* contents in unfolded and native N2′, as taken from Table 2, are given in parentheses. The values above and below the arrows are the reciprocal rate constants. Those for the $U_c \leftrightarrow U_t$ reaction are given for unfolding in 5.0 M urea; all other values refer to refolding in 1.0 M urea.

at the tip of a β hairpin (Fig. 1b), was found to be *cis* in the crystal structure[43,45] of N2, but, in fact, it exists as a mixture of two native, folded forms that differ in the *cis*/*trans* conformation at Pro161. During the refolding of N2, the *cis*/*trans* equilibrium constant changes 75-fold, and the *cis* content increases from 7% in the unfolded form to 90% in the folded form.[42] Accordingly, an additional 10.3 kJ mol^{-1} of conformational folding energy becomes available when Pro161 is *cis*, and this energy is used to drive the change in the *cis*/*trans* equilibrium during folding. Both the *cis* and the *trans* forms of N2 are truly native states. They unfold with identical rates but differ in the rate of refolding, which indicates that the conformational energy that drives the *cis*/*trans* isomerization is available in the folding transition state already. The interrelation between folding and prolyl isomerization in the unfolded and the folded forms of N2 is well described by the box model in Fig. 1c, which relates the conformational folding reactions with the prolyl isomerizations in the unfolded and the folded forms of the protein.

Here, we combined single- and double-mixing kinetic experiments with an exhaustive mutational analysis to elucidate why the folded state of N2 is compatible with both prolyl isomers, how the local conformation in the Pro161 loop is coupled with the *cis*/*trans* isomeric state at Pro161, and how, during folding, conformational energy is accumulated in the folded part of N2 and transferred to Pro161 at the tip of this loop. Asp160 and Pro161 flank the cis peptide bond in wild-type N2 (Fig. 1). We replaced Asp160 with the 19 other natural amino acids and studied how these substitutions affected the stability, the folding kinetics of the *cis*-Pro161 and the *trans*-Pro161 forms, and their interconversion in the unfolded and the folded state of N2. Then, we analyzed the role of the connector peptides that link Pro161 at the tip of the hairpin with the β strands of the hairpin and the energetic linkage between the β strands and the isomerization at Pro161.

The residue before Pro161 affects the overall stability of N2, but, surprisingly, not the *cis*/*trans* equilibration reaction. It is determined by the stability of the β sheet at the base of the Pro161 hairpin. The energetic linkage between Pro161 and this β sheet is established in the *cis* form, because the connector peptides are folded in the *cis* form but unfolded in the *trans* form.

Results and Discussion

The *cis*/*trans* equilibrium at Pro161 in folded N2 is not determined by the local sequence

In the absence of folded structure, such as in short peptides, the *cis*/*trans* equilibria and the isomerization kinetics of Xaa–Pro bonds are determined by the chemical nature of residue Xaa.[46] In wild-type N2, Pro161 is preceded by Asp160. To elucidate the role of residue 160 for Pro161 isomerization in the unfolded and the folded form of N2 as well as for the folding transition state, we created a library of N2 variants in which this position was occupied by all 20 amino acids. As in our previous work,[42] we used

Table 1. Stability data for variants of N2′ with substitutions at position 160

Residue at position 160	T_M (°C)	ΔH_D (kJ mol^{-1})	$[urea]_M$ (M)	m (kJ mol^{-1} M^{-1})	ΔG_D (kJ mol^{-1})	$\Delta\Delta G_D$(2.4 M) (kJ mol^{-1})
Ala	37.2	319	2.40	7.0	16.8	−2.2
Cys	n.d.	n.d.	2.58	6.8	17.5	−0.9
Asp (wt)	38.9	310	2.71	6.6	17.9	0
Glu	36.4	332	2.41	6.8	16.4	−2.1
Phe	37.2	359	2.41	6.8	16.4	−2.1
Gly	35.0	308	2.19	6.9	15.1	−3.6
His	38.2	331	2.65	6.9	18.1	−0.5
Ile	35.0	303	2.11	6.6	14.0	−4.1
Lys	36.8	351	2.45	6.7	16.5	−1.9
Leu	35.1	326	2.22	7.0	15.6	−0.9
Met	36.4	343	2.24	7.1	15.8	−1.1
Asn	39.6	356	2.88	6.3	18.2	0.8
Pro	36.5	319	2.20	6.9	15.2	−3.6
Gln	36.5	342	2.37	6.7	15.9	−2.0
Arg	35.9	285	2.38	6.4	15.1	−2.1
Ser	37.5	305	2.56	6.5	16.7	−1.2
Thr	35.8	299	2.25	6.6	14.8	−3.2
Val	34.6	309	1.99	6.6	13.2	−4.8
Trp	n.d.	n.d.	2.55	6.7	17.2	−1.2
Tyr	36.6	324	2.52	7.0	17.6	−1.4

The stability parameters result from two-state analyses of thermal (columns 2 and 3) and urea-induced unfolding transitions (columns 4–7). For all proteins, the melting temperature (T_M), the van't Hoff enthalpy of denaturation at T_M (ΔH_D), the midpoint of the urea-induced unfolding transition ($[urea]_M$), the cooperativity value (m), and the Gibbs free energy of denaturation (ΔG_D) at 15 °C and 0 M urea are given. $\Delta\Delta G_D$ depicts the differences in ΔG_D between the variant and the wild-type (Asp160) protein at 2.4 M urea, which is in the transition region for most variants. The thermal unfolding transitions were measured in 100 mM K phosphate (pH 8.0), and the urea-induced unfolding transitions were measured in 100 mM K phosphate (pH 7.0) at 15 °C. Thermal unfolding of N2′ variants D160C and D160W could not be determined, because they aggregated irreversibly under these conditions. The precisions of T_M and ΔH_D are ±0.1 °C and ±15 kJ mol^{-1}, and for m and ΔG_D, they are ±0.5 kJ mol^{-1} M^{-1} and ±0.9 kJ mol^{-1}, respectively.

the Q129H variant of N2 as the reference or pseudo-wild-type protein and denote it as N2′. It is more stable than N2, and thus, a wider range of destabilizing substitutions could be investigated.

All variants with the substitutions at position 160 were soluble and folded and showed midpoints of thermal unfolding (T_M) that ranged between 34.6 and 39.6 °C (Table 1). The midpoints of the urea-induced unfolding transitions ($[urea]_M$) ranged between 2.0 and 2.9 M. Figure 2a and b show selected thermal and urea-induced transitions. The cooperativities (ΔH_D and m) were similar for all variants (Table 1), and, apart from the Cys160 and Trp160 variants, unfolding was reversible. T_M and $[urea]_M$ followed the same rank order, as shown by a good correlation between the two parameters (Fig. 2c). The side chain of residue 160 is exposed in folded N2′, which explains the modest effects on stability of these substitutions. The maximal difference in stability is 5.6 kJ mol^{-1}, with 160V being the least stable and 160N being the most stable variant.

Residue 160 affects the folding kinetics of the two Pro161 isomers of N2′ differently

All N2′ variants are completely unfolded above 5 M urea (Fig. 2b). The kinetics of unfolding at 6.0 M urea are virtually independent of the nature of residue 160 (Table 2). Only the Gly160 variant showed a small (1.5-fold) increase in the rate of unfolding. In Fig. 3a, $RT\ln\lambda_u$ is plotted as a function of ΔG_D in the form of a Brønsted plot. It shows that the barrier to unfolding is identical in height for all variants (except the Gly160 variant) and independent of the effects of the individual mutations at position 160 on the equilibrium stability.

The refolding of N2′ consists of two fast phases and a slow phase. The two fast phases (λ_3 and λ_2) reflect the conformational refolding reactions $U_c \rightarrow N_c$ and $U_t \rightarrow N_t$ (Fig. 1c), respectively; the slow phase (λ_1), measured after manual mixing, reflects the isomerization at Pro161 in the folded state ($N_t \leftrightarrow N_c$, Fig. 1c). This slow phase could be measured by fluorescence, because N_t and N_c differ slightly in Trp emission. The values obtained for λ_1 are listed in Table 2. For 18 of the 20 variants, the λ_1 values for the $N_t \leftrightarrow N_c$ reaction differ less than 2-fold, and the increase in Trp fluorescence (α_1) that occurs during this reaction is also similar (12±2%). For the Trp160 variant, we observed a decrease in Trp fluorescence, which is probably caused by a local effect of Pro161 isomerization on Trp160. The highest rate (λ_1) and the highest amplitude (α_1) for Pro161 isomerization in the folded N2′ domain was observed when position 160 was also occupied by a proline. Otherwise, the nature of the residue at position 160 had surprisingly little influence on the kinetics of this isomerization.

In the unfolded protein, *trans*-Pro161 is favored, and 93% of all molecules are in the U_t form (Fig. 1c). As a consequence, the refolding kinetics are dominated by the $U_t \rightarrow N_t$ reaction. The $U_c \rightarrow N_c$ reaction shows a very small amplitude and is difficult to measure because only 7% U_c molecules are present at equilibrium.

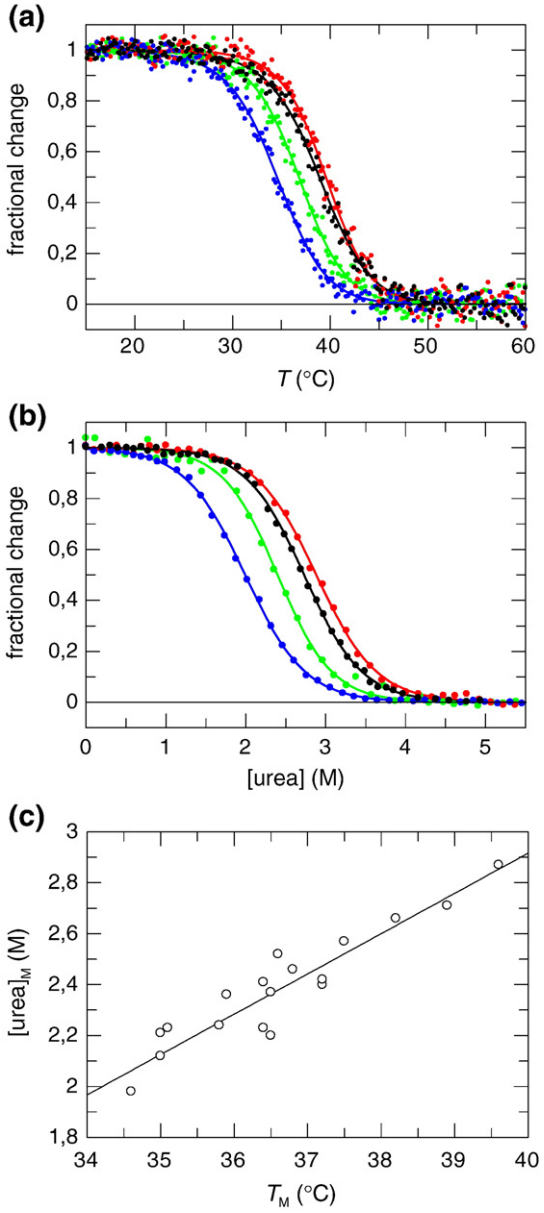


Fig. 2. (a) Thermal and (b) urea-induced equilibrium unfolding transitions of the N2′ domain (black) and the N2′ variants D160A (green), D160N (red), and D160V (blue). The thermal transitions of 4 μM protein were measured by CD at 222 nm, and the urea-induced transitions of 1 μM protein were measured by protein fluorescence at 340 nm after excitation at 280 nm at 15 °C. The fractional changes were obtained from two-state analyses of the data. The results from these analyses are shown in Table 1. The transitions were measured in 100 mM K phosphate (pH 7.0). (c) Correlation between the thermal and the urea-induced unfolding transitions. For all variants, the midpoints of the urea-induced transitions, $[urea]_M$, are plotted as a function of the corresponding midpoints of the thermally induced unfolding transitions, T_M. The midpoints are taken from Table 1. The straight line represents a linear fit to the data.

Table 2. Variants of N2′ with substitutions at position 160

Residue at position 160	Unfolding	Slow refolding		Fast refolding		*cis* content (%)		
	λ_u (s^{-1})	α_1 (%)	λ_1 (s^{-1})	λ_2 (s^{-1})	λ_3 (s^{-1})	N2′ native	N2′ unfolded	Peptide[a]
Ala	4.85	11.1	0.0197	1.35	32.1	86.5	7.5	7.7
Cys	4.82	7.1	0.0153	1.05	36.8	86.4	8.9	8.7
Asp (wt)	4.98	10.8	0.0160	2.09	49.6	89.7	7.0	7.3
Glu	4.92	11.0	0.0164	1.18	26.6	90.7	9.8	9.0
Phe	4.96	12.4	0.0146	0.84	17.3	92.3	19.8	23.0
Gly	7.58	12.5	0.0250	2.56	16.1	75.5	13.1	13.7
His	4.93	10.7	0.0235	1.16	31.2	92.3	12.9	9.5[b]
Ile	4.87	13.6	0.0189	1.10	17.0	90.4	11.3	12.0
Lys	4.97	12.4	0.0195	0.91	27.9	90.5	7.6	6.8
Leu	4.76	12.6	0.0172	0.96	16.4	90.8	12.9	12.0
Met	5.06	12.8	0.0192	0.99	24.4	91.1	10.0	10.0
Asn	4.87	7.5	0.0209	1.98	51.8	90.6	11.3	11.6
Pro	5.21	19.9	0.0274	2.93	13.9	86.9	21.7	6.0
Gln	4.92	11.6	0.0179	0.97	25.7	90.9	11.0	11.5
Arg	4.92	11.6	0.0201	0.98	26.6	90.9	8.3	7.2
Ser	4.77	10.2	0.0195	1.25	33.2	91.5	11.4	10.3
Thr	4.82	12.6	0.0183	0.98	24.2	90.8	10.0	9.4
Val	5.31	11.1	0.0201	0.81	18.4	89.5	10.2	10.4
Trp	5.07	−18.5	0.0159	0.64	12.2	92.2	33.3	37.7
Tyr	4.96	10.7	0.0175	0.90	16.5	92.8	22.6	24.0

Unfolding and refolding kinetics and *cis* contents in the native and the unfolded forms are shown.

The *cis* contents (precisions ≤1%) in the native and unfolded forms of the N2′ variants were determined by double-mixing experiments as described in Materials and Methods. The unfolding kinetics were measured in 6.0 M urea, where λ_u approximates the microscopic unfolding rate constant k_{NU}. The refolding rates were measured in 0.2 M urea, where λ_2 and λ_3 approximate the rate constants k_{UN} for the refolding of the U_t and U_c molecules, respectively. The kinetics of slow refolding (λ_1) follow the $N_t \leftrightarrow N_c$ reaction and were measured after a manual dilution of the denatured N2′ variants (in 5.0 M urea) to 0.2 M urea. The amplitude α_1 is given relative to the final fluorescence. The wild-type protein (wt) has Asp at position 160. All measurements were performed in 100 mM K phosphate (pH 7.0) at 15 °C.

[a] Data for the peptide Ac-Ala-Xaa-Pro-Ala-Lys-NH_2, measured at 23 °C in 20 mM sodium phosphate (pH 6.0), taken from Ref. 46.

[b] Measured at pH 8.0.[46]

To follow both the $U_c \rightarrow N_c$ and the $U_t \rightarrow N_t$ refolding reaction with large amplitudes, we used a stopped-flow double-mixing protocol. In the first step, native N2′ was unfolded by a jump to pH 2.0, where conformational unfolding (90% $N_c \rightarrow U_c$ and 10% $N_t \rightarrow U_t$) is fast and complete within 0.1 s.[42] The subsequent $U_c \leftrightarrow U_t$ interconversion is slow (τ=50–100 s), and therefore, about 90% of all molecules are still in the U_c state at this time. The $U_c \rightarrow N_c$ refolding reaction could thus be measured with a large amplitude when refolding was initiated after 0.1 s of unfolding. After 500 s of unfolding, *cis*/*trans* equilibration at Pro161 has gone to completion, and more than 90% of the molecules are in the U_t state. The refolding of the *trans* form ($U_t \rightarrow N_t$) could therefore be measured with a large amplitude after 500 s of unfolding.

This pH-jump double-mixing procedure also allowed us to determine the refolding kinetics in the absence of a denaturant. The measured rate constants of the two refolding reactions (λ_2 and λ_3, Table 2) are therefore virtually identical with the microscopic rate constants for the $U_t \rightarrow N_t$ and $U_c \rightarrow N_c$ refolding reactions, respectively. The Brønsted plot for the $U_c \rightarrow N_c$ reaction (λ_3, Fig. 3b) shows that the majority of the variants follow a line with a slope of 1, which indicates that, for the U_c molecules (with a *cis*-Pro161), the differences in the equilibrium Gibbs free energy of folding between the variants are present already in the transition state of folding. In other words, in the folding transition state, residue 160 is already in a native-like conformation ($\Phi \approx 1$). The strongest deviations in the Brønsted plot (Fig. 3b) are observed for Tyr, Trp, and Val. We suspect that aromatic residues such as Trp and Tyr and a β-branched residue such as Val lead to local changes in the structure of the transition state.

A different picture emerges for the refolding of the molecules with a *trans*-Pro161 ($U_t \rightarrow N_t$). It is barely affected by the substitutions at position 160, and the corresponding Brønsted plot (λ_2, Fig. 3c) indicates that for most substitutions, there is no correlation between the equilibrium and activation Gibbs free energies. They scatter around the line with a slope of 0. Increased rates of folding were found only for the two most stable variants, Asn160 and Asp160 (the wild-type residue), and for the Gly160 variant, which is less stable but folds more rapidly than the Ala160 reference variant.

In summary, the Brønsted plots in Fig. 3 reveal distinctly different effects of the substitutions at position 160 on the folding of the N2′ molecules with a *cis*- or a *trans*-Pro161. For the *cis* form, the slope is near 1, indicating that the 160 region is native-like ordered in the folding transition state of the $U_c \rightarrow N_c$ reaction. For the *trans* form, the slope is near 0, which suggests that the residue before P161 is unordered when this Pro is *trans*.

The *cis*/*trans* equilibria at Pro161 in the unfolded and folded forms of the variants are not correlated

The *cis*/*trans* equilibria at Pro161 in the folded forms of the variants were also determined in

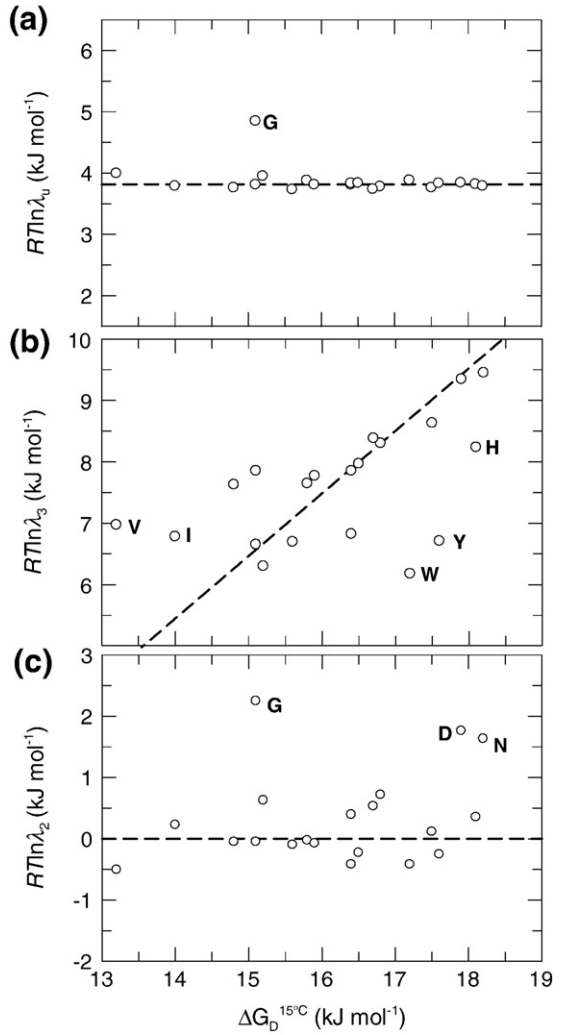


Fig. 3. Brønsted plots for the apparent rate constants (a) λ_u for unfolding and (b) λ_3 for $U_c \rightarrow N_c$ refolding and (c) λ_2 for $U_t \rightarrow N_t$ refolding of the N2′-D160X-protein variants. The broken lines in (a) and (c) have fixed slopes of 0, whereas the broken line in (b) has a fixed slope of 1. Points that deviate strongly from the general trends are identified by the nature of the amino acid at position 160. The λ values are taken from Table 2. They were measured under conditions (6.0 and 0.2 M urea) where they approximate the microscopic rate constants of unfolding and refolding, respectively. All measurements were performed in 100 mM K phosphate (pH 7.0) at 15 °C.

double-mixing experiments. As outlined above, a 0.1-s unfolding pulse at pH 2.0 is sufficient to convert the folded forms N_c and N_t into the corresponding unfolded forms U_c and U_t (Fig. 1c), but the *cis/trans* distribution at Pro161 remains essentially unchanged. The amplitudes of the refolding reactions $U_c \rightarrow N_c$ and $U_t \rightarrow N_t$ after such a short 0.1-s unfolding pulse are thus proportional to the fractions of N_c and N_t, as present originally, and reflect the *cis/trans* equilibrium at Pro161 in the folded protein. When the fraction of N_t is calculated from the amplitude of the $U_t \rightarrow N_t$ reaction, the small signal change that accompanies the subsequent $N_t \leftrightarrow N_c$ equilibration (Table 2) is accounted for as described previously.[42]

To determine the Pro161 *cis/trans* equilibrium in the unfolded state, the equilibration between U_c and U_t was allowed to go to completion by incubating the N2′ variants for 10 min under unfolding conditions (4.4 M urea, pH 7.0). The fractions of U_c and U_t were then determined from the amplitudes of their refolding reactions as described above. Acid unfolding was not used here to avoid potential changes in the *cis/trans* equilibrium caused by the protonation of side chains at low pH.

Table 2 shows the fractions of the molecules with a *cis*-Pro161 in the folded and in the unfolded forms of the N2′ variants. In the unfolded form, the lowest *cis* content (7.5%) is observed for A160–P161 and the highest *cis* content (33.3%) is observed for W160–P161. The *cis/trans* equilibrium in unfolded N2′ is determined only by the local sequence, that is, by the nature of the amino acid at position 160 that precedes Pro161. The *cis* contents in the unfolded variants correlate very well with those determined for a series of peptides with the general sequence Ac-Ala-Xaa-Pro-Ala-Lys-NH_2 with all 20 amino acids at position Xaa[46] (Fig. 4a). This clear correlation indicates that, in the unfolded protein, there are no nonlocal interactions that might affect the *cis/trans* ratio at Pro161. A significant difference was found only for the Pro–Pro bond, which is 6.0% *cis* in the peptide, but 21.7% *cis* in the unfolded P160 variant of N2′. Currently, we have no compelling explanation for this observation. Possibly, a coupling between the isomerizations at Pro160 and Pro161 affects our refolding assays.

In the folded forms of all N2′ variants, the *cis* form of Pro161 dominates, and the corresponding *cis* contents cluster in a narrow range around 90% (Table 2). A lower value (75.5%) was observed only for the G160 variant, which probably shows increased local flexibility. There is only a very weak correlation (if any) between the *cis* contents as measured in the unfolded and the folded protein (Fig. 4b). This indicates that the shift in the *cis/trans* equilibrium upon folding is not caused by the residue that precedes Pro161 and suggests that the energy that is required for this shift does not originate from local contributions, although the mutations at position 160 modulate the stability (ΔG_D) of N2′ between 13 and 18 kJ mol^{-1}. Figure 4c shows that, in fact, the Gibbs free energy that is available to shift the *cis/trans* equilibrium at Pro161 during the folding of a particular variant does not correlate with its stability. This lack of correlation confirms the above conclusion and rules out that the residue at position 160 changes the stability directly by an effect on the *cis/trans* equilibrium at the 160–161 prolyl bond. A constant value of 10.5 kJ/mol is available in most variants to shift the *cis/trans* equilibrium at Pro161; lower values were observed only for the variants in which Pro161 was preceded by another Pro, by Gly, or by aromatic residues.

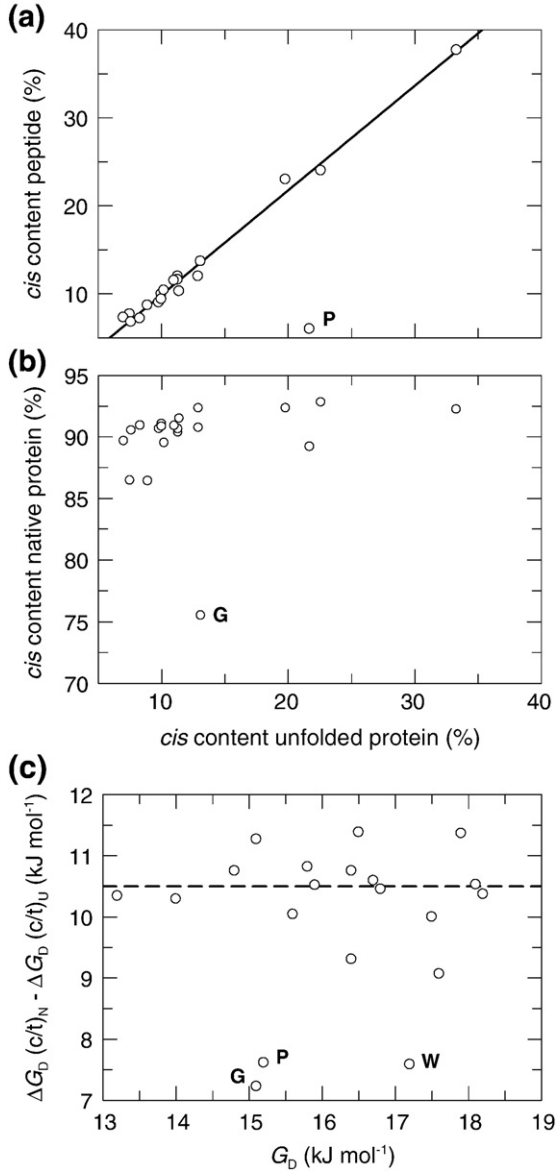


Fig. 4. (a) Correlation between the *cis* content in the unfolded N2′ variants with substitutions at position 160 and the *cis* content in the tetrapeptides acetyl-Ala-Xaa-Pro-Ala-Lys-NH_2.[46] (b) Correlation between the *cis* contents in the native and the unfolded forms of the N2′ variants with substitutions at position 160. (c) Correlation between the Gibbs free energies of denaturation at 15 °C (ΔG_D) with ΔG that results from the shift of the *cis*/*trans* equilibrium at Pro161. The broken line represents a value of 10.5 kJ mol^{-1}. Points that deviate from the general trends are identified by the nature of the amino acid at position 160.

The shift in the *cis*/*trans* equilibrium at Pro161 upon folding is, thus, not correlated with the nature of the amino acid substitution at position 160 as well as with the changes in stability that are caused by these substitutions. This rules out that the energy required to change this equilibrium originates from local interactions in the immediate environment of Pro161.

Mutational strategy to elucidate the origin of the folding energy that drives prolyl isomerization in the Pro161 loop

During folding, conformational energy must be transferred from the folded part of the molecule to the tip of the loop to change the *cis*/*trans* ratio at Pro161 almost 100-fold. To elucidate the molecular path for this transfer of energy, we altered residues in the chain segments that connect the tip of the Pro161 loop with the major body of the N2′ molecule (Fig. 1b). We then examined how these substitutions affect (i) the *cis*/*trans* ratio at Pro161, (ii) the conformational stabilities of the forms with *cis*- or *trans*-Pro161 isomers, and (iii) the folding transition states of these forms.

First, the residues in the connectors between the Asp160–Pro161 bond and the β sheet (residues 158–159 and 162–163, Fig. 1b) were changed to Ala. Then, the residues at the end of the β sheet were substituted by Ala (Q157, K163, T164, and Y165) or by Cys (T156), and finally, two residues that interact with and possibly stabilize the end of the β sheet (L125 and E127) were also changed to Ala (Fig. 1b).

To characterize the role of the connector regions further, we additionally created two insertion and two deletion variants. In the insertion variants InsA157 and InsA162, the connector peptides were extended either on the N- or on the C-terminal side of the 160–161 prolyl bond by single-Ala insertions. In the deletion variants Del159 and Del162, they were shortened by deleting either T159 or V162. As controls, we included a variant with a stabilizing substitution at position 138 (N138G) and a variant with a destabilizing substitution at position 129 (H129Q). Both are remote from the Pro161 loop.

The loop variants show different stabilities and Pro161 *cis*/*trans* equilibria

All variants were folded, but several of them were significantly less stable than the wild-type protein, as indicated by the results from thermal and urea-induced unfolding experiments (Table 3). Manipulations at the positions immediately adjacent to the Asp160–Pro161 bond are only weakly destabilizing. The substitutions of T159 or V162 by Ala or the deletion of T159 decreased the stability by only about 2 kJ mol^{-1}. The G158A substitution destabilized N2′ strongly (by 4.7 kJ mol^{-1}), probably because the backbone conformation at Gly158 (positive dihedral angles) is unfavorable for Ala. Insertions into the connectors between the β sheet and the D160–P161 prolyl bond (after the residues 157 or 161) also reduced the stability by about 4 kJ mol^{-1}. The strongest destabilizations were observed for substitutions at positions 163–165 in the β strand that leads away from the Pro161 loop (Fig. 1b). The Y165A replacement, in particular, reduced ΔG_D by 9.2 kJ mol^{-1}. In the folded protein, Y165 interacts with L125, and, in fact, the L125A substitution also destabilized N2′ by 7.5 kJ mol^{-1}.

Table 3. Stability data for variants of N2′ with substitutions in the 156–165 loop

	Thermal unfolding		Urea-induced unfolding				
Variant	T_M (°C)	ΔH_D (kJ mol^{-1})	$[urea]_M$ (M)	m (kJ mol^{-1} M^{-1})	ΔG_D (kJ mol^{-1})	ΔG_D(2.4 M) (kJ mol^{-1})	Folded N2′ *cis* content (%)
Wild type	37.8	310	2.71	6.6	17.9	2.1	89.7
L125A	30.3	310	1.57	6.5	10.2	−5.4	78.8
E127A	36.6	330	2.61	6.4	16.7	1.3	89.6
T156C	36.3	295	2.46	6.5	16.0	0.4	90.3
Q157A	35.5	312	2.34	6.7	15.7	−0.4	84.8
InsA157	35.1	306	2.14	6.6	14.1	−1.7	68.0
G158A	34.2	318	1.97	6.8	13.4	−2.5	50.4
Del159	35.7	299	2.43	6.6	16.0	0.2	92.9
T159A	36.0	289	2.37	6.4	15.1	−0.2	82.4
D160A	37.2	319	2.40	7.0	16.8	0.0	86.5
Del162	36.7	304	2.61	6.9	18.0	1.5	92.6
V162A	36.6	311	2.42	6.9	16.9	0.1	76.9
InsA162	34.5	306	2.12	6.5	13.8	−1.8	59.9
K163A	33.1	319	2.00	6.3	12.6	−2.5	86.5
T164A	33.1	275	1.93	7.0	13.5	−3.3	80.2
Y165A	30.0	298	1.33	6.6	8.8	−7.1	84.1
H129Q	33.6	300	2.11	7.0	14.8	−2.0	90.0
N138G	41.0	302	3.32	6.5	21.6	6.0	88.9

The stability parameters result from two-state analyses of thermal (columns 2 and 3) and urea-induced unfolding transitions (columns 4–7). For all proteins, the melting temperature (T_M), the van't Hoff enthalpy of denaturation at T_M (ΔH_D), the midpoint of the urea-induced unfolding transition ($[urea]_M$), the cooperativity value (m), and the Gibbs free energy of denaturation (ΔG_D) at 15 °C and 0 M urea are given. ΔG_D(2.4 M) depicts the ΔG_D value at 2.4 M urea. The thermal and the urea-induced unfolding transitions were measured in 100 mM K phosphate (pH 7.0). The precisions of T_M and ΔH_D are ±0.1 °C and ±12 kJ mol^{-1}, and those for m and ΔG_D are ±0.4 kJ mol^{-1} M^{-1} and ±0.9 kJ mol^{-1}, respectively. The *cis* contents (precisions ≤1%) in the native forms of the variants were determined by kinetic double-mixing experiments as described in Materials and Methods. All measurements were performed in 100 mM K phosphate (pH 7.0) at 15 °C.

In summary, mutations in the connectors between the β sheet and Pro161 have moderate effects on stability, shortening by single deletions is much better tolerated than extensions, and the β sheet residues at the base of the Pro161 loop provide major contributions to stability.

The *cis*/*trans* equilibria at the D160–P161 bond in the folded forms of the loop variants, as measured by our double-mixing procedure (0.1 s unfolding at pH 2.0, followed by refolding at pH 7.0), differ considerably for the individual variants (Table 3).

From the ratio of the *cis*/*trans* equilibrium constants in the unfolded and the folded forms of N2′, we calculated for all variants the Gibbs free energies that were necessary to shift the *cis*/*trans* equilibrium. We denote them as the "proline shift energies". They are a measure for the extent of coupling between conformational folding and prolyl isomerization. For the wild-type protein, it amounts to 11.4 kJ mol^{-1}, which is slightly higher than the previously measured value of 10.3 kJ mol^{-1}. Figure 5 compares for the individual substitutions in the Pro161 loop the change in the proline shift energy with the entire change in stability caused by these substitutions. The two control variants with substitutions at positions remote from the Pro161 loop are also included. H129Q is destabilizing and N138G is stabilizing, but both substitutions did not change the proline shift energy. This indicates that changes in the conformational stability at remote positions do not change the *cis*/*trans* equilibrium at Pro161.

The *cis*/*trans* equilibrium at Pro161 is linked, however, with the stability of the two-stranded β sheet that leads into the Pro161 loop. The mutations that destabilized the end of this β sheet (K163A, T164A, and Y165A) also decreased the proline shift energy

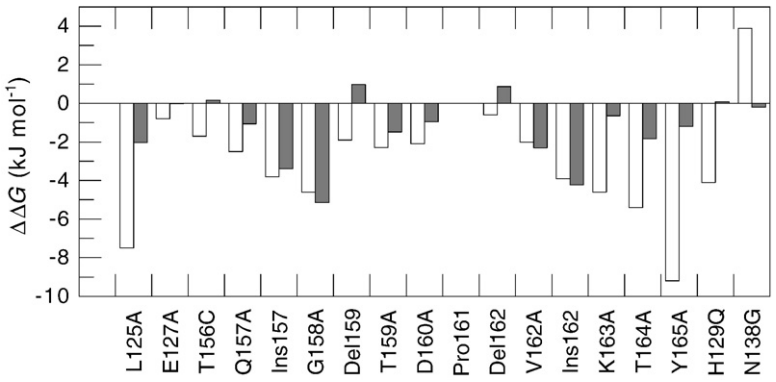


Fig. 5. Effect of the mutations in the Pro161 loop on the stability of N2′ ($\Delta\Delta G_{UN}$, white bars) and on the Gibbs free energy that results from the shift of the *cis*/*trans* equilibrium at Pro161 upon folding (the proline shift energy, gray bars). The $\Delta\Delta G_{UN}$ values at 2.4 M urea and 15 °C are taken from Table 3 ($\Delta\Delta G_{UN} = -\Delta\Delta G_D$). The differences in the proline shift energies are calculated from the "fraction *cis*" values in Table 3, using a fraction *cis* of 7.0% for the unfolded protein. Both $\Delta\Delta G$ values are given relative to the wild-type (Pro161) protein.

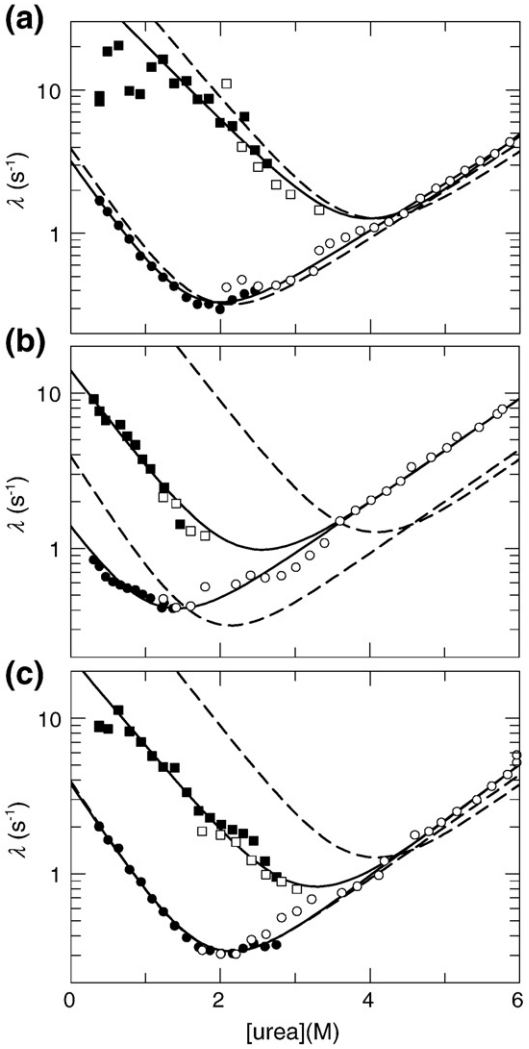


Fig. 6. Fast refolding (filled symbols) and unfolding (open symbols) kinetics of (a) N2′-T156C, (b) N2′-Y165A, and (c) N2′-InsA157. The apparent rate constants λ_3 (squares) and λ_2 (circles) are shown as a function of the urea concentration. Chevrons fitted to the experimental data on the basis of linear two-state models are shown by continuous lines. The results of the analysis are given in Table 4. The folding kinetics were measured after stopped-flow single mixing by the changes in fluorescence above 320 nm (excitation at 280 nm). The broken lines represent the fitted chevrons for wild-type N2′. All measurements were performed in 100 mM K phosphate (pH 7.0) at 15 °C.

(Fig. 5). These decreases amount to 1–2 kJ mol^{-1} and are significantly smaller than the destabilizations caused by these substitutions. Even the L125A mutation decreased the proline shift energy (by 2 kJ mol^{-1}), probably because it destabilized the β sheet indirectly by disrupting interactions with Y165. This linkage between the local stability at the end of the β sheet and the *cis*/*trans* equilibrium at Pro161 suggests that a part of the proline shift energy that becomes available during the folding of N2′ indeed originates from an energetic linkage with this β sheet.

The strongest decreases in the proline shift energy were observed after the G158A substitution and the Ala insertions after residue 157 or 162 in the connector regions. For all the substitutions or insertions in the connector regions, the loss in proline shift energy is approximately equal to the total loss in stability caused by these mutations (Fig. 5). This suggests that these substitutions affect the stability of N2′ because they interfere with the linkage between the conformational stability of the β sheet at the bottom and the prolyl bond at the tip of the Pro161 hairpin.

The substitutions in the Pro161 loop have distinct effects on the folding kinetics of the *cis* and *trans* forms

We measured the unfolding and refolding kinetics of all variants to elucidate how the variations in the Pro161 loop affect the individual stabilities and the individual folding mechanisms of the *cis* and the *trans* isomers of N2′. Such a simultaneous analysis was feasible, because the *cis* and the *trans* forms follow individual chevrons in the kinetic folding experiments and because these two conformational folding reactions are much faster than the isomerizations at Pro161 in the native and unfolded protein, which connect the two folding reactions kinetically (cf., the scheme in Fig. 1c). Figure 6 shows the chevron plots for three representative variants; the others are shown in Fig. S1. The analyses of these chevrons gave the kinetic parameters for unfolding and refolding, as well as the stabilities separately for the *cis* and *trans* forms (Table 4).

The substitutions in the Pro161 loop affected the folding kinetics of N2′ differently, as exemplified by the T156C, Y165A, and InsA157 variants (Fig. 6). The T156C substitution (Fig. 6a) leaves the stabilities as well as the folding kinetics of the *cis* and *trans* isomers virtually unchanged. The Y165A substitution (Fig. 6b) is strongly destabilizing at equilibrium (Table 3), and it affects the two isomers in a similar fashion. It decreases their folding rates, increases their unfolding rates, and, thus, destabilizes both the *cis* and the *trans* form. A peculiar class of variants is represented by Ins157 with an Ala insertion between Q157 and G158 (Fig. 6c). This mutation is overall destabilizing (Table 3), but this results exclusively from a strong deceleration of the refolding of the *cis* form and the corresponding decrease in its stability. The stability and the folding rate of the *trans* form are not affected. In fact, the entire decrease in stability, as measured at equilibrium for the InsA157 variant (Table 3), originates from the decrease in the refolding rate of the *cis* form.

The analysis of the folding kinetics of all variants provided separate sets of kinetic and thermodynamic parameters for the *cis* and *trans* forms of N2′. These values, in particular the changes in the equilibrium stability ($\Delta\Delta G_{UN}$) and in the activation energy of refolding ($\Delta\Delta G^{\ddagger}_{UN}$) caused by the mutations, are listed in Table 4. The $\Delta\Delta G_{UN}$ and $\Delta\Delta G^{\ddagger}_{UN}$ values for the individual positions along the Pro161 loop are compared for the *trans* form of N2′ in Fig. 7a and for the *cis* form in Fig. 7b. This graphical representation shows that the mutational effects on

Table 4. Unfolding and refolding kinetics of variants of N2′ with substitutions in the 156–165 loop

	Unfolding kinetics		Refolding kinetics										
			$U_t \rightarrow N_t$					$U_c \rightarrow N_c$					
Variant	k_{NU}	$m^{\ddagger}_{NU}$	k_{UN}	$m^{\ddagger}_{UN}$	ΔG_D	$\Delta\Delta G^{\ddagger}_{UN}$	$\Delta\Delta G_{UN}$	k_{UN}	$m^{\ddagger}_{UN}$	ΔG_D	$\Delta\Delta G^{\ddagger}_{UN}$	$\Delta\Delta G_{UN}$	Φ_{trans}	Φ_{cis}
Wild type	0.042	1.8	3.89	−4.1	10.9	0.0	0.0	149.2	−3.4	20.7	0.0	0.0		
L125A	0.081	1.9	1.75	−4.7	7.4	1.9	3.5	28.5	−3.2	14.6	4.0	6.1	0.54	0.65
E127A	0.049	1.8	3.33	−4.5	10.1	0.4	0.8	86.5	−3.2	19.7	1.3	1.0		
T156C	0.051	1.8	3.15	−4.1	9.9	0.5	1.0	98.5	−2.9	20.0	1.0	0.7		
Q157A	0.085	1.8	3.34	−4.4	8.8	0.4	2.1	68.9	−3.2	17.7	1.9	3.0	0.17	0.62
InsA157	0.036	2.0	3.72	−4.0	11.1	0.1	−0.2	24.7	−3.2	16.0	4.3	4.7		0.92
G158A	0.030	2.1	3.35	−4.5	11.3	0.4	−0.4	19.7	−3.4	15.3	4.9	5.4		0.90
Del159	0.042	1.9	3.26	−4.4	10.4	0.4	0.5	142.1	−3.5	21.2	0.1	−0.5		
T159A	0.050	1.9	4.16	−4.7	10.6	−0.2	0.3	58.0	−3.2	18.4	2.2	2.3		0.96
D160A	0.034	1.9	1.54	−4.2	9.1	2.2	1.8	83.2	−3.3	19.1	1.4	1.6		
Del162	0.037	1.9	3.31	−4.3	10.8	0.4	0.1	150.2	−3.4	20.6	0.0	0.1		
V162A	0.041	2.0	6.38	−4.1	12.1	−1.2	−1.2	72.0	−3.1	19.3	1.7	1.4		
InsA162	0.031	2.1	4.33	−4.2	11.9	−0.3	−1.0	22.1	−3.0	15.8	4.6	4.9		0.93
K163A	0.060	1.8	1.97	−4.0	8.4	1.6	2.5	27.6	−2.9	16.2	4.0	4.5	0.65	0.90
T164A	0.059	1.9	2.57	−4.6	9.1	1.0	1.8	45.3	−3.4	17.2	2.9	3.5		0.82
Y165A	0.093	1.8	1.30	−3.8	6.3	2.6	4.6	13.8	−3.5	12.1	5.7	8.6	0.57	0.66
H129Q	0.048	2.0	1.90	−3.8	7.7	1.7	3.2	74.4	−3.3	17.6	1.7	3.1	0.54	0.53
N138G	0.042	1.9	20.6	−4.0	14.9	−4.0	−4.0	685.6	−3.2	24.3	−3.7	−3.6	0.99	1.01

The kinetic parameters were determined in 100 mM K phosphate (pH 7.0) at 15 °C, pH 7.0, from the kinetic chevrons shown in Fig. 6 and Fig. S1 after a linear two-state analysis as described in Materials and Methods. The microscopic rate constants of refolding k_{UN} (s^{-1}) and unfolding k_{NU} (s^{-1}) at 0 M urea are given. $m^{\ddagger}_{UN}$ (kJ mol^{-1} M^{-1}) and $m^{\ddagger}_{NU}$ (kJ mol^{-1} M^{-1}) are the kinetic m values for refolding and unfolding, respectively. ΔG_D is the Gibbs free energy of denaturation calculated from the chevron plots. $\Delta\Delta G^{\ddagger}_{UN}$ is the difference in activation free energy of refolding between the variants and the reference protein (N2′) in the absence of denaturants. $\Delta\Delta G_{UN}$ is the difference in Gibbs free energy of refolding between the variants and the reference protein (N2′) in the absence of denaturants. All ΔG and $\Delta\Delta G$ are given in kilojoules per mole. Φ_{trans} and Φ_{cis} are the calculated Φ values in the absence of denaturants. Φ values are not given when $\Delta\Delta G_D$ is smaller than 2 kJ mol^{-1}.

the equilibrium and activation energies are small for the *trans* but large for the *cis* form of N2′.

Substitutions in the connector regions

The mutations in the connector regions have strikingly different effects on the *cis* and on the *trans* form of the protein. The substitutions at positions 158, 159, and 162 strongly destabilize the *cis* form but have almost no effect on the stability of the *trans* form (Fig. 7a and b). This suggests that residues 158–159 and 162 in the connectors between the Asp160–Pro161 prolyl bond and the β sheet (Fig. 1b) are important for the conformational stability of the *cis* but not of the *trans* form of N2′.

The activation energies of refolding ($\Delta\Delta G^{\ddagger}_{UN}$) of the *cis* and *trans* forms are also differently affected by the mutations in the connectors. In the *trans* form, the substitutions in this region do not change the value of $\Delta\Delta G^{\ddagger}_{UN}$. For the *cis* form, however, the increases in ΔG_{UN} (in particular by the G158A substitution) are correlated with equally strong increases in $\Delta G^{\ddagger}_{UN}$ (Fig. 7b). These close similarities between the $\Delta\Delta G_{UN}$ and $\Delta\Delta G^{\ddagger}_{UN}$ values indicate that in the *cis* form, the 157–159 region is equally folded in the native state and in the transition state of refolding.

The finding that, in the *trans* form, both ΔG_{UN} and $\Delta G^{\ddagger}_{UN}$ are unchanged by these substitutions shows that the connector region is in a similar conformation in the unfolded state, the transition state, and the folded state. We suggest that this conformation is unfolded-like. The insensitivity of the *trans* form to the G158A substitution indicates that G158 is equally flexible in all three states, and therefore, the reduction in chain entropy caused by the mutation to alanine is the same in all three forms. In the crystal structure of the *cis* form of N2′, G158 shows positive dihedral angles. Accordingly, the mutation to Ala is strongly destabilizing, because alanine is not well tolerated at such a position. The absence of such a destabilization in the *trans* form suggests that the backbone mobility at position 158 is so high that it accommodates Gly and Ala equally well.

Substitutions in the β sheet

The unfolding and refolding kinetics were also measured for six variants with substitutions in the terminal region of the two-stranded β sheet that leads into the Pro161 loop (Fig. 1b). The T156C substitution left the stability of N2′ almost unchanged, and therefore, it is not considered further in the analysis. All other replacements were destabilizing and reduced the overall stability of N2′ by values between 2.5 and 9.2 kJ mol^{-1} (Table 3). The kinetic analysis revealed that the extents of destabilization were strongly different for the *cis* and the *trans* form. In general, the individual substitutions destabilized the *cis* form more strongly than the *trans* form. The Y165A substitution decreased the stability of the *cis* form by 8.6 kJ mol^{-1} and that of the *trans* form by 4.6 kJ mol^{-1} (Table 4). A strong difference was found also for the substitution of L125, which is not part of the β sheet but interacts with Y165 in folded N2′.

The isomer-specific destabilizations of the *cis* and *trans* forms by mutations in the β sheet (Fig. 7a and

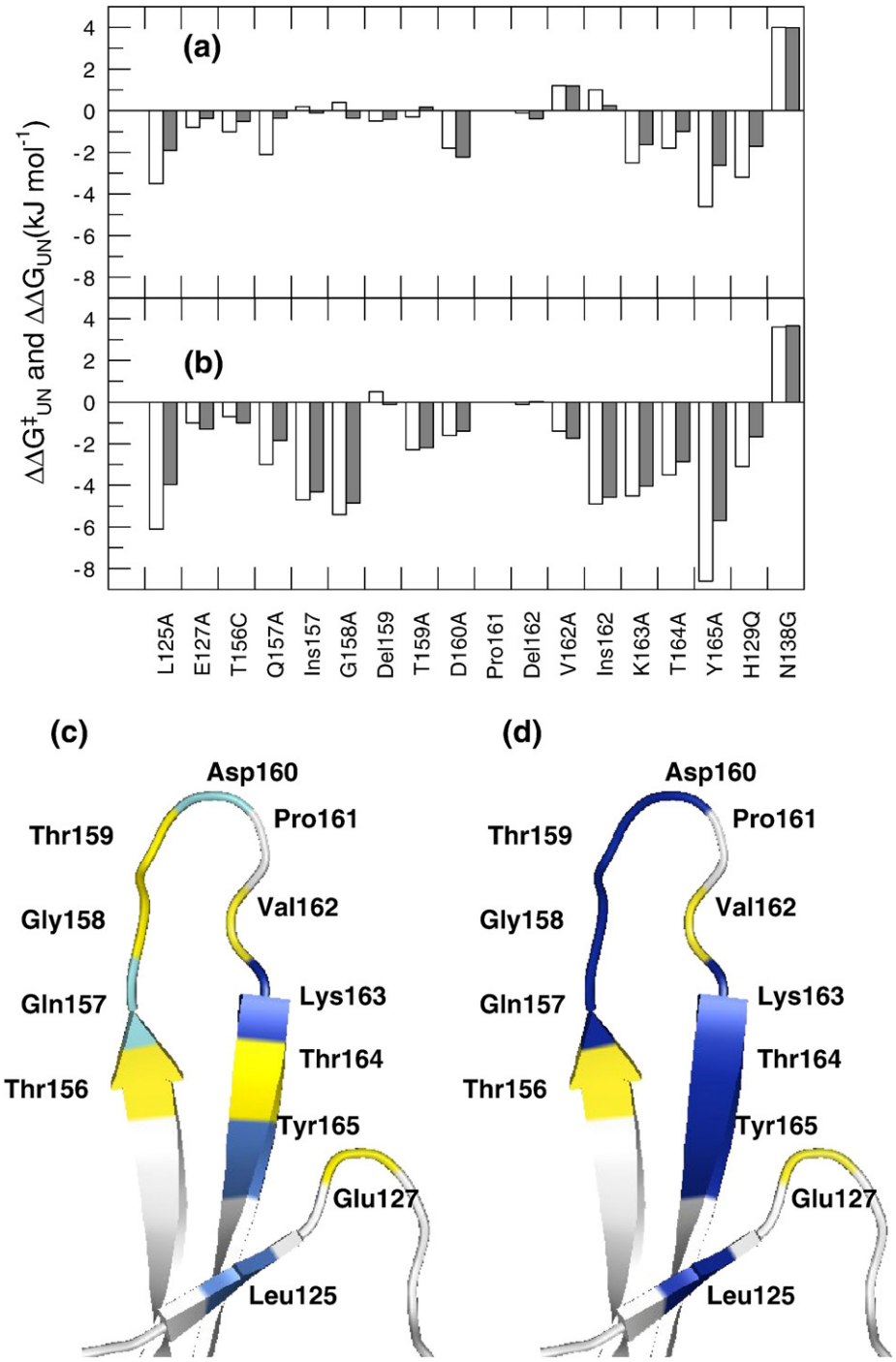


Fig. 7. Comparison of the changes in the equilibrium stability ($\Delta\Delta G_{UN}$, white bars) and in the activation energy of refolding ($\Delta\Delta G^{\ddagger}_{UN}$, gray bars) for the individual positions in the Pro161 loop for (a) the *trans* and (b) the *cis* form. The $\Delta\Delta G_{UN}$ and $\Delta\Delta G^{\ddagger}_{UN}$ values are taken from Table 4. Φ values for (c) the *trans* and (d) the *cis* forms plotted onto the structure of the Pro161 loop. The residues are color-coded according to their Φ values: light blue, $0.0<\Phi<0.3$; marine blue, $0.3<\Phi<0.6$; dark blue, $0.6<\Phi<1.0$; yellow, residues with a $\Delta\Delta G_{UN}<2.0$ kJ mol^{-1}.

b) and the shifts in the *cis*/*trans* equilibrium at Pro161 in folded N2′ (Fig. 5) by the same mutations demonstrate that there is an energetic linkage between the conformational stability of the β sheet and the shift upon folding of the *cis*/*trans* equilibrium at Pro161. In other words, a part of the energy that is required to favor the *cis* state of Pro161 originates from an improved conformational stability of the β sheet when Pro161 is *cis*. The total interaction energy between β sheet formation and Pro161 isomerization is certainly higher than the value that could be probed in this mutational analysis, because the six mutations reduced the stability of the β sheet, but did, of course, not abolish it. More strongly destabilizing mutations or combinations of destabilizing mutations could not be employed in this

analysis, because wild-type N2′ is only marginally stable.

Insertions and deletions in the connector regions

To investigate how the β sheet at the base and the Asp160–Pro161 prolyl bond at the tip of the hairpin communicate energetically, we varied the length of the two connectors by extending or shortening them by one residue. The insertion of an Ala residue on either side of Pro161, between the residues 157 and 158 (InsA157) or between the residues 162 and 163 (InsA162), had almost identical consequences. Both insertions reduced the conformational stability of the *cis* form by 4–5 kJ mol^{-1} each but left the stability of the *trans* form almost unchanged. At the same time, they decreased the fraction of the *cis* form of Pro161 from 90% (in the wild-type protein) to 68% (InsA157) and 60% (InsA162) (Table 3), indicating that 3.4 and 4.2 kJ mol^{-1} of proline shift energy were lost as a consequence of the two insertions. The similarity of all these energy differences (Fig. 5) suggests that the insertions are destabilizing because they increase the flexibility in the connectors and, thus, weaken the intramolecular energetic coupling between the β sheet and *cis*-Pro161. The energy that becomes available in the course of folding to shift the *cis*/*trans* equilibrium at Pro161 is, thus, reduced from 11.4 kJ mol^{-1} as in the wild-type protein to about 7.5 kJ mol^{-1}. In a reciprocal fashion, the conformational stabilities of the *cis* and *trans* isomers of N2′ are affected differently by the insertions in the connectors, presumably because the *cis* but not the *trans* form draws conformational stability from the β sheet at the base of the Pro161 loop.

The destabilization of the *cis* form by the insertions is reflected in an increase in the activation energy for refolding ($\Delta\Delta G^{\ddagger}_{UN}$) of the same magnitude (Fig. 7b), indicating that the energetic coupling along the connectors is fully established in the transition state of folding already. In the *trans* form, the same insertions in the connectors leave the stability and the folding kinetics unchanged, suggesting that such an energetic coupling does not exist, because the connectors are only weakly structured or unfolded when Pro161 is *trans* in the folded wild-type protein.

In reciprocal experiments, we shortened the two connectors by deleting either T159 or V162. These mutations, Del159 and Del162, left the stability and the folding kinetics of both the *cis* and the *trans* isomer virtually unchanged (Table 4; Fig. 7). In contrast to all other mutations, however, the two deletions led to increases in the proline shift energy (Fig. 5). Possibly, the corresponding shortening of the connectors improves the structural and energetic coupling between the β sheet and Pro161.

A side-by-side comparison of $\Delta\Delta G_{UN}$ and $\Delta\Delta G^{\ddagger}_{UN}$ values, as in Fig. 7, provides a good representation of the impact of amino acid substitutions on a protein. Short bars immediately identify positions that are unimportant for both the stability and the folding kinetics of a protein. Long white bars mark positions that are important for stability, and pairs of long white and gray bars identify positions that are important for the folding process, because their large contributions to stability are present in the folding transition state already.

Traditionally, the ratio of $\Delta\Delta G^{\ddagger}_{UN}$ and $\Delta\Delta G_{UN}$, the Φ value, is used to characterize the folding transition state, and for small proteins with a well-defined, narrow ensemble of transition state structures, the Φ value is equated with the fraction of native-like interactions that are established by the residue under investigation in the transition state already.[47,48] Φ values are reliable only when $\Delta\Delta G_{UN}$ is larger than a certain threshold, which is in the range of 2 kJ mol^{-1} for small proteins.[49,50] Φ is a ratio of energy differences and carries no information about their magnitude. The Φ values for the *cis* and *trans* forms of N2′, as calculated from the kinetic data, are given in Table 4 as well, and they are plotted on the structure of the Pro161 loop in Fig. 7c and d. Regions that are insensitive to mutations are marked as well.

The alternative representations of the mutational analysis in Fig. 7 highlight the distinct differences between the *cis* and *trans* forms of N2′. They show that the effects of mutations in the connectors on the equilibrium and activation energies are small in the *trans* but large in the *cis* form. This difference is most pronounced for the G158A mutation and for the insertions into the connectors (Fig. 7a and b). At the level of Φ values (Fig. 7c and d), the differences are evident for the 157–159 connector region, which shows high Φ values for the *cis* form. For the *trans* form, no information is available, because the $\Delta\Delta G_{UN}$ values are too small.

Bar graphs as in Fig. 7a and b are more informative, because data from insertions or deletions can be integrated readily and, in particular, because the mutations that do not change the stability and the activation energy of folding are represented as well. This "negative" information is equally important and is lost in Φ value calculations.

Conclusions

In proline-limited protein-folding reactions, conformational energy is needed to drive prolyl isomerization. In other words, an energetic linkage between the two processes must develop in the course of folding. For the N2 domain of the phage gene-3-protein, we were able to elucidate the structural and energetic basis of this linkage. This was possible because the shift in the *cis*/*trans* equilibrium at Pro161 upon folding could be measured for many protein variants and, in particular, because the stabilities and folding kinetics of the *cis* and *trans* forms of these variants could be determined separately.

Three structural elements are essential for the coupling between conformational folding and Pro161 isomerization: the 160–161 prolyl bond itself, a two-stranded β sheet that ends at T156 and starts at K163 (Fig. 1b), and the two connectors between Pro161 and these two β strands. In the folded wild-

type protein, about 10 kJ mol^{-1} is available to shift the *cis*/*trans* equilibrium at Pro161 towards the *cis* form. This proline shift energy originates largely, if not entirely, from the folding of the β sheet at the base of the Pro161 loop, evidently because the two strands of this sheet can improve their stabilizing interactions when Pro161 is *cis*. This stabilization is propagated to Pro161, because in the *cis* form, the connector peptides are native-like in the folded protein as well as in the folding transition state. Such an energetic linkage does not exist for the *trans* form. Here, the connector peptides are equally unstructured in folded N2′ and in the transition state of folding, and thus, the proline is uncoupled from the conformational stability of the β sheet. As a consequence, mutations in the connectors affected the folding and stability of the *cis* form but not of the *trans* form.

Local unfolding and refolding are thus used to break and to reestablish the energetic communication between a prolyl isomerization and the stability of an element of secondary structure (in this case, a two-stranded β sheet). Such a linkage between prolyl isomerization and local unfolding/refolding is probably not restricted to late steps in protein folding. Prolyl isomerizations are also used as slow molecular switches to control the function of regulatory proteins.[22–30,51] A local unfolding reaction, controlled by Pro213 *cis*→*trans* isomerization, is, in fact, required to activate the gene-3-protein of phage fd for the infection of *Escherichia coli* cells.[26,28] Prolyl isomerizations have also been suggested to regulate the binding of ligands to the Itk kinase[24] and the autoinhibition of the Crk protein.[29]

Our results provide a framework for the understanding of how molecular communication pathways in such proteins can be interrupted and regenerated by prolyl *cis*/*trans* isomerizations. The *cis*/*trans* equilibrium can be changed by interactions with effector proteins and/or by reversible chemical modifications, such as the phosphorylation of Ser–Pro and Thr–Pro motifs, and the intrinsically low rate of isomerization can be modulated by prolyl isomerases such as cyclophilin or Pin1.

Materials and Methods

Expression and purification of variants of N2′

The isolated N2 domain [residues 102–205 of the gene-3-protein of phage fd, extended by $(His)_6$] with the stabilizing mutation Q129H was used as the reference (pseudo-wild-type) protein. All mutated variants of N2′ were expressed and purified as described previously.[42] Site-directed mutagenesis of N2′ was performed by QuikChange (Stratagene, La Jolla, CA). Because the variants N2′-D160G and N2′-Q157A could not be produced with the original expression system, they were linked N-terminally with the SUMO protein Smt3 from *Saccharomyces cerevisiae*,[52] and the fusion protein was overproduced in *E. coli* BL21(DE3) (Stratagene). After lysis of the cells in 50 mM Tris/HCl, 50 mM NaCl, and 10 mM ethylenediaminetetraacetic acid, pH 8.0, with a microfluidizer and centrifugation, the SUMO-N2′ fusion proteins were obtained in soluble form. The proteins were purified by immobilized metal-affinity chromatography on a Ni-NTA column (elution with 250 mM imidazole) and cleaved with 1 μM of a fragment of the SUMO-protease Ulp1 (a gift from Dr. Lima, Sloan Kettering Institute, New York)[52] at 4 °C overnight. The SUMO fragment was removed by a second passage over a Ni-NTA column, and the N2′ variants were further purified by gel filtration on a Superdex 75 prep grade column (GE Healthcare, Uppsala, Sweden).

Equilibrium unfolding transitions

Samples of the N2′ variants (1.0 μM) were incubated for 1 h at 15 °C in 100 mM K phosphate, pH 7.0, and varying concentrations of urea. The fluorescence of the samples was measured in 1-cm cuvettes at 340 nm (10 nm bandwidth) after excitation at 280 nm (5 nm bandwidth) in a Hitachi F4010 fluorescence spectrometer. The data were analyzed according to a two-state model by assuming that ΔG_D as well as the fluorescence emissions of the folded and the unfolded form depend linearly on the urea concentration. A nonlinear least-squares fit with proportional weighting of the experimental data was used to obtain ΔG_D as a function of the urea concentration.[53]

The heat-induced unfolding transitions were measured in a Jasco J-600A spectropolarimeter equipped with a PTC 348 WI peltier element at a protein concentration of 4 μM in 100 mM K phosphate, pH 7.0 or pH 8.0, at a heating rate of 1 °C/min. The transitions were monitored by the increase of the CD signal at 222 nm with 1 nm bandwidth and 10 mm path length. The experimental data were analyzed on the basis of the two-state approximation,[54] with a heat capacity change ΔC_p of 6400 J mol^{-1} K^{-1} (calculated as described in Ref. 55).

Kinetic experiments

All urea-induced unfolding and refolding experiments were performed in 100 mM K phosphate, pH 7.0, at 15 °C at a final protein concentration of 1.0 μM. The denatured (in 5.0 M urea) N2′ variants were diluted 25-fold to 0.2 M urea by manual mixing in a Hitachi F4010 fluorescence spectrometer to measure the kinetics of the slow refolding reaction (λ_1). The two fast folding reactions (λ_2 and λ_3) were measured by using a DX.17MV stopped-flow spectrometer from Applied Photophysics (Leatherhead, UK). The native or the unfolded (in 4.4 M urea) protein was diluted 11-fold with urea solutions of varying concentrations. The kinetics were followed by the change in fluorescence above 320 nm after excitation at 280 nm (10 nm bandwidth) in an observation cell with 2 mm path length. A 0.5-cm cell with acetone was placed between the observation chamber and the photomultiplier to absorb scattered light from the excitation beam. The kinetics were measured at least 8 times under identical conditions and averaged to improve the signal-to-noise ratio. The *cis* contents in the unfolded N2′-D160X variants were determined from the relative amplitudes of the refolding reactions of the *cis* (λ_3) and *trans* (λ_2) isomers at 0.4 M urea, by taking into account that N_c and N_t differ slightly in fluorescence.

In the analysis of the unfolding and refolding kinetics of the individual variants, we assumed that the folding kinetics of the *cis* and the *trans* forms are kinetically isolated by the slow $U_t \leftrightarrow U_c$ and $N_t \leftrightarrow N_c$ isomerizations (Fig. 1c) and that the logarithms of the microscopic rate constants of

unfolding and refolding depend linearly on the urea concentration. The ΔG_{UN} values were determined from the ratio of the rate constants for refolding and unfolding [$\Delta G_{UN} = -RT\ln(k_{UN}/k_{NU})$]. $\Delta\Delta G_{UN}$ is the difference between the ΔG_{UN} values of the mutant and the wild-type protein. The $\Delta\Delta G^{\ddagger}_{UN}$ values were derived from the ratio of the refolding rate constants of the mutant and the wild-type protein [$\Delta\Delta G^{\ddagger}_{UN} = -RT\ln(k_{UN}(mt)/k_{UN}(wt))$], and Φ is the $\Delta\Delta G^{\ddagger}_{UN}/\Delta\Delta G_{UN}$ ratio.

Double-mixing stopped-flow experiments (interrupted unfolding) were performed to determine the *cis*/*trans* ratio in the folded forms of the N2′ variants at 15 °C. In the first step, 33 μM folded N2′ (in 100 mM K phosphate, pH 7.0) was diluted 11-fold with 100 mM glycine buffer to initiate unfolding at a final pH of 2.0. Under these conditions, conformational unfolding is complete within 10 ms and Pro161 *cis*/*trans* equilibration shows a time constant of 55 s.[42] After 100 ms, the N2′ variants were, thus, fully unfolded, but the *cis*/*trans* ratio remained virtually the same as in the folded protein. Refolding was induced after a delay time of 100 ms by a further 6-fold dilution to final folding conditions of 0.5 μM protein in 100 mM K phosphate, pH 7.0. The *cis* content in the folded protein was calculated from the ratio of the amplitudes of the refolding reactions of the *cis* (λ_3) and *trans* (λ_2) isomers, taking into account that N_c and N_t differ slightly in fluorescence. The experiments were repeated 20 times for all variants. Every measurement was analyzed separately and the determined *cis* contents differed by less than 1.5%.

The Gibbs free energy that is necessary to shift the *cis*/*trans* equilibrium (the proline shift energy) is equal to $-RT\ln(K_N/K_U)$, where K_N and K_U are the measured equilibrium constants for Pro161 *cis*/*trans* isomerization in the unfolded state ($K_U = [U_t]/[U_c]$) and in the native state ($K_N = [N_t]/[N_c]$).

Acknowledgements

We thank the members of our group for many discussion of this work. It was supported by grants from the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2009.02.021

References

1. Steinberg, I. Z., Harrington, W. F., Berger, A., Sela, M. & Katchalski, E. (1960). The configurational changes of poly-L-proline in solution. *J. Am. Chem. Soc.* **82**, 5263–5279.
2. Cheng, H. N. & Bovey, F. A. (1977). Cis–trans equilibrium and kinetic studies of acteyl-L-proline and glycyl-L-proline. *Biopolymers*, **16**, 1465–1472.
3. Grathwohl, C. & Wüthrich, K. (1981). NMR studies of the rates of proline cis–trans isomerization in oligopeptides. *Biopolymers*, **20**, 2623–2633.
4. 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.
5. Hagerman, P. J., Schmid, F. X. & Baldwin, R. L. (1979). Refolding behavior of a kinetic intermediate observed in the low pH unfolding of ribonuclease A. *Biochemistry*, **18**, 293–297.
6. Schmid, F. X. & Baldwin, R. L. (1978). Acid catalysis of the formation of the slow-folding species of RNase A: evidence that the reaction is proline isomerization. *Proc. Natl Acad. Sci. USA*, **75**, 4764–4768.
7. 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.
8. Schmid, F. X. (1982). Proline isomerization in unfolded ribonuclease A. *Eur. J. Biochem.* **128**, 77–80.
9. Goto, Y. & Hamaguchi, K. (1982). Unfolding and refolding of the constant fragment of the immunoglobulin light chain. *J. Mol. Biol.* **156**, 891–910.
10. Nall, B. T. (1985). Proline isomerization and protein folding. *Comments Mol. Cell. Biophys.* **3**, 123–143.
11. Schreiber, G. & Fersht, A. R. (1993). The refolding of *cis*- and *trans*-peptidylprolyl isomers of Barstar. *Biochemistry*, **32**, 11195–11203.
12. Lilie, H., Rudolph, R. & Buchner, J. (1995). Association of antibody chains at different stages of folding: prolyl isomerization occurs after formation of quaternary structure. *J. Mol. Biol.* **248**, 190–201.
13. 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, UK.
14. Pappenberger, G., Bachmann, A., Muller, R., Aygun, H., Engels, J. W. & Kiefhaber, T. (2003). Kinetic mechanism and catalysis of a native-state prolyl isomerization reaction. *J. Mol. Biol.* **326**, 235–246.
15. Schmid, F. X. (2005). Prolyl isomerization in protein folding. In *Protein Folding Handbook* (Buchner, J. & Kiefhaber, T., eds), pp. 916–945, Wiley-VCH, Weinheim, Germany.
16. Schmid, F. X. (1983). Mechanism of folding of ribonuclease A. Slow refolding is a sequential reaction via structural intermediates. *Biochemistry*, **22**, 4690–4696.
17. Kiefhaber, T., Schmid, F. X., Willaert, K., Engelborghs, Y. & Chaffotte, A. (1992). Structure of a rapidly formed intermediate in ribonuclease T1 folding. *Protein Sci.* **1**, 1162–1172.
18. Veeraraghavan, S., Rodriguez-Gdiharpour, S., MacKinnon, C., Mcgee, W. A., Pierce, M. M. & Nall, B. T. (1995). Prolyl isomerase as a probe of stability of slow-folding intermediates. *Biochemistry*, **34**, 12892–12902.
19. Aumüller, T. & Fischer, G. (2008). Bioactivity of folding intermediates studied by the recovery of enzymatic activity during refolding. *J. Mol. Biol.* **376**, 1478–1492.
20. Sakata, M., Chatani, E., Kameda, A., Sakurai, K., Naiki, H. & Goto, Y. (2008). Kinetic coupling of folding and prolyl isomerization of beta2-microglobulin studied by mutational analysis. *J. Mol. Biol.* **382**, 1242–1255.
21. 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), pp. 198–203, AAAS, Washington, DC.
22. 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.
23. 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. *Nat. Struct. Biol.* **9**, 900–905.
24. Andreotti, A. H. (2003). Native state proline isomerization: an intrinsic molecular switch. *Biochemistry*, **42**, 9515–9524.
25. Fischer, G. & Aumüller, T. (2003). Regulation of peptide bond cis/trans isomerization by enzyme catalysis and its implication in physiological processes. *Rev. Physiol., Biochem. Pharmacol.* **148**, 105–150.
26. Eckert, B., Martin, A., Balbach, J. & Schmid, F. X. (2005). Prolyl isomerization as a molecular timer in phage infection. *Nat. Struct. Mol. Biol.* **12**, 619–623.
27. Vogel, M., Bukau, B. & Mayer, M. P. (2006). Allosteric regulation of Hsp70 chaperones by a proline switch. *Mol. Cell*, **21**, 359–367.
28. Eckert, B. & Schmid, F. X. (2007). A conformational unfolding reaction activates phage fd for the infection of *Escherichia coli*. *J. Mol. Biol.* **373**, 452–461.
29. Sarkar, P., Reichman, C., Saleh, T., Birge, R. B. & Kalodimos, C. G. (2007). Proline cis–trans isomerization controls autoinhibition of a signaling protein. *Mol. Cell*, **25**, 413–426.
30. Lu, K. P., Finn, G., Lee, T. H. & Nicholson, L. K. (2007). Prolyl cis–trans isomerization as a molecular timer. *Nat. Chem. Biol.* **3**, 619–629.
31. Fischer, G. & Schmid, F. X. (1999). Peptidyl–prolyl cis/trans isomerases. In *Molecular Biology of Chaperones and Folding Catalysts* (Bukau, B., ed.), pp. 461–489, Harwood Academic Publishers, New York, NY.
32. Göthel, S. F. & Marahiel, M. A. (1999). Peptidyl–prolyl cis–trans isomerases, a superfamily of ubiquitous folding catalysts. *Cell. Mol. Life Sci.* **55**, 423–436.
33. Galat, A. (1999). Variations of sequences and amino acid compositions of proteins that sustain their biological functions: an analysis of the cyclophilin family of proteins. *Arch. Biochem. Biophys.* **371**, 149–162.
34. Schmid, F. X. (2002). Prolyl isomerases. *Adv. Protein Chem.* **59**, 243–282.
35. 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.
36. Higgins, K. A., Craik, D. J., Hall, J. G. & Andrews, P. R. (1988). Cis–trans isomerization of the proline residue in insulin studied by ^{13}C NMR spectroscopy. *Drug Des. Delivery*, **3**, 159–170.
37. Chazin, W. J., Kördel, J., Drakenberg, T., Thulin, E., Brodin, P., Grundström, T. & Forsén, S. (1989). Proline isomerism leads to multiple folded conformations of calbindin D9k: direct evidence from two-dimensional NMR spectroscopy. *Proc. Natl Acad. Sci. USA*, **86**, 2195–2198.
38. 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.
39. Alexandrescu, A. T., Hinck, A. P. & Markley, J. L. (1990). Coupling between local structure and global stability of a protein: mutants of staphylococcal nuclease. *Biochemistry*, **29**, 4516–4525.
40. Adjadj, E., Naudat, V., Quiniou, E., Wouters, D., Sautiere, P. & Craescu, C. T. (1997). Solution structure of Lqh-8/6, a toxin-like peptide from a scorpion venom—structural heterogeneity induced by proline cis/trans isomerization. *Eur. J. Biochem.* **246**, 218–227.
41. Feng, Y., Hood, W. F., Forgey, R. W., Abegg, A. L., Caparon, M. H., Thiele, B. R. *et al.* (1997). Multiple conformations of a human interleukin-3 variant. *Protein Sci.* **6**, 1777–1782.
42. Jakob, R. & Schmid, F. X. (2008). Energetic coupling between native-state prolyl isomerization and conformational protein folding. *J. Mol. Biol.* **377**, 1560–1575.
43. 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. *Nat. Struct. Biol.* **5**, 140–147.
44. Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics*, **14**, 51–55.
45. 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.
46. 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.
47. Fersht, A. R., Matouschek, A. & Serrano, L. (1992). The folding of an enzyme. 1. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* **224**, 771–782.
48. Paci, E., Vendruscolo, M., Dobson, C. M. & Karplus, M. (2002). Determination of a transition state at atomic resolution from protein engineering data. *J. Mol. Biol.* **324**, 151–163.
49. Garcia-Mira, M. M., Boehringer, D. & Schmid, F. X. (2004). The folding transition state of the cold shock protein is strongly polarized. *J. Mol. Biol.* **339**, 555–569.
50. Fersht, A. R. & Sato, S. (2004). Phi-value analysis and the nature of protein-folding transition states. *Proc. Natl Acad. Sci. USA*, **101**, 7976–7981.
51. OuYang, B., Pochapsky, S. S., Dang, M. & Pochapsky, T. C. (2008). A functional proline switch in cytochrome P450cam. *Structure*, **16**, 916–923.
52. Mossessova, E. & Lima, C. D. (2000). Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol. Cell*, **5**, 865–876.
53. 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.
54. Mayr, L. M., Kiefhaber, T. & Schmid, F. X. (1993). Prolyl isomerizations as rate-determining steps in the folding of ribonuclease T_1. In *Protein Folding: In Vivo and In Vitro* (Cleland, J. L., ed.), 1st edit., vol. 526, pp. 142–155. American Chemical Society, Washington, DC.
55. Privalov, P. L. & Gill, S. J. (1988). Stability of protein structure and hydrophobic interaction. *Adv. Protein Chem.* **39**, 191–234.

Teilarbeit E:

Jakob, R.P., Zoldák, G. and Schmid, F.X**.**
Chaperone domains adapt prolyl isomerases for a generic function in protein folding., *Manuskript in Vorbereitung*

Running title: Chaperone domains adapt prolyl isomerases for a generic function

Chaperone domains adapt prolyl isomerases for a generic function in protein folding

Roman P. Jakob, Gabriel Zoldák and Franz X. Schmid*

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

*Corresponding author

Franz X. Schmid
Biochemie
Universität Bayreuth
D-95440 Bayreuth, Germany

Tel: ++49 921 553660
Fax: ++49 921 553661
E-mail: fx.schmid@uni-bayreuth.de

Abbreviations used: FKBP12, FKBP52, human FK-506 binding protein of 12 kDa and 52 kDa; FkpA, periplasmic FKBP of E. coli; SlyD, product of the slyD (sensitive-tolysis) gene, cytosolic *E. coli* chaperone; SlyD*, 1-165 fragment of SlyD; IF, "insert-in-flap" domain of SlyD; FKBP12+IF, variant of FKBP12 in which the flap was replaced by the IF domain of SlyD; N2, N2 domain of the gene 3 protein of phage fd; Abz, amino benzoyl; pNA, 4-nitroanilide.

Abstract

The *cis*/*trans* isomerization of peptide bonds before proline (prolyl bonds) is an intrinsically slow reaction. It determines the rate of many protein folding reactions, and it is catalyzed by prolyl isomerases. Cellular folding is usually assisted by prolyl isomerases that belong to the FKBP family, such as trigger factor, SlyD and FkpA. In addition to the catalytic prolyl isomerase domain, these enzymes contain chaperone domains, which increase the affinity for refolding protein substrates. To determine the sequence specificities of these prolyl isomerases, we used two libraries, one of peptide variants and of protein variants. In both libraries, the amino acid before a conserved proline was replaced by every naturally occurring amino acid. For the peptide substrates, the FKBP-type prolyl isomerases show a high substrate specificity towards the residue before the proline, with a strong preference for hydrophobic side chains. For the protein substrates, the catalytic prolyl isomerase efficiencies are overall improved and, at the same time, the high substrate specificity of the prolyl isomerase site is almost abolished. The binding of a refolding protein chain to the chaperone domains apparently overrides the inherently high sequence specificity of the prolyl isomerase site and thus solves the problem of generic substrate recognition during protein folding.

Key words: trigger factor; SlyD; folding helpers; folding catalysis; folding mechanism, chaperone, ribosome

Introduction

The isomerization of peptide bonds before proline (Xaa-Pro bonds, prolyl bonds) is an intrinsically slow reaction. The *cis*/*trans* equilibrium and the kinetics of interconversion are largely determined by the chemical nature of the amino acid Xaa that precedes proline (Figure 1a) [1-3]. Prolyl isomerizations determine the rates of many protein folding reactions [4-6], and they are catalyzed by prolyl isomerases [7-11]. Most of the prolyl isomerases that assist in cellular protein folding contain a catalytic domain that is homologous to human FKBP12 (Figure 1b). The trigger factor (Figure 1d) [12-15], SlyD (Figure 1c) [16-19] and FkpA [20-22] of prokaryotes and FKBP52 of eukaryotes [23] belong to this family of folding enzymes. In addition to the catalytic FKPB domain, these prolyl isomerases contain chaperone domains, which increase the affinity for folding protein substrates.

The classical assay for prolyl isomerases exploits the specificity of proteases (such as chymotrypsin or trypsin), which do not accept peptides with *cis* prolyl bonds as substrates [7]. For tetrapeptides of the general formula succinyl-Ala-Xaa-Pro-Phe-4-nitroanilide, the molecules with a *cis* Xaa-Pro bond are cleaved by chymotrypsin only after the *cis* $\rightarrow$ *trans* isomerization. The concentration of the protease must be high in this coupled assay to warrant a rapid cleavage of the *trans* peptide. By using such an assay, it was found that members of the FKBP family show high substrate specificities with a strong preference for hydrophobic residues at the position Xaa before proline [24-26]. The initial characterization of human FKBP12 indicated that it catalyzes Phe-Pro isomerization about 1000-fold better than Glu-Pro isomerization [24].

Prolyl peptide bonds mostly occur in coil regions or in chain reversals, such as β turns, often at or near the protein surface. As a consequence, polar or charged residues are strongly overrepresented at the position preceding proline [27], and thus they should be poor substrates for the folding enzymes of the FKBP family. The sequence specificities of prolyl isomerases towards refolding protein chains are not known, due to the lack of a suitable assay system.

The protease-coupled peptide assay can be used only for enzymes that are resistant to proteolysis, and therefore it cannot be applied to most multi-domain prolyl isomerases, including the trigger factor and SlyD. Recently, however, a sensitive protease-free peptide assay for prolyl isomerases was developed. For this assay, tetrapeptides are used that carry an aminobenzoyl (Abz) group at the aminoterminus and a para nitroanilide (pNA) group at the carboxyterminus. In these peptides, the fluorescence of the Abz group is sensitive to the isomeric state of the prolyl bond.

To determine the sequence specificities of prolyl isomerases in short tetrapeptides as well as in refolding proteins, we employed a peptide and a protein library. In a tetrapeptide library of the general formula Abz-Ala-Xaa-Pro-Phe-pNA, the position Xaa before proline was occupied by all 20 natural amino acids. To create a corresponding protein library, we used the N2 domain of the gene-3-protein of phage fd as a model system. Pro161 is largely *cis* in folded N2 [28,29] and more than 90 % of all refolding N2 molecules undergo *trans* → *cis* isomerization at Pro161 in the rate-limiting step of folding [30]. By directed mutagenesis, we created a library of 20 N2 variants in which Pro161 is preceded by all 20 naturally occurring amino acids [31]. This library was then used to characterize the sequence specificity of prolyl isomerases in a folding reaction.

We find different substrate specificities of the prolyl isomerases for peptides and refolding proteins. In the presence of a chaperone domain, the catalytic efficiency is greatly improved and, at the same time, the high substrate specificity of the prolyl isomerase site, as observed for the peptide substrates, is almost abolished. Binding to the chaperone domains apparently overrides the inherent sequence specificity of the prolyl isomerase site and thus solves the problem of generic substrate recognition during protein folding.

Results

The substrate specificities of prolyl isomerases in the presence of chaperone domains

By using a limited set of tetrapeptides in a chymotrypsin-coupled assay, Harrison and Stein discovered the high substrate specificity of FKBP12 with regard to the residue Xaa preceding proline [24,25]. Our new protease-free peptide assay confirms this high specificity. 60 nM FKBP12 accelerates the isomerization at Leu-Pro 12-fold, but leaves the isomerization rate of Glu-Pro virtually unchanged (Figure 2a). From measurements of the isomerization rate as a function of the FKBP12 concentration (as in Figure 2), the catalytic efficiencies (k_{cat}/K_M) for all 20 Xaa-Pro sequences could be derived (Table 1). They differ about 500-fold between Leu-Pro (the best) and Asp/Glu-Pro (the worst). The k_{cat}/K_M values determined with the new protease-free assay are about five- to tenfold higher than the values determined previously for a limited subset of peptides by the protease-coupled assay [24,25]. We assume that the presence of the aminobenzoyl group in our peptides improves the interaction with FKBP12. The relative activities towards the different peptide substrates, as obtained by the two assays are, however, very similar (Table 1).

FKBP12 is a poor catalyst of prolyl isomerization in protein folding [32]. Its efficiency is strongly improved when the chaperone domain of SlyD (the IF domain) is grafted into a loop of FKBP12 near the active site to create the chimeric protein FKBP12-IF [33]. The insertion of this chaperone domain did not change the activity and the high specificity of the prolyl isomerase active site of FKBP12 towards Xaa-Pro in our tetrapeptide library (Table 1). Only the activity towards the Glu and Asp containing peptides was slightly improved (4- to 5-fold).

Next, we employed the protein library of N2 variants with all amino acids before Pro161 to examine the substrate specificity of FKBP12 in a protein folding reaction. As mentioned, FKBP12 alone is a poor catalyst of protein folding, and therefore up to 2.5 μM FKBP12 had to be used for measuring catalysis (Figure 3a). In the case of Leu160-N2, catalysis is clearly visible, for Asp160-N2 catalysis could not be detected, suggesting that the k_{cat}/K_M value is smaller than $0.5 \cdot 10^3$ $M^{-1}s^{-1}$. The catalytic efficiencies of FKBP12 towards 19 variants of N2 are shown in Table 1. Figure 4a compares them with the corresponding values measured for the tetrapeptide library. The activities of FKBP12 towards Xaa-Pro in the library of refolding proteins are two to three orders of magnitude lower than the corresponding values for Xaa-Pro in the peptide library, but the substrate specificity towards residue Xaa is similarly high in the folding proteins and in the tetrapeptides.

The kinetic parameters change drastically when the IF domain with its chaperone function is inserted into FKBP12. Now, concentrations as low as 1-10 nM of FKBP12+IF are sufficient to strongly increase the rate of folding for all members of the N2 protein library, as shown for Leu160-N2 and Asp160-N2 in Figure 3b. For library members with polar amino acids before Pro161, such as Ser, Thr, Asn, and Arg, the k_{cat}/K_M value is about 1000-fold increased when the IF domain is inserted into FKBP12. The k_{cat}/K_M values for all variants are listed in Table 1, the activities as measured for the tetrapeptide and protein libraries are compared in Figure 4b. This comparison reveals that, in the presence of the IF domain, the catalytic efficiency of the FKBP domain in protein folding is not only much higher than in the assays with the tetrapeptide library but also almost identical for all members of the N2 library. Most variants show k_{cat}/K_M values between $2 \cdot 10^6$ and $5 \cdot 10^6$ $M^{-1}s^{-1}$, suggesting that the nature of the residue Xaa before proline is no longer important for the catalytic efficiency of FKBP12 when the IF chaperone domain is present.

The chaperone domain of the chimeric protein FKBP12-IF was derived from the prolyl isomerase SlyD of *E. coli*. To examine the function of the IF domain in its natural environment, we used our peptide and protein libraries to measured the substrate specificities of SlyD* as well. SlyD* is a C-terminally truncated form of SlyD that consists of the FKBP

and chaperone domains. The data are shown in Table 1 and Figure 4c. As observed for the chimera FKBP12-IF, the activity of SlyD* towards polar residues before Pro was strongly increased, when protein substrates were used, leading to values that approach those for the hydrophobic residues (Figure 4c). Surprisingly, the catalytic efficiency of SlyD* in protein folding was about fivefold lower than the efficiency of the artificial chimera FKBP12-IF.

Trigger factor is located at the bacterial ribosome, close to the exit site for nascent protein chains, and presumably involved in *de-novo* protein folding. It consists of a ribosome-binding domain, a prolyl isomerase domain of the FKBP type, and a chaperone domain (Figure 1d), and it catalyzes proline-limited protein folding reactions *in vitro* very well [13,34,35]. As the other FKBPs, trigger factor shows a high substrate specificity towards Xaa-Pro bonds in tetrapeptides (Table 1) and, in its activity profile, it strongly resembles FKBP12-IF (Figure 4b,d). In the protein folding assays with the library of N2 variants, the activity of trigger factor is very high and almost independent of the nature of residue Xaa before proline. This is exemplified by the less than threefold difference in the k_{cat}/K_M values for the Leu160 and Asp160 variants of N2 (Figure 3d) and the similar heights of the blue bars in Figure 4d.

Enzyme kinetics of catalyzed protein folding

k_{cat}/K_M is a composite value that characterizes the efficiency of an enzyme at low substrate concentration. To obtain the kinetic parameters k_{cat} and K_M separately, we measured the enzyme kinetics of catalyzed folding. In these experiments, the initial rate of folding was determined as a function of the substrate protein concentration. Catalyzed and uncatalyzed folding occur in the presence of a folding catalyst and must be accounted for in the analysis. To dissect the refolding kinetics into the contributions from catalyzed and uncatalyzed folding, we used a procedure originally developed for prolyl isomerization in peptides [36]and adapted for catalyzed protein folding [33,34]. The range of substrate concentrations is restricted to less than 10 μM in these experiments, because the uncatalyzed reaction dominates at high substrate concentration.

We measured the enzyme kinetics of catalyzed folding for three representative members of our library of N2 variants: the Ala160 variant with the "default" amino acid alanine, the Asp160 variant with a charged residue, and the Leu160 variant with a hydrophobic residue before the proline. As the enzymes we employed FKBP12-IF, SlyD* and trigger factor. The corresponding Michaelis-Menten plots are shown in Figure 5. The scatter in the data is considerable, because uncatalyzed folding contributes, but in all cases, the initial rates of catalyzed folding show saturation behavior and apparently obey the Michaelis-Menten

equation. The kinetic constants k_{cat} and K_M derived from this analysis are shown in Table 2. The results for the three prolyl isomerases and the three substrate proteins are remarkably similar. The k_{cat} values are in the range of 2 – 3 s^{-1} for all three enzymes, the K_M values are in the range of 0.6 to 3.8 μM. For all three enzymes, the K_M for the Ala160 and Leu160 variants are similar, and the K_M for the Asp160 variant is about twofold larger. The k_{cat}/K_M values calculated from the individual parameters agree well with the composite k_{cat}/K_M values determined from experiments at increasing enzyme concentrations, as in Figure 3. This confirms that folding catalyzed by the various prolyl isomerases in this study can be analyzed by the Michaelis-Menten equation.

For FKBP12 without a chaperone domain, a saturation behavior as in Figure 5 could not be observed in the range of substrate concentrations that are accessible for measuring the catalyzed folding of N2 (Figure 6), which indicates that the substrate affinity of FKBP12 for folding protein chains is very low. Kinetic constants could also not be determined for the catalysis of prolyl isomerization, because of the limited solubility of the tetrapeptides of our library.

Discussion

In the absence of a chaperone domain, prolyl isomerases of the FKBP family bind both peptide and protein substrates with a low affinity, but a very high specificity with regard to the amino acid Xaa before the proline. Catalysis is up to 1000-fold more efficient when the proline is preceded by a hydrophobic residue instead of a negatively charged residue. This high local sequence specificity is abolished when proline-limited protein folding reactions are catalyzed by prolyl isomerases that contain a chaperone domain.

The presence of the chaperone domain has two important consequences for catalyzed protein folding. Its affinity for unfolded or non-native protein chains determines the measured K_M value of catalyzed folding and shifts it into the range of about 1 μM, and the catalytic efficiency becomes independent of the local sequence context of the prolyl bond to be isomerized.

Apparently, these are generic effects of the chaperone domains. The chaperone domains of SlyD and trigger factor differ strongly in size and folding pattern (Figure 1), yet they change the catalytic properties of the corresponding prolyl isomerase domains in the same fashion. Moreover, when the chaperone domain of SlyD is artificially introduced into human

FKBP12, a folding enzyme is created that is virtually indistinguishable from trigger factor with regard to substrate specificity and catalytic efficiency.

The chaperone domains of trigger factor and SlyD bind to protein substrates with similar affinity, resulting in K_M values around 1 μM. The catalysis by trigger factor of the folding of other proteins and its inhibition by permanently unfolded proteins were also governed by K_M and K_I values in the range of 1 μM [34,35]. A moderate affinity is probably important for the proper function of the chaperone domains of folding enzymes. Too low an affinity would compromise substrate recognition, too high an affinity would interfere with substrate transfer to the prolyl isomerase active site.

The turnover numbers of catalyzed folding obtained for the different prolyl isomerases in this study are also very similar (2 – 3 s^{-1}) and agree well with the value of 1.3 s^{-1}, determined previously for the catalysis of folding of ribonuclease T1 by trigger factor [34]. For FKBP12, Park et al. derived estimates for k_{cat} and K_M of 600 s^{-1} and 0.5 mM, respectively, for the catalysis of Leu-Pro isomerization in a tetrapeptide at 5 °C [26]. A comparison of these values with the k_{cat} of 2 – 3 s^{-1} and the K_M of about 1 μM, as observed here for the catalysis of proline-limited folding by the chaperone-containing prolyl isomerases indicates that these enzymes bind very well to their substrates, but the catalytic turnover is decelerated by almost three orders of magnitude. The very low k_{cat} values certainly do not reflect the rate of the catalyzed *cis*/*trans* isomerization in the protein substrates. We suggest that the k_{cat} value is related with the dissociation of the substrate protein from the chaperone binding site and the efficiency of productive binding in the prolyl isomerase site. Permanently unfolded substrate proteins dissociate from the chaperone domain of trigger factor with a rate of approximately 8 s^{-1} [37], which is about fourfold higher than the observed k_{cat} value. Probably, not all dissociation events from the chaperone site lead to productive binding at the prolyl isomerase site. In some cases, non-productive binding of chain regions without a proline might occur as well as premature rebinding to the chaperone site. It is also unknown, whether complete dissociation from the chaperone domain is required for reaching the prolyl isomerase active site.

This leads to a simple model for the mechanism of action of prolyl isomerases that use chaperone domains for substrate binding. Initially, the chaperone domain binds to non-native protein substrates with an adjusted, micro molar affinity. The binding equilibrium is highly dynamic, which allows efficient intramolecular transfer to the prolyl isomerase site, as well as the clearance of the binding site from misfolded proteins, if necessary. Productive transfer to

the prolyl isomerase active site occurs as a stochastic process, it must compete with unproductive binding and with re-binding to the chaperone site. This suggests that both measured kinetic parameters of catalyzed folding, k_{cat} and K_M, are determined by the chaperone domain of these folding enzymes. K_M reflects its affinity for the refolding protein and k_{cat} is largely determined by the rate of substrate dissociation from the chaperone domain. The actual catalysis of isomerization at the prolyl isomerase site is several orders of magnitude faster than substrate delivery and therefore its inherently high sequence specificity does not affect the overall catalytic efficiency.

Materials and Methods

Expression and purification of variants of N2 and the prolyl isomerases

As the wild-type protein the isolated N2 (residues 102-205 of mature G3P, extended by (His)6) domain with the mutation Q129H was used. The mutated variants of N2 were expressed and purified as described [30,31]. Recombinant FKBP12, SlyD and trigger factor were produced and purified as described [38]

Peptide assay for prolyl isomerases

The prolyl isomerase activities of the proteins were measured by a protease-free assay (Zoldak et al, manuscript in preparation). For the assay, the peptide substrates Abz-Ala-Xaa-Pro-Phe-pNa (3 M) were dissolved in anhydrous trifluoroethanol containing 0.5 M LiCl. Under these conditions, about 50% of the peptide molecules are in the *cis* conformation. Upon 600-fold dilution into aqueous buffer, the *cis* content decreases to about 10 %. The kinetics of the decrease in *cis* content was measured by the change in fluorescence at 416 nm after excitation at 316 nm in a Jasco FP-6500 fluorescence spectrophotometer. The assays were performed in 50 mM Hepes/NaOH, (pH 7.8) at 15 °C. Under these conditions, the *cis*-to-*trans* isomerization of the prolyl bond was a mono-exponential process.

Folding experiments

For the folding experiments a Hitachi F4010 fluorescence spectrometer was used. N2 variants were unfolded by incubating the protein in 0.1 M K-phosphate, pH 7.0, and 5.0 M urea at 15 °C. Refolding at 15 °C was initiated by a 25-fold dilution to a final concentration of 0.33 µM in 0.1 M K-phosphate, pH 7.0, 0.2 M urea, and the desired concentration of the

prolyl isomerase. The folding reaction was followed by the increase in protein fluorescence at 340 nm (10 nm band width) after excitation at 280 nm (5 nm band width). Under these conditions N2 refolding was a monoexponential process and, in all folding experiments, the small contribution of the prolyl isomerases to the fluorescence was subtracted from the measured values. The rate constants of folding were determined by using GraFit 3.0 (Erithacus Software, Staines, UK).

Enzyme kinetics of catalyzed folding

In the Michaelis-Menten kinetic experiments the initial velocities of N2 folding were determined from the progress curves of folding in the presence of 5 nM FKBP12+IF, SlyD(1-165) and trigger factor, respectively, under the conditions described above. Measurements were carried out between 0.1 and 10 μM N2. Below 0.1 μM the signal-to-noise ratio was too low. Both uncatalyzed and catalyzed folding occurs in these experiments. The relative contribution of uncatalyzed folding increases linearly with N2 concentration, and the initial rate of catalyzed folding would be progressively overestimated when determined simply from the initial slope of the progress curve of folding. Kofron *et al.* [36] developed a method to account for both uncatalyzed and enzyme-catalyzed prolyl isomerization in a peptide. We used this method and analyzed the data as described [34,38].

Acknowledgements

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

References

1. Grathwohl, C. & Wüthrich, K. (1981). NMR studies of the rates of proline cis-trans isomerization in oligopeptides. *Biopolymers* **20**, 2623-2633.
2. 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**, 449-60.
3. Fischer, G. (2000). Chemical aspects of peptide bond isomerisation. *Chemical Society Reviews* **29**, 119-127.
4. 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.
5. Schmid, F. X. & Baldwin, R. L. (1978). Acid catalysis of the formation of the slow-folding species of RNase A: evidence that the reaction is proline isomerization. *Proc. Natl. Acad. Sci. U.S.A.* **75**, 4764-4768.
6. Schmid, F. X. (2005). Prolyl Isomerization in Protein Folding. In *Protein Folding Handbook* (Buchner, J. K., T., ed.), pp. 916-945. Wiley-VCH, Weinheim.
7. 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.
8. Fischer, G. (1994). Peptidyl-prolyl cis/trans isomerases and their effectors. *Angew. Chem. Int. Ed. Engl.* **33**, 1415-1436.
9. Göthel, S. F. & Marahiel, M. A. (1999). Peptidyl-prolyl cis-trans isomerases, a superfamily of ubiquitous folding catalysts [Review]. *Cell. Mol. Life Sci.* **55**, 423-436.
10. 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.
11. Schmid, F. X. (2002). Prolyl isomerases. *Adv.Protein Chem.* **59**, 243-282.
12. Crooke, E. & Wickner, W. (1987). Trigger factor: a soluble protein that folds pro-OmpA into a membrane-assembly-competent form. *Proc. Natl. Acad. Sci. USA* **84**, 5216-5220.
13. Stoller, G., Rücknagel, K. P., Nierhaus, K., Schmid, F. X., Fischer, G. & Rahfeld, J.-U. (1995). Identification of the peptidyl-prolyl cis/trans isomerase bound to the Escherichia coli ribosome as the trigger factor. *EMBO J.* **14**, 4939-4948.
14. Hesterkamp, T. & Bukau, B. (1996). Identification of the prolyl isomerase domain of Escherichia coli trigger factor. *FEBS Lett.* **385**, 67-71.
15. Ferbitz, L., Maier, T., Patzelt, H., Bukau, B., Deuerling, E. & Ban, N. (2004). Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature* **431**, 590-596.
16. Roof, W. D., Horne, S. M., Young, K. D. & Young, R. (1994). slyD, a host gene required for phi X174 lysis, is related to the FK506-binding protein family of peptidyl-prolyl cis-trans-isomerases. *J. Biol. Chem.* **269**, 2902-10.
17. Wülfing, C., Lombardero, J. & Plückthun, A. (1994). An Escherichia coli protein consisting of a domain homologous to FK506- binding proteins (FKBP) and a new metal binding motif. *J. Biol. Chem.* **269**, 2895-901.
18. Roof, W. D. & Young, R. (1995). Phi X174 lysis requires slyD, a host gene which is related to the FKBP family of peptidyl-prolyl cis-trans isomerases. *FEMS Microbiol. Rev.* **17**, 213-8.

19. Hottenrott, S., Schumann, T., Plückthun, A., Fischer, G. & Rahfeld, J. U. (1997). The Escherichia coli SlyD is a metal ion-regulated peptidyl-prolyl cis/trans-isomerase. *J. Biol. Chem.* **272**, 15697-15701.
20. Bothmann, H. & Plückthun, A. (2000). The periplasmic Escherichia coli peptidylprolyl cis,trans-isomerase FkpA - I. Increased functional expression of antibody fragments with and without cis-prolines. *J. Biol. Chem.* **275**, 17100-17105.
21. Ramm, K. & Plückthun, A. (2000). The periplasmic Escherichia coli peptidylprolyl cis,trans-isomerase FkpA - II. Isomerase-independent chaperone activity in vitro. *J. Biol. Chem.* **275**, 17106-17113.
22. Ramm, K. & Plückthun, A. (2001). High enzymatic activity and chaperone function are mechanistically related features of the dimeric E. coli peptidyl-prolyl-isomerase FkpA. *J. Mol. Biol.* **310**, 485-98.
23. Peattie, D. A., Harding, M. W., Fleming, M. A., Decenzo, M. T., Lippke, J. A., Livingston, D. J. & Benasutti, M. (1992). Expression and Characterization of Human FKBP52, an Immunophilin That Associates with the 90-kDa Heat Shock Protein and Is a Component of Steroid Receptor Complexes. *Proc.Natl.Acad.Sci.USA* **89**, 10974-10978.
24. Harrison, R. K. & Stein, R. L. (1990). Substrate specificities of the peptidyl prolyl cis-trans isomerase activities of cyclophilin and FK-506 binding protein: evidence for the existence of a family of distinct enzymes. *Biochemistry* **29**, 3813-3816.
25. Stein, R. L. (1993). Mechanism of enzymatic and nonenzymatic prolyl cis- trans isomerization. *Adv. Protein Chem.* **44**, 1-24.
26. Park, S. T., Aldape, R. A., Futer, O., DeCenzo, M. T. & Livingston, D. J. (1992). PPIase catalysis by human FK506-binding protein proceeds through a conformational twist mechanism. *J. Biol. Chem.* **267**, 3316-24.
27. Reimer, U. & Fischer, G. (2002). Local structural changes caused by peptidyl-prolyl cis/trans isomerization in the native state of proteins. *Biophys. Chem.* **96**, 203-212.
28. 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. *Nat. Struct. Biol.* **5**, 140-147.
29. 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.
30. Jakob, R. & Schmid, F. X. (2008). Energetic coupling between native-state prolyl isomerization and conformational protein folding. *J. Mol. Biol.* **377**, 1560-1575.
31. Jakob, R. P. & Schmid, F. X. (2009). Molecular determinants of a native-state prolyl isomerization. *J. Mol. Biol.* **387**, 1017-31.
32. Tropschug, M., Wachter, E., Mayer, S., Schönbrunner, E. R. & Schmid, F. X. (1990). Isolation and sequence of an FK506-binding protein from N. crassa which catalyses protein folding. *Nature* **346**, 674-676.
33. Knappe, T. A., Eckert, B., Schaarschmidt, P., Scholz, C. & Schmid, F. X. (2007). Insertion of a chaperone domain converts FKBP12 into a powerful catalyst of protein folding. *J. Mol. Biol.* **368**, 1458-68.
34. Scholz, C., Stoller, G., Zarnt, T., Fischer, G. & Schmid, F. X. (1997). Cooperation of enzymatic and chaperone functions of trigger factor in the catalysis of protein folding. *EMBO J.* **16**, 54-58.
35. Scholz, C., Mücke, M., Rape, M., Pecht, A., Pahl, A., Bang, H. & Schmid, F. X. (1998). Recognition of protein substrates by the prolyl isomerase trigger factor is independent of proline residues. *J. Mol. Biol.* **277**, 723-732.

36. 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**, 6127-6134.
37. Maier, R., Scholz, C. & Schmid, F. X. (2001). Dynamic association of trigger factor with protein substrates. *J. Mol. Biol.* **314**, 1181-1190.
38. Scholz, C., Eckert, B., Hagn, F., Schaarschmidt, P., Balbach, J. & Schmid, F. X. (2006). SlyD proteins from different species exhibit high prolyl isomerase and chaperone activities. *Biochemistry* **45**, 20-33.
39. Harrison, R. K. & Stein, R. L. (1992). Mechanistic Studies of Enzymic and Nonenzymic Prolyl Cis- Trans Isomerization. *J. Am. Chem. Soc.* **114**, 3464-3471.
40. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1993). Atomic Structures of the Human Immunophilin FKBP-12 Complexes with FK506 and Rapamycin. *J. Mol. Biol.* **229**, 105-124.
41. Weininger, U., Haupt, C., Schweimer, K., Graubner, W., Kovermann, M., Bruser, T., Scholz, C., Schaarschmidt, P., Zoldak, G., Schmid, F. X. & Balbach, J. (2009). NMR solution structure of SlyD from Escherichia coli: spatial separation of prolyl isomerase and chaperone function. *J. Mol. Biol.* **387**, 295-305.

Table 1. Catalytic efficiencies of FKBP12, FKBP12+IF1, SlyD* and trigger factor for prolyl isomerization in the library of peptide substrates and during refolding of the N2 variants

residue before Pro	cis→trans Isomerization in tetrapeptide[a]					N2 refolding[b]			
	FKB12[c]	FKBP12	FKBP12+IF	SlyD*	trigger factor	FKBP12	FKBP12+IF	SlyD*	trigger factor
Xaa	k_{cat}/K_M ($mM^{-1}s^{-1}$)	k_{cat}/K_M ($mM^{-1}s^{-1}$)	k_{cat}/K_M ($mM^{-1}s^{-1}$)	k_{cat}/K_M ($mM^{-1}s^{-1}$)	k_{cat}/K_M ($mM^{-1}s^{-1}$)	k_{cat}/K_M ($\mu M^{-1}s^{-1}$)	k_{cat}/K_M ($\mu M^{-1}s^{-1}$)	k_{cat}/K_M ($\mu M^{-1}s^{-1}$)	k_{cat}/K_M ($\mu M^{-1}s^{-1}$)
Ala	53·± 3	230 ± 10	300 ± 60	230 ± 2	1600 ± 100	0.003 ± 0.0006	5.2·± 0.3	0.93·± 0.05	3.6·± 0.2
Cys		1200 ± 90	1000 ± 50	34 ± 7	2100 ± 200	0.0039 ±0.0005	4.3 ± 0.3	0.60·± 0.05	2.0·± 0.1
Asp		5.0 ± 0.2	28 ± 5	25 ± 1	26 ± 1	<0.0005	1.6·± 0.1	0.33·± 0.04	1.7 ± 0.7
Glu	0.6 ± 0.05	5.5 ± 0.5	25 ± 1	3.6 ± 0.1	30 ± 1	<0.0005	3.8·± 0.1	0.33·± 0.03	2.0·± 0.1
Phe	620· ± 140	1130 ± 30	1200 ± 200	2620 ± 100	3200 ± 200	0.019 ±0.002	7.6·± 0.7	1.55·± 0.09	4.4·± 0.5
Gly	1.2·± 1.0	190 ± 20	60 ± 6	3 ± 0.3	90 ± 5	0.0024 ±0.001	2.3·± 0.5	0.98·± 0.06	3.4·± 0.3
His	28·± 10	170 ± 17	90 ± 190	380 ± 10	160 ± 30	0.0035 ±0.0002	3.1·± 0.2	0.84·± 0.10	3.4·± 0.5
Ile		770 ± 90	390 ± 10	1510 ± 40	600 ± 50	0.0093 ±0.001	4.0·± 0.3	1.41·± 0.18	4.4·± 0.4
Lys	28·± 5	74 ± 14	170 ± 20	330 ± 10	420 ± 40	0.0048 ±0.0005	3.6·± 0.2	0.62·± 0.08	3.6·± 0.2
Leu	640 ± 50	2300 ± 200	1000 ± 600	1100 ± 100	1000 ± 100	0.016 ±0.002	3.5·± 0.4	1.27·± 0.12	4.4·± 0.6
Met		630 ± 50	1200 ± 100	1100 ± 40	810 ± 80	0.013 ±0.001	4.1·± 0.2	1.19·± 0.12	3.6·± 0.3
Asn		260 ± 45	260 ± 40	90 ± 10	90 ± 30	0.0031 ±0.0004	4.2·± 0.2	0.57·± 0.06	3.8·± 0.6
Pro		8.6 ± 2	14 ± 1	1.1 ± 0.1	36 ± 1	0.0038 ±0.001	2.3·± 0.3	0.24·± 0.02	1.6·± 0.2
Gln		88 ± 10	74 ± 7	22 ± 2	200 ± 20	0.0068 ±0.0005	3.4·± 0.2	0.39·± 0.07	3.0·± 0.2
Arg		260 ± 30	130 ± 20	460 ± 10	530 ± 40	0.0044 ±0.0003	4.8·± 0.6	0.68·± 0.08	3.4·± 0.3
Ser		260 ± 30	160 ± 10	190 ± 10	1400 ± 100	0.0036 ±0.0004	3.3·± 0.2	0.84·± 0.04	3.1·± 0.4
Thr		350 ± 50	120 ± 10	200 ± 10	1100 ± 100	0.0036 ±0.0007	3.7·± 0.2	0.59·± 0.07	2.8·± 0.2
Val	170·± 80	940 ± 30	450 ± 40	590 ± 100	480 ± 50	0.005 ±0.001	3.3·± 0.2	0.79·± 0.09	4.6·± 0.3
Trp		220 ± 20	570 ± 90	1200 ± 40	680 ± 70	n.d.	n.d.	n.d.	n.d.
Tyr		770 ± 90	2400 ± 200	2320 ± 140	1800 ± 100	0.011 ±0.001	5.6·± 0.5	1.61·± 0.16	5.1·± 0.6

[a] The k_{cat}/K_M values were obtained as described in the legend of Figure 2.

[b] The kinetic experiments were performed as described in the legend of Figure 3. [c]data for the peptide substrate Suc-Ala-Xaa-Pro-Phe-pNA, as taken from ref. [39]

Table 2: Enzyme kinetic parameters for the catalysis of N2-D160X refolding by FKBP12+IF1, SlyD and TF

N2-variant	N2 refolding	Michaelis-Menten experiments[b]		
	k_{cat}/K_M[a] ($M^{-1}s^{-1}$)	turnover number k_{cat} (s^{-1})	Mchaelis constant K_M (μM)	Composite value k_{cat}/K_M ($M^{-1}s^{-1}$)
FKBP12+IF1				
Ala	$5.2 \cdot 10^6$	3.9	0.78	$5.0 \cdot 10^6$
Asp	$1.6 \cdot 10^6$	2.6	1.87	$1.4 \cdot 10^6$
Leu	$3.5 \cdot 10^6$	2.9	0.83	$3.5 \cdot 10^6$
SlyD				
Ala	$0.93 \cdot 10^6$	1.9	1.9	$0.98 \cdot 10^6$
Asp	$0.33 \cdot 10^6$	2.2	3.8	$0.57 \cdot 10^6$
Leu	$1.27 \cdot 10^6$	2.7	1.8	$1.5 \cdot 10^6$
trigger factor				
Ala	$3.6 \cdot 10^6$	2.2	1.06	$2.1 \cdot 10^6$
Asp	$1.7 \cdot 10^6$	2.9	2.14	$1.4 \cdot 10^6$
Leu	$4.4 \cdot 10^6$	2.3	0.58	$4.0 \cdot 10^6$

The kinetic experiments were performed as described in the legend to Figure 5.

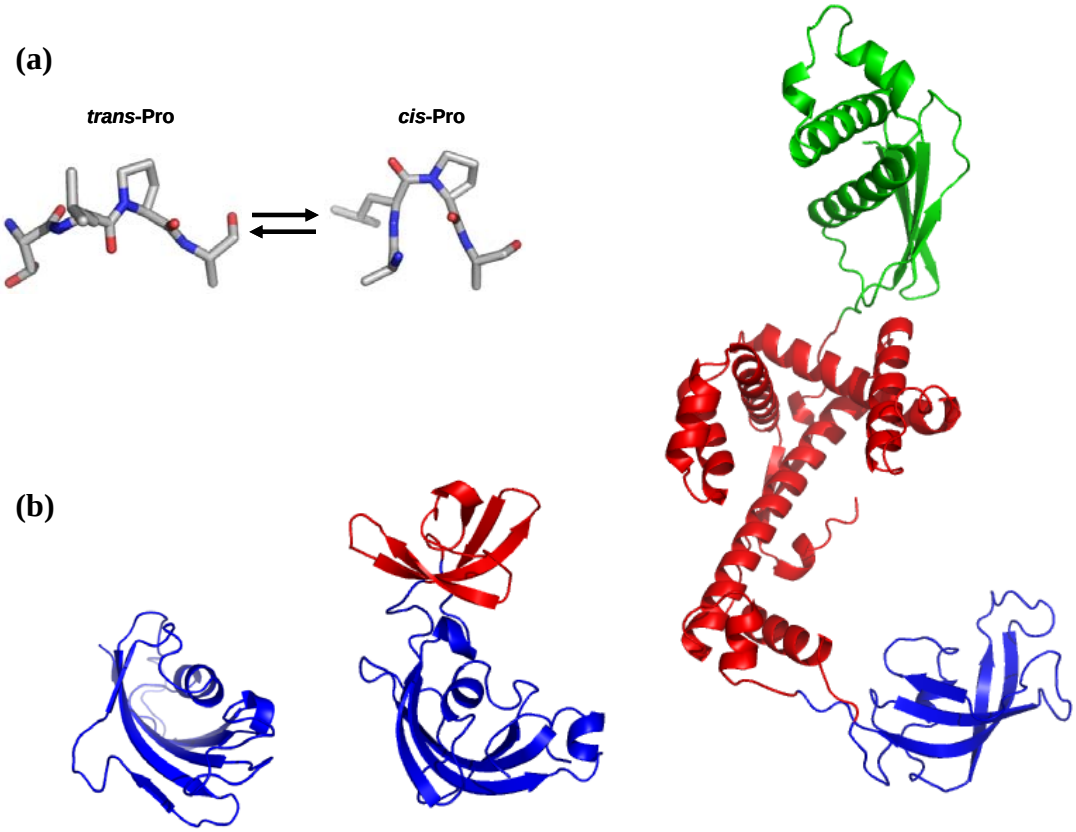


Figure 1 (a) *Trans* and *cis* isomer of a prolyl peptide bond. (b) Backbone structure of human FKBP12 (left), SlyD (middle) and trigger factor (right). The prolyl isomerase domains are shown in blue, the chaperone domains in red, the ribosome-binding domain of trigger factor is shown in green. The PDB files 1FKF for FKBP12 [40], 2k8I for *E. coli* SlyD [41], 1W26 for *E. coli* trigger factor [15]were used to prepare the Figures.

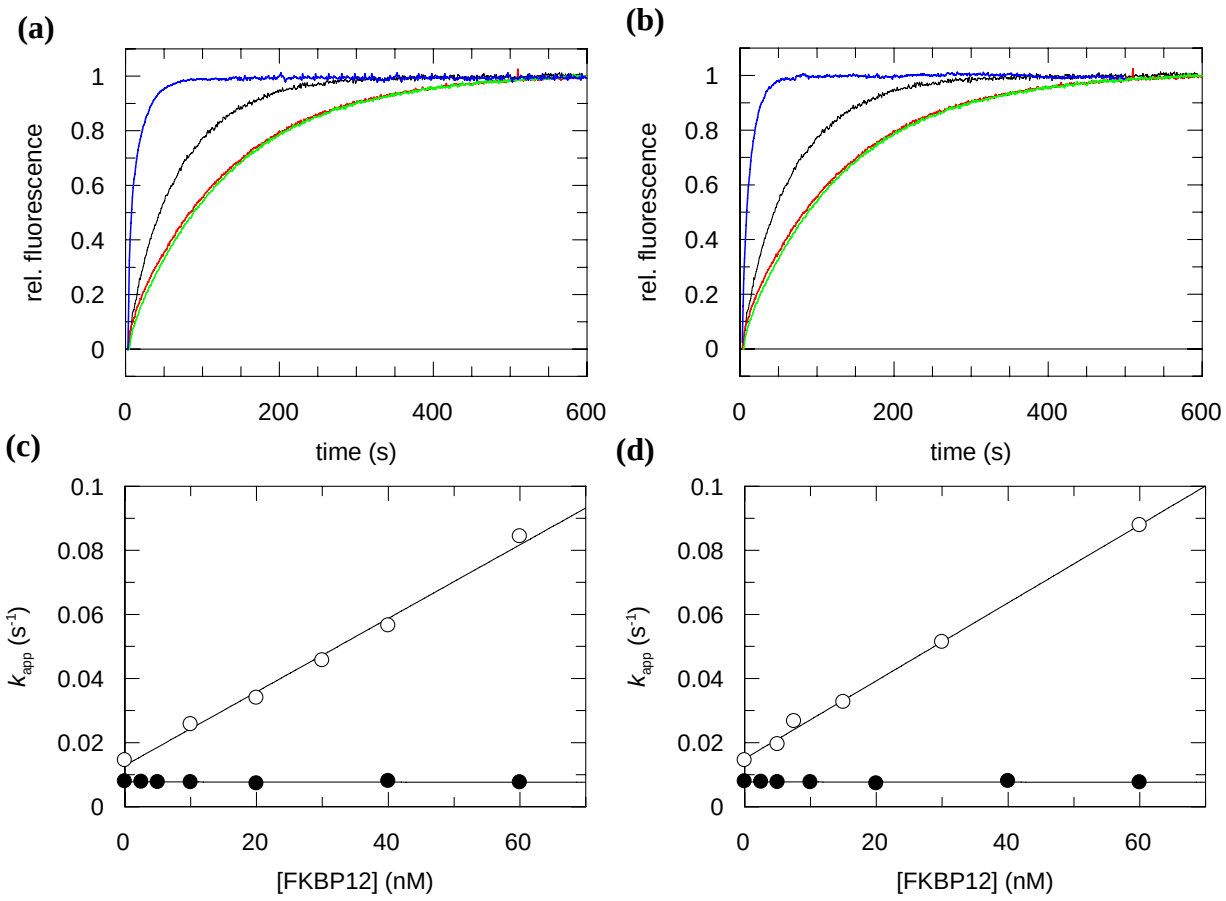


Figure 2. Catalysis of the *cis*/*trans* isomerization of 5 µM Abz-Ala-Phe-Pro-Phe-pNa and 5 µM Abz-Ala-Glu-Pro-Phe-pNa in the absence (black and green curve) and in the presence of 60 nM prolyl isomerase (blue and red curve, respectively for (a) FKBP12 and (b) FKBP12-IF. Measurements were performed in 50 mM HEPES, pH 7.8 at 15 °C. Fluorescence was measured at 416 nm (bandwidth 5 nm) after excitation at 316 nm (bandwidth 3 nm). The reaction was started by the addition of the peptide dissolved in the dry 0.55 M LiCl/TFE. (c,d) Dependence of the apparent rate constants, k_{app}, for the *cis*/*trans* isomerization reaction of Abz-Ala-Phe-Pro-Phe-pNa (open circles) and Abz-Ala-Glu-Pro-Phe-pNa (closed circles) on the concentration of (c) FKBP12 and (d) FKBP12-IF. Values of 1.1 ± 0.2 × 10^6 $M^{-1}s^{-1}$ (FKBP12) and 1.2 ± 0.6 × 10^6 $M^{-1}s^{-1}$ (FKBP12+IF) were obtained for k_{cat}/K_M from the slopes of the lines for Abz-Ala-Phe-Pro-Phe-pNa. The dependence of the apparent rate constants, k_{app}, for the *cis*/*trans* isomerization of Abz-Ala-Glu-Pro-Phe-pNa (closed circles) on enzyme concentration is also shown but the catalytic efficiency for this peptide was obtained by the variation of the enzyme concentration at µM range.

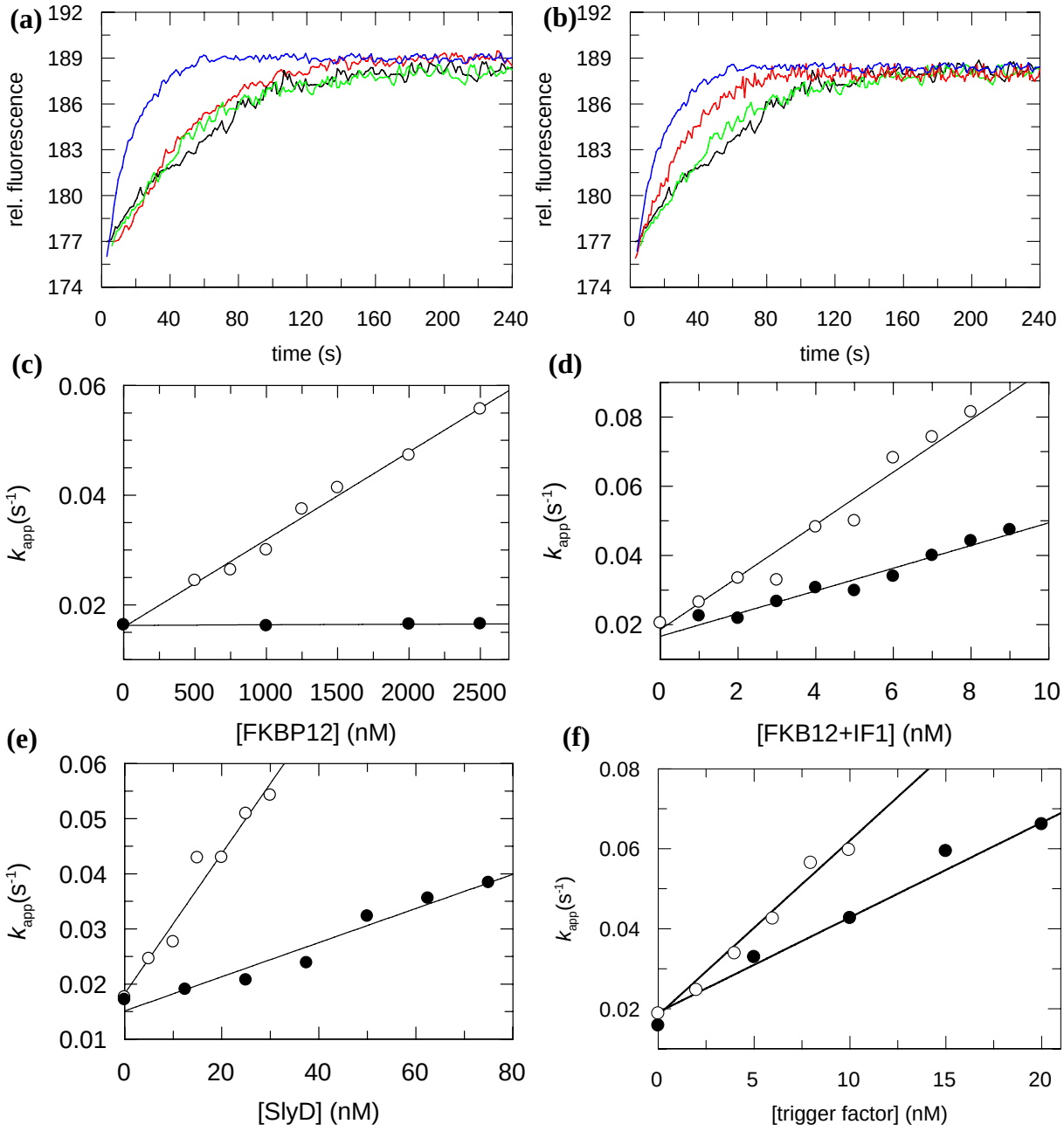


Figure 3. Refolding kinetics of 0.33 µM N2-160L and 0.33 µM N2-160E in the absence (black and green curves) and in the presence (blue and red curves, respectively) of (a) 2 µM FKBP12 and (b) 6 nM FKBP12-IF. (c-e) Catalytic efficiencies of (c) FKBP12, (d) FKBP12+IF, (e) SlyD* and (f) trigger factor towards the proline-limited slow refolding reaction of N2-160L (○) and N2-160E (●). The measured rate constants k_{app} of N2 folding are shown as a function of the prolyl isomerase concentration. The k_{cat}/K_M values derived from the slopes are given in Table 1. Refolding kinetics were measured at 15 °C in 0.1 M potassium phosphate, 0.2 M urea, pH 7.0 by the change in fluorescence at 340 nm, after excitation at 280 nm.

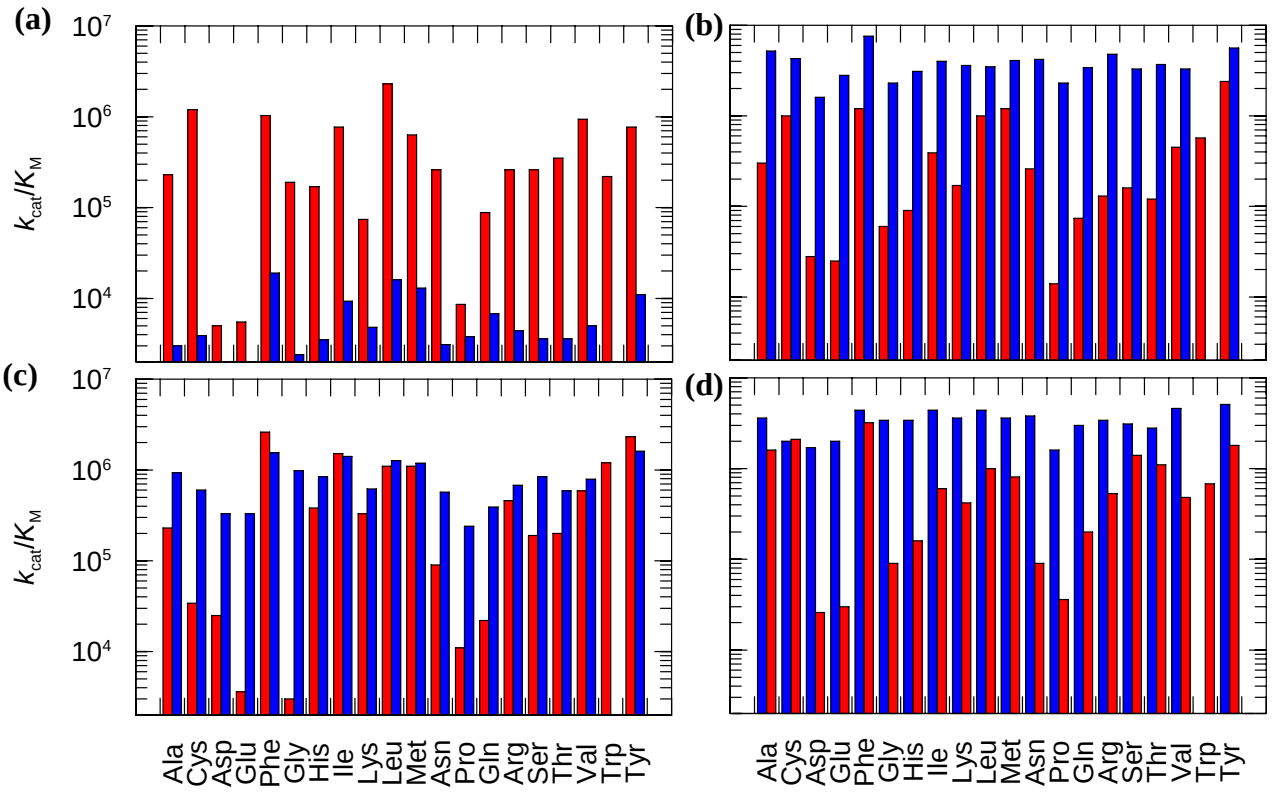


Figure 4. Catalytic effiencies k_{cat}/K_M of the various prolyl isomerases for the *cis/trans* isomerization of Abz-Ala-Xaa-Pro-Phe-pNa (blue bars) and fot the refolding of N2-D160Xaa (red bars). (a) hFKBP12, (b) FKBP12+IF, (c) SlyD*, (d) trigger factor. The measurements were performed as described in Figures 2 and 3. The catalytic efficiencies were derived from the dependence of the apparent rate constants on the enzyme concentration.

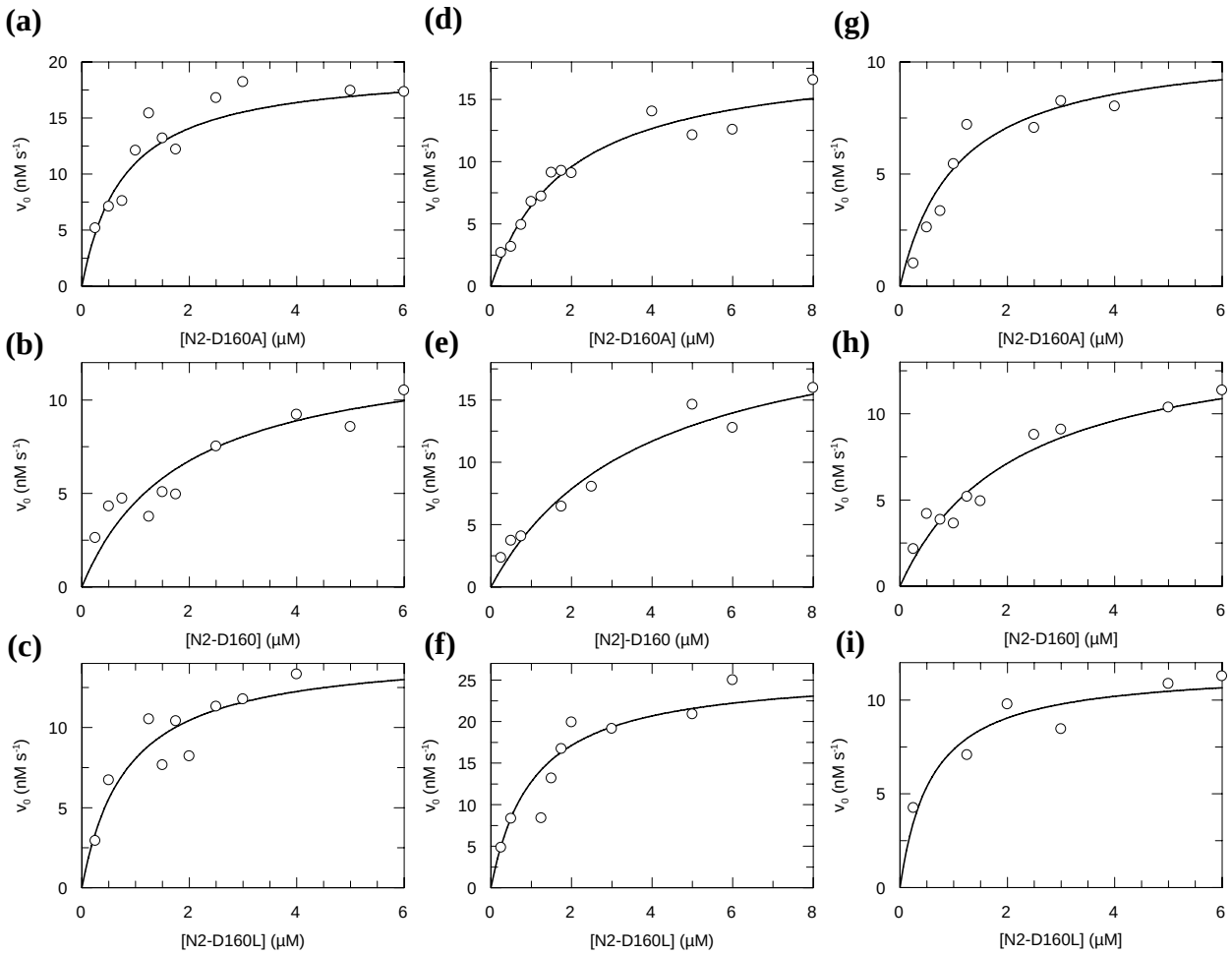


Figure 5. Michaelis-Menten kinetics of the refolding of (a, d, g) Ala160-N2, (b, e, h) Asp160-N2 and (c, f, i) glu160-N2 catalyzed by (a-c) FKBP12+IF, (d-f) SlyD* and (g-i) trigger factor. The initial velocities of catalyzed refolding at 15 °C are shown as a function of the concentration of N2. The concentrations were 5 nM for FKBP12+IF and trigger factor, and 10 nM for SlyD*, and the buffer consisted of 0.2 M urea, 0.1 M K phosphate (pH 7.0). The values for K_M and k_{cat} are presented in Table 3.The initial folding rates were determined and analyzed as described in Materials and Methods.

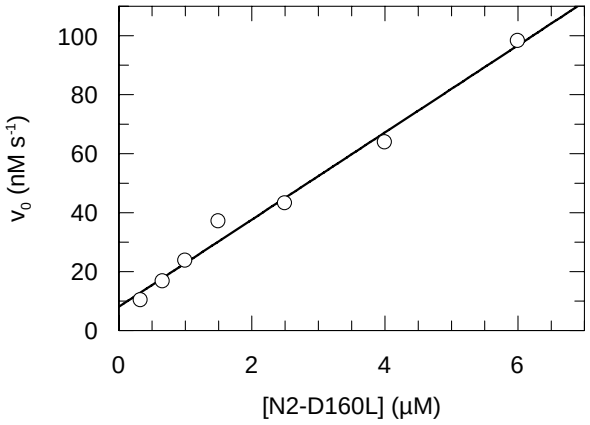


Figure 6. Initial velocity ν_0 of the catalyzed refolding of N2-D160L as a function of the concentration of N2-D160L in the presence of 1 µM hFKBP12. Refolding was measured at 15°C in of 0.2 M urea, 0.1 M K phosphate (pH 7.0). The initial folding rates were determined and analyzed as described in Materials and Methods.

Teilarbeit F:

Jakob, R.P., Zierer, B.K., Hofmann, S., Lorenz, S., Dobbek, H. and Schmid, F.X**.**

Protein stabilization by elimination of a cis-proline-containing loop and optimization of type I' β-turns., *Manuskript in Vorbereitung*

Running title: Turn optimizations in the phage gene-3-protein

Protein stabilization by elimination of a cis-proline-containing loop and optimization of type I' β-turns

Roman P. Jakob, Bettina K. Zierer, Stefanie Hofmann, Stefan H. Lorenz,
Holger Dobbek and Franz X. Schmid*

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

*Corresponding author
Franz X. Schmid
Biochemie
Universität Bayreuth
D-95440 Bayreuth, Germany
Tel: ++49 921 553660
Fax: ++49 921 553661
E-mail: fx.schmid@uni-bayreuth.de

Abbreviations used: G3P, gene-3-protein of phage fd; N2, N-terminal domain of G3P; N2', N2 containing the stabilizing mutation Q129H; $[urea]_M$, midpoint of a urea-induced unfolding transition; CD, circular dichroism; T_M, midpoint of a thermal unfolding transition; ΔH_D, van´t Hoff enthalpy of denaturation at T_M; ΔG_D, Gibbs free energy of denaturation; m, cooperativity value of a denaturant-induced equilibrium unfolding transition; A, λ, amplitude and apparent rate constant of a folding reaction; τ ($=\lambda^{-1}$), time constant of a folding reaction; k_{NU}, microscopic rate constant and m_{NU} ($= \delta \ln k_{NU}/\delta[urea]$), kinetic m-value of unfolding; k_{UN}, m_{UN}, microscopic rate constant and kinetic m-value of refolding.

Abstract

The N2 domain of the gene-3-protein (G3P) of phage fd shows an irregular loop of seven residues (157-163) that connects two consecutive β strands. The mobility of this loop is very high, and its conformation is so weak that it can maintain Pro161 at its tip in the *cis* form in only about 90 % of the molecules. Refolding of N2 is complex because the molecules with *cis* or *trans* isomers at Pro161 both fold to native-like forms, albeit with different rates. We reduced the folding complexity and strongly increased the conformational stability of N2 by using a consensus sequence design to shorten the seven-membered irregular loop around Pro161 to a four-membered type I' turn without a proline. An additional *in-vitro* selection for N2 variants with increased stability yielded sequences for this turn that strongly resembled those derived from the consensus design. Two other type I' turns of N2 could also be stabilized by consensus design. All three turns are critical for the refolding of N2. They are formed early during refolding and must be native-like to decrease the height of the folding transition state. Accordingly, refolding was strongly accelerated by the turn optimizations. The crystal structure of the variant with all three stabilized turns confirms the successful replacement of the 157-164 loop by a type I' turn. The other turns maintained their type I' conformation in the course of the sequence optimization.

Key words: protein design, folding mechanism, consensus design, folding transition state, protein stability, protein engineering, β-turn

Introduction

Turns and loops are elements of local secondary structure in proteins that allow the polypeptide chain to revert its direction within a few residues[1]. It has been discussed for a long time whether turns are active elements for protein folding as nuclei for early structure formation or passive in the sense that they tolerate the early formation of ordered secondary structure in their vicinity, such as in β-hairpins. In either case, they are expected to affect the stability and the folding kinetics of proteins [2; 3; 4; 5;6; 7; 8; 9] and the role of β-hairpins in nucleating protein folding has been investigated for several proteins [2; 10; 11; 12; 13; 14; 15; 16].

The N2 domain of the gene-3-protein (G3P) of phage fd is a small autonomously folding protein[17; 18]. At its top (in the orientation shown in Figure 1), it shows an irregular loop of seven residues (residues 157-163) that connects two consecutive β strands. This loop with a *cis* proline in its middle shows puzzling properties[19; 20]. The two β strands leading into and out of the loop are well ordered, but the loop itself shows a high mobility (high B factors) in the crystal structure, and its conformation is so weak that it can maintain Pro161 in the *cis* form in only about 90 % of the molecules. In fact, the connector peptides between Pro161 and the two β strands are folded when this Pro is *cis*, but unfolded when it is in the *trans* conformation. Despite this weak and heterogeneous structure in the folded state and the incomplete coupling between conformational folding and the *cis*/*trans* isomerization at Pro161, this loop was found to be important for the folding of the N2 domain. Several residues at the ends of the β strands and in the loop showed mutational Φ values close to one, which indicates that structure formation in this loop is in fact essential for reaching the folding transition state [20].

Here, we asked whether the folding complexity in the 157-163 region of N2 could be simplified and, possibly, the stability improved by shortening this seven-membered loop around Pro161 to a five-membered G1 bulge or to a four-membered type I' turn, both without proline. Our designs, based on data from the statistical analysis of turns and loops led to several strongly stabilized variants. The best design contained a type I' turn with a sequence that almost matches the consensus sequence for this turn type.

The N2 domain contains two additional type I` turns (marked in red in Figure 1), and we examined whether they could as well be optimized by a design towards the consensus sequence. To elucidate the structural consequences of these turn optimizations, we solved the crystal structure of a variant with optimal sequences for all three type I' turns. To reveal the importance of these turns for the folding mechanism, we measured the unfolding and refolding kinetics as a function of the urea concentration. The results indicate that the turn optimizations strongly increased the stability of the N2 domain and, at the same time, greatly

accelerated refolding, confirming that correct turn formation is essential for early steps of folding.

Results and Discussion

Stability and folding of the N2 domain with designed turn 3

The N2 domain comprises the residues 102 to 205 of the phage G3P (Figure 1a). As in our previous work [19], we used the stabilized Q129H variant of N2 as the reference or pseudo wild-type protein and denote it as N2'. The N2 domain is an independently folding unit, and the analysis of the folding mechanism revealed a native state *cis*/*trans* equilibrium at Pro161, which is located at the tip of an extended β-loop, which we call "loop 3" (residues 157-163) (Figure 1b). This exposed loop is not engaged in interactions with other parts of the protein, it shows high *B*-values in crystal structures [21; 22] and a high flexibility in NMR measurements (data not shown).

Here we asked, whether the N2 domain can be stabilized by shortening the irregular loop 3. Whether two β-strands are best connected by five residues (as in G1 bulges) or by four residues (as in type I′ turns), is determined largely by the angle formed between the ends of the two strands that lead into and out of the turn. In loop 3 these two strands show an angle of about 40 °, which would be compatible with both a G1 bulge and a type I' turn. We therefore designed two series of variants, in which the seven residues of the 157-163 loop were shortened to either four residues to generate a type I' turn or five residues to generate of G1 bulge (Figure 1b). The sequences were designed following the comprehensive consensus analysis by Thornton and coworkers [23; 24].

For the three central residues of a G1 bulge, the first position is in 80 % of all cases occupied by a Gly residue, at position 2 charged groups are favoured, and at position 3 Asn and Asp occur frequently, probably because they can engage in hydrogen bonds with residues surrounding the bulge region. Following these preferences, we shortened loop 3 to a G1 bulge by replacing the residue stretch 158-162 by the sequences GDN, GQN or GQD. In addition, we also inserted the sequence ADG, which, in an extensive analysis, could be characterized as the optimal G1 bulge for the first loop of the WW domain of Pin1[25].

All four G1 bulge variants were found to be folded and soluble. The variant GDN resembled the wild-type protein in its stability, and the variants GQN and GQD were slightly destabilized (by about 2 kJ mol^{-1}, Figure 2a). Interestingly, the sequence ADG from the Pin1 WW domain optimization also strongly stabilized the N2 domain and increased the midpoint of its thermal unfolding transition (T_M) by 5 °C (Figure 2a). Similar results were obtained when the N2 variants were unfolded by urea (Table 1). The midpoints of the transitions

$[urea]_M$ followed the same rank order as the T_M values, again with ADG being the most strongly stabilizing replacement for loop 3 ($\Delta\Delta G_D$ = 5.1 kJ mol^{-1}).

Subsequently, we replaced the native seven-residues loop sequence (QGTDPVK) by type I' turn sequences (QXGK). Four-residue type I' β-turns are common in protein *β*-hairpins, because the geometry of this turn is well compatible with the right-handed twist of many *β* sheets. The positions of the turn are often labelled as (–B1) – (L1) – (L2) – (+B1) [24], where L1 and L2 are the central turn residues and –B1 and +B1 connect to the β-strands preceding and following the turn, respectively. The backbone Φ and Ψ angles for L1 lie in the α_L region of the Ramachandran plot, and therefore Asn, Asp, and Gly are found at L1 in 60 % of all cases. For position L2, the backbone conformation is even more restricted. The required backbone angles (Φ ~ 90° and Ψ ~ 0°) are disallowed for all residues other than Gly, and, as a consequence, Gly is by far the most abundant residue at position L2 [26]. Therefore, all our designed type I' variants contained a Gly at the L2 position, for the L1 position, we used Gly or Asn. Both N2 variants with the wild-type residues at the positions –B1 and +B1 (QNGK and QGGK) were more stable than the most stable G1 bulge variant and their T_M values were 7.5 °C (Gly) and 8.6 °C (Asn) higher than the T_M value of the wild-type protein (Figure 2). This shows that loop shortening leads to a strong stabilization of the N2 domain. The strongest effect was observed when the seven-residue loop was shortened to a four-residue type I' turn with a Asn-Gly sequence at the central L1 and L2 positions (Table 1), as found in the statistical analysis of amino acid preferences for these positions.

The statistical analysis of type I' turns revealed preferred residues for the flanking positions -B1 and +B1 as well, although the preferences are not as clear-cut as those for the central positions L1 and L2 [27]. At position +B1, Lys as in turn 3 of the N2 domain is in fact the preferred residue and, consequently, we retained this residue. At position -B1 Val is preferred and found in 20 % of all type I' turns. For other non-polar residues, the preferences are less pronounced. At the –B1 position of turn 3, we therefore introduced a Val and, alternatively, a Tyr residue. Val (in the VNGK variant) in fact led to a shift of the T_M from 46.4 °C for the QNGK variant to 48.5 °C, equivalent to an increase in stability by 9.0 kJ mol^{-1}. The mutation to Tyr at the –B1 position left the stability of N2 unchanged (Table 1). All variants were also unfolded by urea (Figure 2b), and the midpoints of the transitions $[urea]_M$ follow the same rank order as the corresponding T_M values (Table 1). VNGK is the most stable variant, with a $[urea]_M$ of 4.6 M and an increase in ΔG_D by 11.4 kJ mol^{-1}.

Together, these design experiments show that the stability of the N2 domain could be strongly increased by shortening a seven-membered loop to a four-residue type I' β-turn. The most stable variant was obtained by inserting the consensus sequence VNGK, as obtained from the statistical analysis. We assume that this mutation towards the consensus worked so well because this turn is solvent exposed with few intra-molecular contacts.

To elucidate the origin of this stabilization, we measured the folding kinetics of the turn 3 variants. The folding of the wild type form (N2') is a complex process, because the folded form shows a *cis*/*trans* equilibrium at Pro161, which is located in β-loop 3. The refolding kinetics are therefore composed of two phases that reflect the folding of the species with a *cis* and a *trans* Pro161, respectively, with the *cis* form refolding 30-times faster than the *trans* form[19]. Native-like interactions of the two β strands that lead into loop 3 are important early in folding and must be established before the folding transition state is reached. These interactions are more easily formed and stabilized when Pro161 is in the *cis*-conformation, and therefore these molecules refold much faster. This suggests that turn 3 is in fact important for folding. The unfolding rates of the *cis* and the *trans* species are identical, which indicates that these two species are truly native forms and that loop 3 is native-like already in the folding transition state.

The kinetic experiments with the designed variants gave a clear result. All the transformations of loop 3 into a G1 bulge or a type I' turn left the unfolding kinetics of N2' unchanged (Figure 3a). The refolding kinetics, however, were strongly accelerated. The kinetics were monophasic, and a split into a fast and a slow phase was no longer observed, evidently because Pro161 is removed in these shortened turn 3 variants. (Figure 3b) All the changes in the equilibrium stability between the variants are reflected in corresponding changes in the rates of refolding.

These differences are most clearly seen when the measured rates of folding are plotted as a function of the denaturant concentration in the form of chevron plots (Figure 3c). All turn 3 variants show a common unfolding limb but differ in the refolding limb of the chevron. The increased stability of these proteins thus results exclusively from an increase in the rate of refolding, which indicates that most of the stabilizing interactions of the turn are established early in the folding process, before the passage through the transition state This is in line with the previous folding experiments with the wild-type protein. They showed, that mutations in the loop 3 and in the adjacent β strands predominantly affect refolding, resulting in high Φ – values for these residues. Refolding of the best variant (VNG) is about 100-fold faster than the $U_t \rightarrow N_t$ reaction of the wild-type protein, which accounts for about 90 % of the refolding wild-type molecules.

Stability and folding of the N2' domain with designed turn 1 and turn 2 sequences

The results described above indicate that the N2 domain could be strongly stabilized by shortening the 157-163 loop to a type I' turn. Wild-type N2' contains two type I' turns. The first one (turn 1, residues136-39) is buried, the other one (turn 2, residues 144-147) is mostly surface accessible (Figure 1). In their sequences, turn 1 (FQNN) and turn 2 (RQGA) differ strongly from the consensus sequence VNGK. Here, we asked whether N2' could be

stabilized by optimizing these turns towards the consensus sequence as well. Therefore, we produced six variants of turn 1 and seven variants of turn 2 with single and multiple amino acid replacements.

In turn 1, N2' deviates from the consensus sequence at all four positions. Thus, we replaced every amino acid individually by the statistically most preferred one, leading to the F136V, Q137N, N138G, and N139K variants. The variant N138G had been found before in an *in vitro* evolution experiment with the aim of generating a stable variant of the phage fd gene-3-protein without disulfide bonds. In this work, protein libraries had been employed to select stabilizing amino acid replacements that could compensate for the loss of thermodynamic stability as a result of the disulfide replacements [22; 28]. The substitution N138G at position L2 increased the T_M of the isolated N2' by 3.1 degrees and ΔG_D by 3.2 kJ mol^{-1}, evidently because Gly is the only residue that is compatible with the backbone angles required at this position. The other three single mutations (F136V, Q137N, N139K) towards the turn I' consensus sequence were destabilizing (Figure 4a), The F136V substitution alone decreased ΔG_D by 15.5 kJ mol^{-1}. The single substitutions Q137N at position L1 and N138G at position L2 were also introduced in combination, to examine whether they act in a synergistic manner. Indeed, the $\Delta\Delta G_D$ of this variant (FNGN) is 2 kJ mol^{-1} higher than the sum of the $\Delta\Delta G_D$ values of the single variants. Still, the FNGN variant is less stable than wild type protein.

Turn 1 is fully buried, and all side chains (beside N138) are involved in multiple van der Waals or hydrogen-bonding interactions. Evidently, the substitutions by the statistically most favoured residues could not compensate for the disruption of these interactions. Thus, optimization of buried β-turns remains challenging, because potential changes in side chain interactions must be accounted for.

Turn 2 (residues 144-147) of N2' is mostly surface accessible and shows the sequence RQGA. Gly at position L2 as in the statistically preferred sequence VNGK is thus present already. Again, the other three turn residues were replaced individually by the most preferred ones. The substitutions at the first two positions, R144V and Q145N, indeed stabilized N2' (Figure 4b), by 3.2 kJ mol^{-1} and 1.1 kJ mol^{-1}, respectively (Table 1). The mutation A147K was slightly destabilizing (by -1.2 kJ mol^{-1}). In folded N2', the side chain of residue 147 is in close contact to residues 117-122, and amino acids with large side chains, such as Lys, are probably not well accommodated.

In combination, the two substitutions R144V and Q145N (in the VNGA variant) stabilized N2' by 5.4 kJ mol^{-1}, which is 1.1 kJ mol^{-1} higher than the sum of the two individual substitutions. When the third substitution, N147K, was included as well (to reach the consensus sequence VNGK) again a destabilization by 2.2 kJ mol^{-1} was found. This indicates

that the destabilization of the wild-type protein by the single N147K substitution is not caused by a cross-turn repulsion with Arg144, which is no longer present in the VNGK variant.

Modelling with the swiss pdb viewer indicated that a Ser or Thr residue at the last (+B1) position of turn 2, might be donating a favourable hydrogen bond to the backbone amide of Arg144. The substitution A147T was, however, slightly destabilizing (-0.6 kJ mol^{-1}, Table 1). Finally, we examined whether a β-branched residue such as Val might be well suited for this position, because the last position of a type I' turn is already part of the subsequent β-strand. Indeed, the A147V replacement slightly stabilized N2 (by 0.8 kJ mol^{-1}, Table 1). Thus, the replacement of the wild-type sequence RQGA by VNGV stabilized turn of 2 N2' by 6.2 kJ mol^{-1} (Table 1). The stabilities of all the designed variants of turn 2 were also determined from urea-induced transitions (Table 1). The $[urea]_M$ values correlated very well with the corresponding T_m values (Figure 4c), which suggests that the mutations do not change the two-state character and the cooperativity of unfolding. The most stable sequence for turn 2 (VNGV) differs from the consensus sequence only at the +B1 position. It is partially buried in the wild-type protein, which probably explains why a Val is preferred over Lys. These results confirm those obtained for turn3, and suggest that the consensus sequence or sequences very close to the consensus are indeed very well suited to optimize surface-exposed type I' turns.

For several of the optimized turn 1 and turn 2 variants we also measured the folding kinetics (Figure 5). These proteins showed separate refolding reactions for the *cis* and *trans* forms of Pro161 as the wild-type protein, because Pro161 is present in all the variants. The most stable turn 1 variant (FQGN) showed unchanged unfolding kinetics (Figure 5a), but fivefold increases in the rates of the two refolding reactions. An identical behaviour was found for the variant VNGA of turn 2. Here, also the unfolding kinetics remained unchanged, and the refolding reactions were accelerated (Figure 5c).

The subsequent introduction of the A147V mutation into turn 2, to create the VNGV variant, led to interesting changes in the folding kinetics. It increased the rates of refolding (about fourfold) as well as of unfolding (about threefold) (Figure 5b). Such overall increases in the dynamics of folding are usually attributed to a specific stabilization of the folding transition state relative to the unfolded and the folded state. The molecular origin for this effect remains unclear. Together with the results obtained for turn 3, this indicates that all three turns are important for the folding pathway of the N2 domain. They form early, and they must be native-like when the transition state of folding is crossed.

Combination of the most stable designed turn sequences

Finally, we combined the most stable individual turn sequences FQGN for turn 1, VNGV for turn 2, and VNGK for turn 3 in the single variant 3T-N2'. The T_M value of this protein was strongly increased from 37.8 °C to 54.9 °C (Figure 6a), but thermal unfolding was only

partially reversible, because the protein had to be heated to 80 °C to complete unfolding. The combinatorial variant was also strongly stabilized towards unfolding by urea, and the midpoint of this reversible unfolding transition increased from 2.7 to 6.5 M, equivalent to an increase in ΔG_D by 17.0 kJ mol^{-1} (Table 1). This is 3-4 kJ mol^{-1} less than the sum of the $\Delta\Delta G_D$ values for the individual turn variants. This difference originates from the reduced cooperativity of unfolding of the 3T-N2' variant (m = 5.1 kJ mol^{-1} M^{-1}) relative to the wild-type protein (m = 6.7 kJ mol^{-1} M^{-1}). Such a decrease in the measured cooperativity might originate from less perfect packing in the folded state or from residual structure in the unfolded form, which reduces the interaction of the protein backbone with the denaturant. Alternatively, the unfolding of this strongly stabilized variant might no longer obey the two-state approximation. Still, the combinatorial mutant shows a strongly enhanced stability towards both thermal and denaturant-induced unfolding.

The folding kinetics of the combinatorial variant were measured as well (Figure 6c) Unfolding and refolding were monoexponential reactions between 4 and 8.5 M urea, resulting in a chevron with a minimum at 6.5 M urea (Figure 6c), in good agreement with the midpoint of the equilibrium unfolding transition in Figure 6b. In the T3 variant, Pro161 is deleted due to turn 3 optimization, and therefore the refolding kinetics were monoexponential, unlike the refolding of the wild-type protein, which showed separate refolding kinetics for the species with the *cis* and *trans* forms of Pro161.

The strong stabilization of T3-N2' originates predominantly from an acceleration of the refolding reaction, as expected from the strong effects on refolding observed before for the species with individually optimized turns (Figures 3c and 5). This confirms that early structure formation in all three turns is important to reach the folding transition state of N2'. Unfolding is slightly accelerated as well, which reflects the effect of the A147V substitution in turn 2 (cf. Figure 5b).

The folding kinetics of variants that differ so strongly in stability as 3T-N2' and N2' are difficult to compare. The transition states might not be conserved, and the two-state approximation might not be adequate for the strongly stabilized variant.. Therefore, in a simple approach, we compared the rates of equilibration between the folded and unfolded form at the respective transition midpoints, where intermediates are not likely to be populated. This comparison shows that the N and U forms of the T3 variant with the stabilized turns equilibrates 40-times faster than the *cis* form of N2' and 230-times faster than the *trans* form. It indicates that under conditions where N and U are equally stable, the molecules with

optimized turns find the native state much more often, evidently because the optimized turns stabilize the folding transition state.

Below 4.5 M urea, the refolding limb of the chevron for T3-N2' becomes non-linear. At the same time, it loses amplitude and becomes too fast for reliable measurements after stopped-flow mixing. It is thus possible that, because of the strong turn stabilizations, the T3 variant forms a folding intermediate even at high urea concentration. We speculate that this intermediate is related to a much less stable intermediate that leads to the minor curvature in the chevron of the wild-type protein at low urea concentration (Figure 6c) [19]. Similar effects were found for a stabilized variant of ubiquitin [7; 29].

The structure of the Gen-3-Protein with combined most stable type I' turns

Attempts to crystallize the isolated 3T-N2' domain were unsuccessful. Therefore, we introduced the three most stable turn sequences into a N1-N2 fragment of the gene-3-protein that we had previously obtained from an *in-vitro* selection for improved stability [30]. This protein (3T-G3P*) could be crystallized, and its structure was solved at 2.2 Å resolution. Details of the crystallization and structure determination are summarized in Table 4. The crystals showed $P2_12_12_1$ symmetry with two molecules in the asymmetric unit. Their backbone atoms superimposed with a RMS of 0.5 Å. To interpret the consequences of the turn replacements, we used the structure of the wild-type protein (1G3P) [21] for comparison and aligned the N2 domains individually.

In the reference protein, the seven residues of loop 3 show high B-factors (Figure 7a), suggesting that they are flexible. The replacement by the optimized four-residue turn I' β turn strongly reduced this flexibility (Figure 7a). These results were confirmed by NMR measurements with the isolated N2 domain. Here, residues 157-163 were found to have a high flexibility on the ps-ns timescale in the wild type protein, which was strongly decreased in the variants with stabilizing turn I' sequences (Ulrich Weininger and Roman Jakob, data not shown).

The shortening of loop 3 from seven to four residues in fact led to a type I' turn, as indicated by the backbone dihedral angles in this region (Figure 7b and Table S1). This confirms our design and explains why the sequences close to the consensus sequence for type I' turns provided the strongest stabilization of the N2' domain. Neither of the turn residues is involved in additional side chain interactions to other parts of the molecule, which shows that the protein is stabilized only by improving the local interactions in the turns.

The turns 1 and 2 show low B factors in the reference protein already, and they are unchanged in the stabilized variant (Figure 7a). The backbone conformation of turn 2 is also unchanged in the 3T variant (Figure 7c), and the substituted amino acids are not engaged in

side chain interactions. The valine residues at the positions 144 and 147 optimize the ends of the β strands that lead into turn 2 and appear to reflect the intrinsic β-sheet forming propensity of the residues at the -B1 and +B1 positions at the end of the β-strand. The comparison of the structures of turn1 indicates that the N138G substitution (Figure 7d), which is present in the two stabilized variants, leads to a change in the backbone angles at position 138 towards the Φ/Ψ combination that is optimal for this position in a type I' turn (Table S1).

Generally, we could not detect structural rearrangements outside the re-designed turns. This suggests that, as expected, the turn optimizations stabilized the protein by improving the local interactions in the turns. The rather strict requirement for peculiar backbone angles at the four turn positions determines the often observed preference for residues with correspondingly high propensities for these angles. This is particularly obvious for Gly at position L2. These results are consistent with earlier substitutions in turns of α-spectrin SH3 [31], fibroblast growth factor-1 [32], lysozyme [33], and ubiquitin [6].

Optimal replacement of loop 3 by *in-vitro* selection

To examine whether the type I' turn consensus sequence VNGK is in fact the optimal sequence for turn 3, we employed the *in-vitro* selection method *Proside* (*p*rotein *s*tability *i*ncreased by *d*irected *e*volution) [34; 35] to find sequences that maximally stabilize turn 3 of the N2' domain.

In *Proside*, 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.[34; 35] The *Proside* strategy can also be used to stabilize G3P itself, in the absence of a guest protein.[17; 28]

Phages libraries were constructed in which the four residues of turn 3 were randomized by saturation mutagenesis. These libraries were then used to search for stabilized variants by *Proside*. During the selection, the temperature was increased in 15 selection rounds in steps from 48 °C to 59 °C, and the chymotrypsin concentration from 0.25 to 2.5 μM. 16 single clones from two independent libraries were sequenced. Seven of the 16 clones showed the sequence INGR, eight showed IDGR. These sequences agree well with the statistical analysis of type I' type turn sequences [7; 24]. Both contain a hydrophic β branched Ile residue at the first, a Gly at the third and a positively charged residue (Arg) at the fourth position. The selected variants only differ at the second position. Here, Asn and Asp were found, both of which can readily adopt the Φ and Ψ angles typical of this position. N2' variants with these

turn 3 sequences were expressed and their stabilities measured. A single clone with the sequence IAGK was found, which, due to its single occurrence, was not characterized further. Several spontaneously occurring *second site* mutations in G3P were found as well in the selections (at the positions 13, 29, 50 101, 129, 209). They had been found before in other selections for stabilized variants of G3P [17; 28].

For these selections, the wild type gene-3-protein was used, with Q129 instead of H129, as in N2', which served as the reference protein in the design experiments. The T_M value of wild-type N2 is at 33.7 °C. It is increased to virtually identical values of 43.6 °C for the INGR variant and 43.3 °C for the IDGR variant (Figure 8a), which corresponds to increases in ΔG_D by 9.9 and 9.6 kJ mol^{-1}, respectively. Similar results were obtained from urea-induced transitions. The $[urea]_M$ increased by 1.9 M (INGR) and 1.8 M (IDGR) (Figure 8b). Asp and Asn seem to be almost equally well suited for the second position of turn 3, as expected from the sequence analysis of type I' turns. The strong stabilization by about 10 kJ mol^{-1}) is only slightly higher than the value found for the designed turn3-variant VNGK, which indicates that both methods are suitable to obtain stabilized turn variants. It remains open, why the variant with the designed VNGK sequence was not retrieved in the selections, although its stability is only slightly lower stability. Apparently, 10 kJ mol^{-1} is the maximal gain in stability that can be achieved for this turn. It is due not only to the optimization of the turn, but also to the restriction of flexibility as a consequence of shortening the loop from seven to four residues.

Consequences of the turn optimizations in the N2 domain for the biological function during phage infection

The N2 domain of the G3P initiates the infection of *Escherichia coli* by filamentous phages by binding to the tip of an F pilus[36; 37]. After pilus retraction, N1 interacts with TolA, the ultimate phage receptor at the surface of *E. coli*[38; 39; 40]. To examine the consequences of the turn optimizations in N2 for the infectivity of the phage, we introduced the mutations for the most stable turn sequences individually and in combination into the phage genome and used the corresponding phage variants to infect aliquots of the same culture of *E. coli* (Figure 9). All variants were active. The optimization of turn 1 left the phage infectivity virtually unchanged, the optimization of turn 3 led to a weak (3-fold) decrease in infectivity. This indicated that the *cis*/*trans* equilibrium at Pro161, which is present in the wild type form of N2 but absent in the optimized turn 3, has no obvious relevance for the biological function.

In contrast, the phage with the optimized turn 2 sequence VNGV was 200-fold less infectious than the wild-type phage. This turn is important for phage function. Arg144 at the first position of turn 2 is part of the binding site for the F-pilus [36], and the mutation to a Val probably reduces the affinity for the pilus and thus the infectivity of the phage. Obviously,

this turn is optimized for function but not for stability. It has repeatedly been observed that protein stabilizations by substitutions in functionally important regions interfere with function. In the WW domain of Pin1, engineering of a functionally important loop stabilized the protein but eliminated binding to phosphorylated substrates[25].

The phage with all three stabilized turns showed a 2000-fold decrease in activity, 10-fold stronger than the effect caused by the mutations in turn 2. Turns 1 and 3 are not involved in binding to the pilus, and, individually, the optimizations of these turns did not reduce the infectivity significantly (Figure 9). Possibly, an excessively high conformational stability of N2 interferes with later steps in the infection process. It was suggested that partial unfolding of the N2 domain during the infection at 37 °C might increase the mobility of the N1 domain at the tip of G3P, and thus the surface of the *E. coli* cell could be scanned more efficiently for the TolA protein, which is the ultimate receptor for the phage [41].

Conclusions

The irregular seven-residue loop 3 in the N2 domain of the phage gene-3-protein is a peculiar element of structure. Together with the two strands that lead into and out of this loop, it forms a critical nucleus in the folding transition state. At the same time, it is unfavourable for the stability because about 10 kJ/mol of conformational folding energy must be transferred into this loop to shift the *cis*/*trans* equilibrium at Pro161 at the tip of this loop. In fact, the deletion of this proline and the shortening of the loop to a four-residue type I' turn stabilized the protein by almost 10 kJ/mol and strongly increased the rate of folding. The Pro161 loop is important for reaching the folding transition state not because it contributes critical structure to the transition state, but because it prevents the consolidation of its structure as long as it is not in its native conformation. In fact, before reaching the transitions state, 10 kJ/mol of conformational energy must be available already and transferred to the Pro161 loop to shift the *cis*/*trans* equilibrium.

An essential role of turns for the folding has been demonstrated for several small proteins [7; 29; 31; 32; 42; 43; 44; 45], and there is no doubt that turn and hairpin formation are essential early processes in protein folding. It remains open, however, whether turns play an "active" role as early initiation sites for local folding or, alternatively, a "passive" role such that they do not allow folding to proceed as long as they have not formed a native-like conformation. Our results with the N2 domain rather point to such a passive role.

As in preceding experiments [6; 8; 25; 32; 46; 47], we find that consensus sequence optimization of turns, possibly in combination with *in-vitro* selection, provides excellent means for increasing the stability of a protein. In the case of the N2 domain the optimization of three loops and turns led to a stabilization by 14 kJ/mol.

Materials and Methods

Expression and purification of variants of N2'

The isolated N2 domain (residues 102-205 of the gene-3-protein of phage fd, extended by $(His)_6$) with the stabilizing mutation Q129H was used as the reference (pseudo wild-type) protein for the rational design. All mutated variants of N2' were expressed and purified as described[19; 48] Site-directed mutagenesis of N2' was performed by Quik-Change (Stratagene).

Equilibrium unfolding transitions

Samples of the N2' variants (1.0 µM) were incubated for 1 h at 15 °C in 100 mM K phosphate, pH 7.0 and varying concentrations of urea. The fluorescence of the samples was measured in 1-cm cuvettes at 340 nm (10 nm band width) after excitation at 280 nm (5 nm band width) in a Hitachi F4010 fluorescence spectrometer. The experimental data were analyzed according to a two-state model by assuming thatΔG_D as well as the fluorescence emissions of the folded and the unfolded form depend linearly on the urea concentration. A nonlinear least-squares fit with proportional weighting of the experimental data was used to obtain ΔG_D as a function of the urea concentration [48].

The heat-induced unfolding transitions were measured in a Jasco J-600A spectropolarimeter equipped with a PTC 348 WI peltier element at a protein concentration of 4 µM in 100 mM K phosphate, pH 7.0 or pH 8.0 at a heating rate of 1 °C/min. The transitions were monitored by the increase of the CD signal at 222 nm with 1 nm band width and 10 mm path length. The experimental data were analyzed on the basis of the two-state approximation [49], with a heat capacity change ΔC_p of 6000 J mol^{-1} K^{-1} (calculated as described in ref. [50]).

Kinetic experiments

All urea-induced unfolding and refolding experiments were performed in 100 mM K phosphate, pH 7.0, at 15 °C at a final protein concentration of 1.0 µM. The folding reactions were measured by using a DX.17MV stopped-flow spectrometer from Applied Photophysics (Leatherhead, UK). The native or the unfolded protein was diluted 11-fold with urea solutions of varying concentrations. The kinetics were followed by the change in fluorescence above 320 nm after excitation at 280 nm (10 nm bandwidth) in an observation cell with 2 mm path length. A 0.5-cm cell with acetone was placed between the observation chamber and the photomultiplier to absorb scattered light from the excitation beam. The kinetics were measured at least eight times under identical conditions and averaged to improve the signal-to-noise ratio.

In the analysis of the unfolding and refolding kinetics of the turn1- and turn2-variants of N2' variants we assumed that the folding kinetics of the *cis* and the *trans* forms are kinetically isolated by the slow $U_t \leftrightarrow U_c$ and $N_t \leftrightarrow N_c$ isomerizations (Figure 1c) and that the logarithms of the microscopic rate constants of unfolding and refolding depend linearly on the urea concentration.

Protein Crystallization and Structure Determination of the gene-3-protein with the combined

For crystallization a gene-3-protein variant[22] with the combined most stable turns was used. The protein was crystallized by vapor diffusion using the hanging drop setup at 16 °C. The reservoir solution (500 µl) contained 15-20 % PEG 3350, 0.05 M $CaCl_2$, 0.2 M NH_4Cl in 0.1 M Tris buffer, pH 8.5. Two drops were set up per reservoir by mixing 2 µl of protein solution (20 and 30 mg/ml in 20 mM Tris-HCl, pH 8.0, respectively) with 2 µl of reservoir solution. The crystals were very sensitive, and therefore cross-linked with 0.025% glutaraldehyde for 15 min. Crystals were then shock frozen after addition of 20 % (v/v) PEG400 as cryoprotectant. Crystals were measured at the PSF beamline ID14-1 of Freie Universität Berlin at synchrotron BESSY, Berlin. The dataset was processed and scaled using XDS [51]. The crystals belong to space group $P2_12_12_1$ with cell dimensions a = 56.5 Å, b = 87.0 Å, c = 96.3 Å, and contain two molecules per asymmetric unit. The structure was determined using Patterson search techniques. The search model was the chain of molecule A of the crystal structure of G3P from the filamentous phage fd (PDB accession code 3DGS) and the best replacement solution was obtained with the program AMoRe[52]. Model building and structure refinement was done with COOT [53]and refmac5 [54].

Data Bank accession codes

The coordinates and structure factor amplitudes are deposited in the *protein data bank* with ID code xx.

Creation of the phage libraries

For the substitution of the seven amino acid loop3 in G3P and selection of stabilized G3P variants with a four amino acid turn we used a phage derived from fCKCBS,[55] containing wild type G3P* without a guest protein. Libraries with the positions of one turn randomized were created by primers that exclude the codons of Cys at the positions of the disulfide. For the randomization of the loop3 the primers were 5′-ACT GTT TAT ACG GGC ACT GTT ACT NNN NNN NNN NNN ACT TAT TAC CAG TAC ACT CCT -3′ and 5′- AGC CGC CGC CAG CGT TAA CC -3′. The fragments obtained from the PCR experiments with the

above-mentioned primers were annealed to a single-stranded form of the phage DNA and used to transform E. coli XL1Blue (Stratagene, La Jolla, USA).

Proside selections

Phages were isolated from *E. coli* TG1 culture medium by polyethylene glycol-precipitation and stored in phosphate-buffered saline containing 0.01% (w/v) NaN_3. The *Proside* selections were performed with about 10^9 phage in 50 µl of 100 mM K phosphate, 100 µM $CaCl_2$ (pH 8.0). After incubation of the phage solution for 2 min at the respective temperature, the proteolysis was performed with 2.5 µM chymotrypsin for 15 min 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* TG1 with $A_{600}\approx1$. Phage were propagated at 37 °C for 8 h. To determine the relative infectivities of the phage of every selection cycle and to measure the proteolytic stabilities of the selected phage variants, two 25 µl samples with about 10^9 phage each in 100 mM potassium phosphate, 100 µM $CaCl_2$ (pH 8.0), one in the absence and one in the presence of 2.5 µM chymotrypsin, were incubated for 15 min at the respective temperature. Aliquots (5 µl) of these samples were added to 495 µl of a culture of *E. coli* TG1 ($A_{600}\approx1$) and incubated at 37 °C for 30 min 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.

Determination of phage infectivity

Phage were isolated from *E. coli* TG1 culture by polyethylene glycol-precipitation and filtered (0.22 µm pore size) to eliminate *E. coli* cells from the phage solution. For the determination of the relative phage concentration, the single-stranded phage DNA was isolated and the DNA content compared. Solutions with an equal phage concentration were prepared. To determine the infectivities of the different phage variants relative to wild-type phage, 5 µl aliquots of the prepared phage solutions were added to 495 µl of *E. coli* TG1 culture ($A_{600}\approx1$) and incubated at 37 °C for 30 min for infection. The number of infected cells and thus the number of infectious phage was determined by plating serial dilutions on dYT^{cam} agar.

Acknowledgements

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

References

1. Rose, G. D., Young, W. B. & Gierasch, L. M. (1983). Interior turns in globular proteins. *Nature* **304**, 654-7.
2. Montelione, G. T., Winkler, M. E., Burton, L. E., Rinderknecht, E., Sporn, M. B. & Wagner, G. (1989). Sequence-specific 1H-NMR assignments and identification of two small antiparallel beta-sheets in the solution structure of recombinant human transforming growth factor alpha. *Proc Natl Acad Sci U S A* **86**, 1519-23.
3. Dyson, H. J., Rance, M., Houghten, R. A., Lerner, R. A. & Wright, P. E. (1988). Folding of immunogenic peptide fragments of proteins in water solution. I. Sequence requirements for the formation of a reverse turn. *J Mol Biol* **201**, 161-200.
4. Klimov, D. K. & Thirumalai, D. (2002). Stiffness of the distal loop restricts the structural heterogeneity of the transition state ensemble in SH3 domains. *J Mol Biol* **317**, 721-37.
5. Nauli, S., Kuhlman, B. & Baker, D. (2001). Computer-based redesign of a protein folding pathway. *Nat Struct Biol* **8**, 602-5.
6. Simpson, E. R., Meldrum, J. K., Bofill, R., Crespo, M. D., Holmes, E. & Searle, M. S. (2005). Engineering enhanced protein stability through beta-turn optimization: insights for the design of stable peptide beta-hairpin systems. *Angew Chem Int Ed Engl* **44**, 4939-44.
7. Simpson, E. R., Meldrum, J. K. & Searle, M. S. (2006). Engineering diverse changes in beta-turn propensities in the N-terminal beta-hairpin of ubiquitin reveals significant effects on stability and kinetics but a robust folding transition state. *Biochemistry* **45**, 4220-30.
8. Trevino, S. R., Schaefer, S., Scholtz, J. M. & Pace, C. N. (2007). Increasing protein conformational stability by optimizing beta-turn sequence. *J Mol Biol* **373**, 211-8.
9. Marcelino, A. M. & Gierasch, L. M. (2008). Roles of beta-turns in protein folding: from peptide models to protein engineering. *Biopolymers* **89**, 380-91.
10. Kim, J., Brych, S. R., Lee, J., Logan, T. M. & Blaber, M. (2003). Identification of a key structural element for protein folding within beta-hairpin turns. *J Mol Biol* **328**, 951-61.
11. Krantz, B. A., Dothager, R. S. & Sosnick, T. R. (2004). Discerning the structure and energy of multiple transition states in protein folding using psi-analysis. *J Mol Biol* **337**, 463-75.
12. Platt, G. W., Simpson, S. A., Layfield, R. & Searle, M. S. (2003). Stability and folding kinetics of a ubiquitin mutant with a strong propensity for nonnative beta-hairpin conformation in the unfolded state. *Biochemistry* **42**, 13762-71.
13. Rotondi, K. S. & Gierasch, L. M. (2003). Local sequence information in cellular retinoic acid-binding protein I: specific residue roles in beta-turns. *Biopolymers* **71**, 638-51.
14. Rotondi, K. S., Rotondi, L. F. & Gierasch, L. M. (2003). Native structural propensity in cellular retinoic acid-binding protein I 64-88: the role of locally encoded structure in the folding of a beta-barrel protein. *Biophys Chem* **100**, 421-36.
15. Viguera, A. R. & Serrano, L. (2001). Bergerac-SH3: "frustation" induced by stabilizing the folding nucleus. *J Mol Biol* **311**, 357-71.
16. Viguera, A. R. & Serrano, L. (2003). Hydrogen-exchange stability analysis of Bergerac-Src homology 3 variants allows the characterization of a folding intermediate in equilibrium. *Proc Natl Acad Sci U S A* **100**, 5730-5.
17. 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.

18. Martin, A. & Schmid, F. X. (2003). The folding mechanism of a two-domain protein: folding kinetics and domain docking of the gene-3-protein of phage fd. *J. Mol. Biol.* **329**, 599-610.
19. Jakob, R. P. & Schmid, F. X. (2008). Energetic coupling between native-state prolyl isomerization and conformational protein folding. *J Mol Biol* **377**, 1560-75.
20. Jakob, R. P. & Schmid, F. X. (2009). Molecular determinants of a native-state prolyl isomerization. *J Mol Biol* **387**, 1017-31.
21. 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.
22. Kather, I., Jakob, R., Dobbek, H. & Schmid, F. X. (2008). Changing the determinants of protein stability from covalent to non-covalent interactions by in vitro evolution: a structural and energetic analysis. *J Mol Biol* **381**, 1040-54.
23. Chan, A. W., Hutchinson, E. G., Harris, D. & Thornton, J. M. (1993). Identification, classification, and analysis of beta-bulges in proteins. *Protein Sci* **2**, 1574-90.
24. Hutchinson, E. G. & Thornton, J. M. (1994). A revised set of potentials for beta-turn formation in proteins. *Protein Sci* **3**, 2207-16.
25. Jager, M., Zhang, Y., Bieschke, J., Nguyen, H., Dendle, M., Bowman, M. E., Noel, J. P., Gruebele, M. & Kelly, J. W. (2006). Structure-function-folding relationship in a WW domain. *Proc Natl Acad Sci U S A* **103**, 10648-53.
26. Ramirez-Alvarado, M., Blanco, F. J., Niemann, H. & Serrano, L. (1997). Role of beta-turn residues in beta-hairpin formation and stability in designed peptides. *J Mol Biol* **273**, 898-912.
27. Griffiths-Jones, S. R., Maynard, A. J. & Searle, M. S. (1999). Dissecting the stability of a beta-hairpin peptide that folds in water: NMR and molecular dynamics analysis of the beta-turn and beta-strand contributions to folding. *J Mol Biol* **292**, 1051-69.
28. Kather, I., Bippes, C. A. & Schmid, F. X. (2005). A stable disulfide-free gene-3-protein of phage fd generated by in vitro evolution. *J Mol Biol* **354**, 666-78.
29. Rea, A. M., Simpson, E. R., Meldrum, J. K., Williams, H. E. & Searle, M. S. (2008). Aromatic residues engineered into the beta-turn nucleation site of ubiquitin lead to a complex folding landscape, non-native side-chain interactions, and kinetic traps. *Biochemistry* **47**, 12910-22.
30. Kather, I., Jakob, R. P., Dobbek, H. & Schmid, F. X. (2008). Increased folding stability of TEM-1 beta-lactamase by in vitro selection. *J Mol Biol* **383**, 238-51.
31. Vega, M. C., Martinez, J. C. & Serrano, L. (2000). Thermodynamic and structural characterization of Asn and Ala residues in the disallowed II' region of the Ramachandran plot. *Protein Sci* **9**, 2322-8.
32. Lee, J., Dubey, V. K., Longo, L. M. & Blaber, M. (2008). A logical OR redundancy within the Asx-Pro-Asx-Gly type I beta-turn motif. *J Mol Biol* **377**, 1251-64.
33. Takano, K., Yamagata, Y. & Yutani, K. (2000). Role of amino acid residues at turns in the conformational stability and folding of human lysozyme. *Biochemistry* **39**, 8655-65.
34. Martin, A., Schmid, F. X. & Sieber, V. (2003). Proside: a phage based method for selecting thermostable proteins. *Methods in Molecular Biology* **230**, 57-70.
35. Sieber, V., Pluckthun, A. & Schmid, F. X. (1998). Selecting proteins with improved stability by a phage-based method. *Nat Biotechnol* **16**, 955-60.
36. 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.

37. 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**, 143-9.
38. Click, E. M. & Webster, R. E. (1997). Filamentous phage infection: required interactions with the TolA protein. *J Bacteriol* **179**, 6464-71.
39. 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.
40. Riechmann, L. & Holliger, P. (1997). The C-terminal domain of TolA is the coreceptor for filamentous phage infection of E. coli. *Cell* **90**, 351-60.
41. Eckert, B. & Schmid, F. X. (2007). A conformational unfolding reaction activates phage fd for the infection of Escherichia coli. *J Mol Biol* **373**, 452-61.
42. Fowler, S. B. & Clarke, J. (2001). Mapping the folding pathway of an immunoglobulin domain: structural detail from Phi value analysis and movement of the transition state. *Structure* **9**, 355-66.
43. Garcia-Mira, M. M., Boehringer, D. & Schmid, F. X. (2004). The folding transition state of the cold shock protein is strongly polarized. *J Mol Biol* **339**, 555-69.
44. Gu, H., Kim, D. & Baker, D. (1997). Contrasting roles for symmetrically disposed beta-turns in the folding of a small protein. *J Mol Biol* **274**, 588-96.
45. Kim, D. E., Fisher, C. & Baker, D. (2000). A breakdown of symmetry in the folding transition state of protein L. *J Mol Biol* **298**, 971-84.
46. Ohage, E. C., Graml, W., Walter, M. M., Steinbacher, S. & Steipe, B. (1997). Beta-turn propensities as paradigms for the analysis of structural motifs to engineer protein stability. *Protein Sci* **6**, 233-41.
47. Steipe, B., Schiller, B., Pluckthun, A. & Steinbacher, S. (1994). Sequence statistics reliably predict stabilizing mutations in a protein domain. *J Mol Biol* **240**, 188-92.
48. 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.
49. Mayr, L. M. & Schmid, F. X. (1993). Stabilization of a Protein by Guanidinium Chloride. *Biochemistry* **32**, 7994-7998.
50. Privalov, P. L. & Gill, S. J. (1988). Stability of protein structure and hydrophobic interaction. *Advances in Protein Chemistry* **39**, 191-234.
51. Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Crystallogr* **26**, 795-800.
52. Navaza, J. (2001). Implementation of molecular replacement in AMoRe. *Acta Crystallogr D Biol Crystallogr* **57**, 1367-72.
53. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32.
54. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-55.
55. 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**, 607-18.
56. 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**, 140-147.

Table 1. Stability data for variants of N2'variants with substitutions in turn1 to 3

N2' variant	Thermal unfolding				Urea-induced unfolding				
Turn1 (136-319)	T_m (°C)	ΔH_D (kJ mol^{-1})	$\Delta G_D^{40°C}$ (kJ mol^{-1})	$\Delta\Delta G_D^{40°C}$ (kJ mol^{-1})	$[Urea]_M$ (M)	*m*-Wert (kJ mol^{-1} M^{-1})	$\Delta G_D^{15°C}$ (kJ mol^{-1})	$\Delta G_D^{15°C}$(3.0 M) (kJ mol^{-1})	$\Delta\Delta G_D^{15°C}$(3.0 M) (kJ mol^{-1})
FQNN (wt)	37.87 ± 0.14	318 ± 13	- 2.2	0.0	2.73	6.7 ± 0.2	17.9 ± 0.6	- 1.8	0.0
VQNN	24.43 ± 0.25	292 ± 18	- 17.7	- 15.5	0.53	7.2 ± 0.7	3.8 ±0.3	- 17.8	-16.0
F**N**NN	30.11 ± 0.09	267 ± 5	- 9.7	- 7.5	1.50	6.7 ± 1.1	10.0 ± 2.2	- 10.0	- 8.2
FQ**G**N	41.01 ± 0.06	306 ± 6	1.0	3.2	3.32	6.5 ± 0.8	21.6 ± 1.4	2.1	3.9
FQN**K**	35.92 ± 0.06	293 ± 5	- 4.0	- 1.8	2.23	5.6 ± 0.4	12.6 ± 0.8	- 4.3	- 2.5
FQN**V**	27.74 ± 0.13	227 ± 6	- 10.7	- 8.5	1.06	6.8 ± 1.3	7.2 ± 2.5	- 13.3	- 11.5
F**NG**N	35.85 ± 0.02	291 ± 2	- 4.1	-1.9	2.14	5.6 ± 0.7	12.0 ± 1.6	- 4.8	- 3.0
Turn2 (144-147)									
RQGA (wt)	37.87 ± 0.14	318 ± 13	- 2.2	0.0	2.73	6.7 ± 0.2	17.9 ± 0.6	- 1.8	0.0
VQGA	41.22 ± 0.07	297 ± 6	+ 1.0	+ 3.2	3.37	5.5 ± 0.2	18.5 ± 0.7	+ 2.0	+ 3.8
R**N**GA	39.88 ± 0.07	290 ± 5	- 1.1	+ 1.1	3.15	5.4 ± 0.3	16.9 ± 1.0	+ 0.8	+ 2.6
RQG**K**	35.85 ± 0.06	307 ± 5	- 3.9	- 1.2	2.40	5.7 ± 0.3	13.6 ± 0.8	- 3.4	- 1.6
VNGK	41.03 ± 0.07	311 ± 7	+ 1.0	+ 3.2	3.27	6.0 ± 0.2	19.8 ± 0.7	+ 1.6	+ 3.4
VNGA	43.00 ± 0.06	347 ± 6	+ 3.2	+ 5.4	3.66	5.6 ± 0.2	20.5 ± 1.1	+ 3.7	+ 5.5
VNGV	43.80 ± 0.05	338 ± 6	+ 4.0	+ 6.2	3.96	5.5 ± 0.3	21.7 ± 0.9	+ 5.2	+ 7.0
VNGT	42.45 ± 0.07	336 ± 7	+ 2.6	+ 4.8	3.68	5.5 ± 0.3	20.3 ± 1.0	+ 3.8	+ 5.6
Turn3 $\underline{Q}_{157}$GTDPV$\underline{K}_{163}$									
wt	37.8±0.14	318±13	-2.2	0.0	2.71	6.6±0.2	17.9±0.6	-1.8	0.0
$\underline{Q}_{157}$ **GG** $\underline{K}_{163}$	45.3±0.09	324±9	5.1	7.3	4.03	6.1±0.5	24.2±1.2	6.3	8.1
$\underline{Q}_{157}$ **NG** $\underline{K}_{163}$	46.4±0.13	340±13	6.4	8.6	4.13	6.2±0.2	25.4±0.5	6.8	8.6
$\underline{\mathbf{V}}_{157}$ **NG** $\underline{K}_{163}$	48.5±0.08	284±6	6.8	9.0	4.58	6.1±0.3	27.8±1.4	9.6	11.4
$\underline{\mathbf{Y}}_{157}$ **NG** $\underline{K}_{163}$	46.3±0.09	321±7	5.9	8.1	4.25	6.2±0.3	26.4±0.9	7.8	9.6
$\underline{Q}_{157}$ **GDN** $\underline{K}_{163}$	38.4±0.08	305±7	-1.6	0.6	2.86	6.5±0.5	18.6±0.5	-0.9	0.9
$\underline{Q}_{157}$ **GQN** $\underline{K}_{163}$	35.8±0.07	293±6	-4.2	-2.0	2.29	6.8±0.3	15.6±1.4	-4.8	-3.0
$\underline{Q}_{157}$ **GQD** $\underline{K}_{163}$	35.1±0.07	310±6	-5.2	-3.0	2.28	7.0±0.5	16.0±1.3	-5.0	-3.2
$\underline{Q}_{157}$ **ADG** $\underline{K}_{163}$	42.9±0.09	308±10	2.8	5.0	3.58	5.7±0.3	20.5±1.2	3.3	5.1
3T-N2'	54.89 ± 0.08	304 ± 8	+ 11.8	+ 14.0	6.54	5.1 ± 0.4	33.0 ± 2.0	+15.2	+17.0

The stability parameters result from two-state analyses of the thermal unfolding transitions (columns 2–5) and urea-induced unfolding transitions (columns 6–10). For all proteins, the melting temperature (T_M), the van't Hoff enthalpy of denaturation at T_M (ΔH_D), the midpoint of the urea-induced unfolding transition ($[urea]_M$), the cooperativity value (*m*), the Gibbs free energy of denaturation (ΔG_D) at 15 °C in 0.0 M urea and the Gibbs free energy at 3.0 M urea ($\Delta\Delta G_D$) are

given. The thermal and the urea-induced unfolding transitions were measured in 100 mM K phosphate (pH 7.0). All measurements were performed in 100 mM K phosphate (pH 7.0) at 15 °C. The parameter for 3T-N2' are apparent values, because its thermal unfolding is largely irreversible.

Table 2. Unfolding and refolding kinetics of N2' variants with substitutions in turns 1 to 3

N2'-variant	k_{UN} (s^{-1})	k_{NU} (s^{-1})	m_{UN} (kJ mol^{-1} M^{-1})	m_{NU} (kJ mol^{-1} M^{-1})	m (kJ mol^{-1} M^{-1})	$[urea]_M$ (M)	$\Delta G_D^{15°C}$ (3.0M) (kJ mol^{-1})
wt(cis)	36.1±0.22	0.43±0.05	-3.4±0.06	1.9±0.04	5.3	3.8	4.8
wt(trans)	0.7±0.06	0.40±0.05	-4.1±0.03	1.8±0.05	6.0	1.8	-7.1
Turn1 (136-139)							0
FQ**G**N(cis)	180±23	0.50±0.09	-3.2±0.17	2.0±0.22	5.2	4.5	7.8
FQ**G**N(trans)	3.9±0.17	0.45±0.06	-4.0±0.14	1.9±0.05	5.9	2.5	-2.9
							0
Turn2 (144-147)							0
VNGA(cis)	294±77	0.46±0.05	-3.3±0.30	1.8±0.30	5.1	4.8	9.2
VNGA(trans)	8.9±0.9	0.42±0.08	-4.2±0.14	1.9±0.07	6.1	2.8	-1.2
VNGV(cis)	2213 ±389	0.82±0.18	-3.4±0.80	2.0±0.07	5.4	5.3	12.4
VNGV(trans)	32.1±4.0	1.29±0.34	-4.1±0.20	1.9±0.05	6.0	2.9	-0.6
Turn3 Q_{157}GTDPVK_{163}							0
Q_{157} **GG** K_{163}	76.7±4.8	0.34±0.05	-3.4±0.04	2.3±0.04	5.7	4.15	6.6
Q_{157} **NG** K_{163}	195±21	0.39±0.05	-3.8±0.04	2.0±0.02	5.8	4.24	7.2
V_{157} **NG** K_{163}	331±25	0.25±0.07	-3.1±0.03	2.2±0.08	5.3	4.81	9.6
Q_{157} **GDN** K_{163}	5.8±0.6	0.51±0.04	-3.8±0.07	2.3±0.03	6.1	2.67	-2
Q_{157} **ADG** K_{163}	64.4±8.1	0.46±0.06	-3.9±0.03	2.1±0.05	6.0	3.66	4
3T-N2'	41400±8000	1.53±0.11	-3.4±0.14	1.8±0.25	5.2	6.4	17.6

The kinetic data were determined by fluorescence in stopped-flow experiments. The microscopic rate constants of refolding k_{UN} at and unfolding k_{NU} at 3 M urea are given. m_{UN} and m_{NU} are the kinetic m values for refolding and unfolding, respectively. The transition midpoint $[urea]_M$ is calculated as $[urea]_M = \ln(k_{NU}/k_{UN})/(m_{UN}-m_{NU})$. The cooperativity values (m) are calculated from the kinetic m values ($m = RT(m_{NU}-m_{UN})$), the Gibbs free energies of denaturation at 15 °C, $\Delta G_D^{15°C} = -RT\ln(k_{NU}/k_{UN})$. The data were analyzed according to a two-state model. The concentrations of proteins were 1 µM. in 100 mM K phosphate (pH 7.0).

Table 3. Stability data for the selected variants of N2 with substitutions in turn 3

N2 variant	Thermal unfolding				Urea-induced unfolding				
	T_m (°C)	ΔH_D (kJ mol^{-1})	$\Delta G_D^{40°C}$ (kJ mol^{-1})	$\Delta\Delta G_D^{40°C}$ (kJ mol^{-1})	$[Urea]_M$ (M)	*m*-Wert (kJ mol^{-1} M^{-1})	$\Delta G_D^{15°C}$ (kJ mol^{-1})	$\Delta G_D^{15°C}$(3.0 M) (kJ mol^{-1})	$\Delta\Delta G_D^{15°C}$(3.0 M) (kJ mol^{-1})
wild type	33.65 ± 0.06	299 ± 6	- 6.5	0.0	2.12	7.0 ± 0.2	14.3 ± 0.4	- 6.2	0
INGR	43.59 ± 0.06	307 ± 5	+ 3.4	+9.9	4.03	5.4 ± 0.3	21.9 ± 1.2	+ 5.6	11.8
IDGR	43.27 ± 0.05	307 ± 4	+ 3.1	+9.6	3.87	5.8 ± 0.2	22.7 ± 0.8	+ 5.0	11.2

The stability parameters result from two-state analyses of the thermal unfolding transitions (columns 2–5) and urea-induced unfolding transitions (columns 6–10). For all proteins, the melting temperature (T_M), the van't Hoff enthalpy of denaturation at T_M (ΔH_D), the midpoint of the urea-induced unfolding transition ($[urea]_M$), the cooperativity value (*m*), the Gibbs free energy of denaturation (ΔG_D) at 15 °C in 0.0 M urea and the Gibbs free energy at 3.0 M urea ($\Delta\Delta G_D$) are given. The thermal and the urea-induced unfolding transitions were measured in 100 mM K phosphate (pH 7.0). All measurements were performed in 100 mM K phosphate (pH 7.0) at 15 °C.

Table 4. Statistics on diffraction data and structure refinement

Data set	3T-G3P
Total/unique reflections	136796/26332
R_s*	0.068 (0.351)
Resolution, Å	30–2.2 (2.3–2.2)
Completeness, %	97.4 (91.6)
$(I)/(\sigma I)$	13.4 (2.0)
Model R/R_{free}-factor, %†	25.9/31.9
rmsd from ideal geometry	
Bonds, Å	0.013
Angles, °	1.8
Cruickshank DPI for coordinate error	0.218
Ramachandran statistics, %	
Most favored/additional/generously allowed/disallowed regions	81.4/18.3/0.3/0.0

In the refinement statistics, Friedel mates were merged.

*$R_s = \Sigma_h \Sigma_i I_i(h) - \langle I(h) \rangle / \Sigma_h \Sigma_i I_i(h)$, where i are the independent observations of reflection h.

†The R_{free} factor was calculated from 5% of the data, which were removed at random before the refinement was carried out.

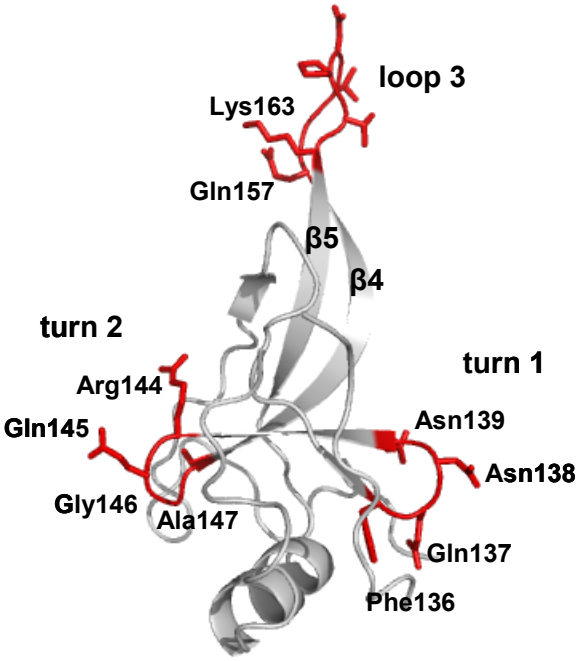


Figure 1. (a) Tertiary structure of the N2 domain of G3P (coordinates from [21]). The side-chains of turns 1 and 2 and of loop 3 are shown in stick representation. The β sstrands 4 and 5 are labeled. The figure was prepared by using Pymol

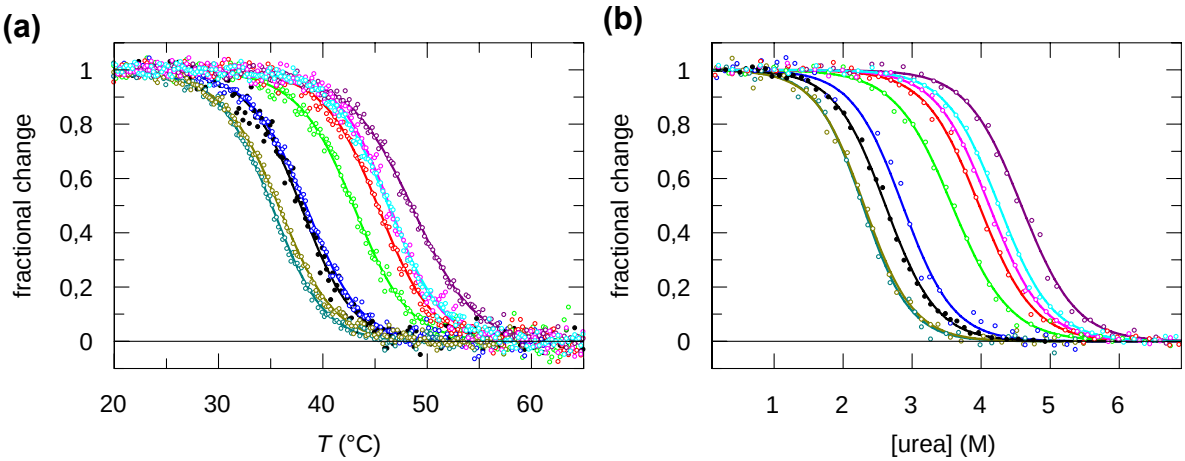


Figure 2. (a) Thermal and **(b)** urea-induced equilibrium unfolding transitions of the N2' domain (black) and the N2' turn3- variants (—) GG; (—) VNG; (—) GDN; (—) GQN; (—) GQD; (—) ADG; (—) NG and (—)YNG. The thermal transitions of 4 µM protein were measured by CD at 222 nm, the urea-induced transitions of 1 µM protein were measured by protein fluorescence at 340 nm after excitation at 280 nm at 15 °C. The fractional changes were obtained from two-state analyses of the data. The results from these analyses are shown in Table 1.

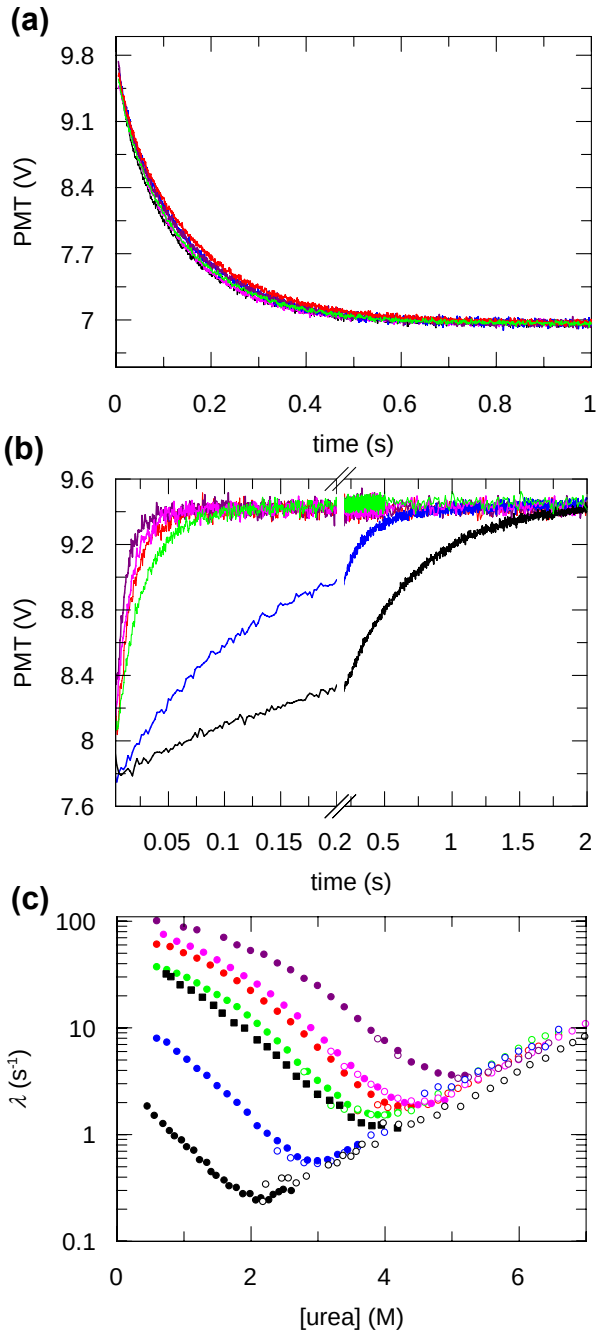


Figure 3: **(a)** Unfolding kinetics in 6.0 M urea and (b) refolding kinetics in 0.6 M urea of (—) N2′ and the N2’ turn3 variants (—) GG, (—) VNG, (—) NG, (—) GDN, (—) ADG. **(c)** Refolding (filled symbols) and unfolding (open symbols) kinetics of turn3 variants (●,○) GG, (●,○) VNG, (●,○) NG, (●,○) GDN, (●,○) ADG and (■,●,○) N2’ are shown as a function of the urea concentration. For N2’, two refolding reactions, of the *cis* (■) and *trans* (●) forms of Pro161 were measured. The folding kinetics were followed after stopped-flow mixing by the changes in fluorescence above 320 nm (excitation at 280 nm). The results of the analysis are given in Table 2. All experiments were performed in 100 mM K phosphate (pH 7.0).

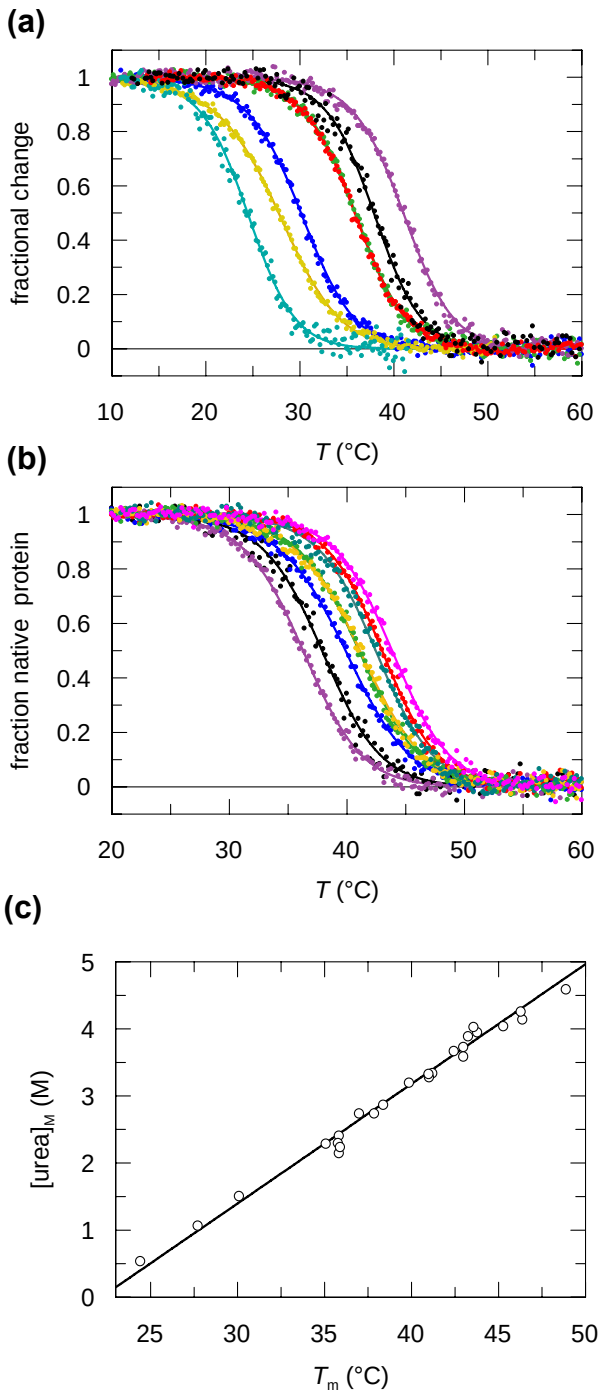


Figure 4. Thermal unfolding transitions of the N2' domain (black) and **(a)** the N2' turn 1 variants (—) VQGA, (—) RNGA, (—) RQGK, (—) VNGA, (—) VNGV, (—) VNGK and (—) VNGT and **(b)** the N2' turn 2-variants (—) FQGN, (—) FNGN, (—) FNNN, (—) FQNK, (—) FQNV, und (—) VQNN. The thermal transitions of 4 µM protein were measured by CD at 222 nm in 100 mM K phosphate (pH 7.0). The fractional changes were obtained from two-state analyses of the data. The results from these analyses are shown in Table 1. **(c)** Correlation between the thermal and the urea-induced unfolding transitions. For all variants the midpoints of the urea-induced transitions, $[urea]_M$, are plotted as a function of the corresponding midpoints of the thermally induced unfolding transitions, T_M. The midpoints are taken from Table 1. The straight line represents a linear fit to the data.

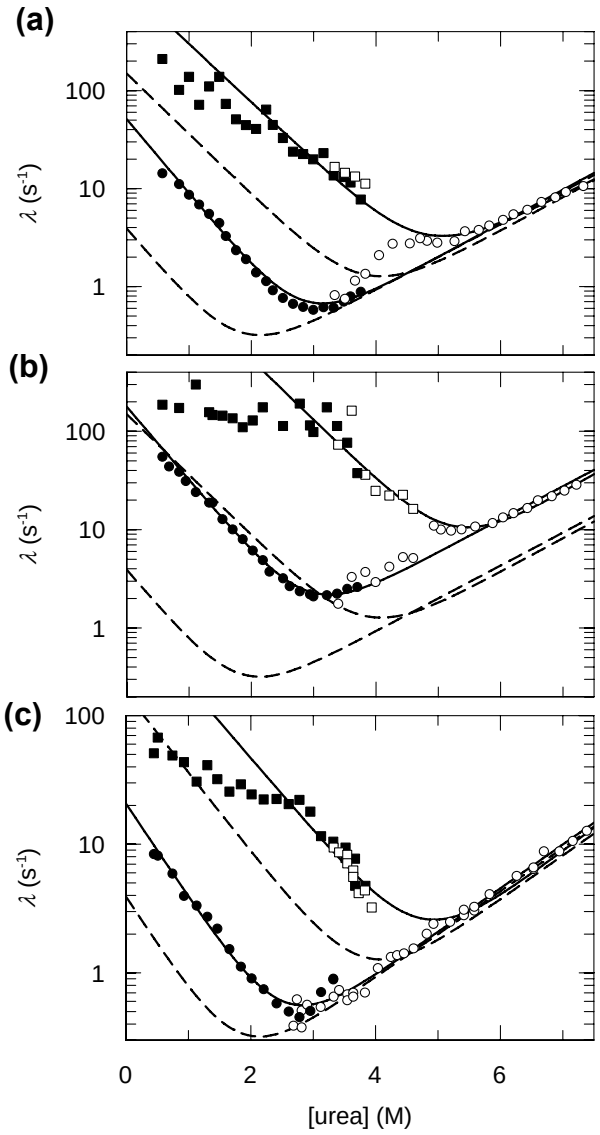


Figure 5. Folding kinetics of **(a)** N2'-T2-VNGA, **(b)** N2'-T2-VGNV and **(c)** N2'-T1-FQGN. The apparent rate constants λ_3 (squares) and λ_2 (circles) are shown as a function of the urea concentration. Chevrons fitted to the experimental data on the basis of linear two-state models are shown by the continuous lines. The results of the analysis are given in Table 2. The folding kinetics were measured after stopped-flow mixing by the changes in fluorescence above 320 nm (excitation at 280 nm). The dashed lines represent the fitted chevrons for wild-type N2'. All measurements were performed in 100 mM K phosphate (pH 7.0) at 15 °C.

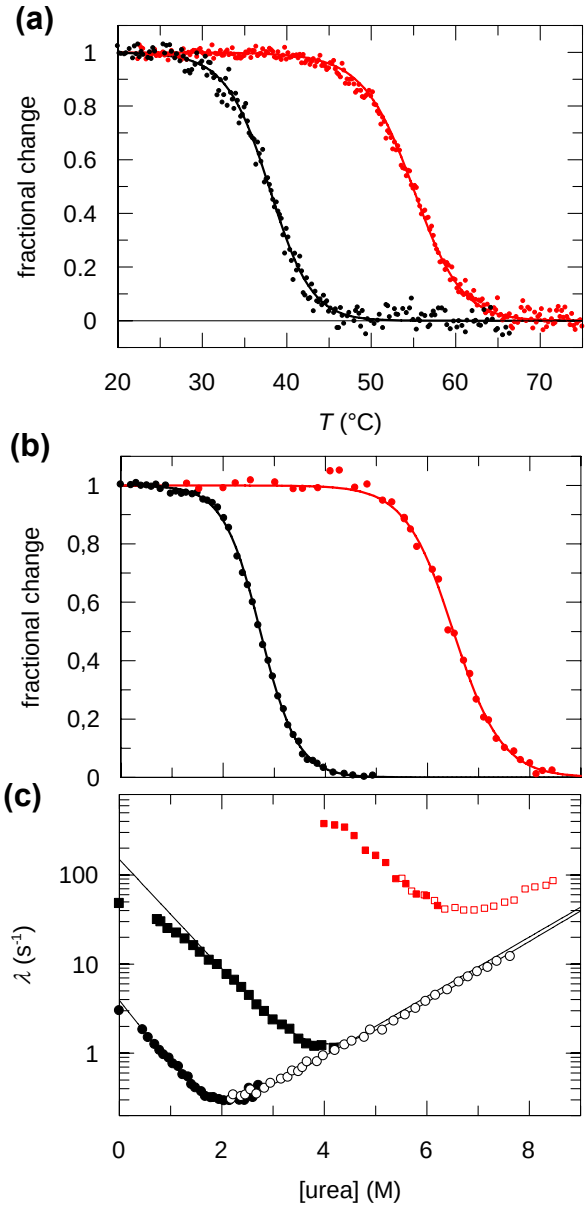


Figure 6. (a) Thermal and **(b)** urea-induced unfolding transitions of the variant obtained by manual combination of all stabilizing mutations for the turns 1, 2, and 3 (red) and of wild-type N2' (black) The thermal transitions of 4 µM protein were measured by CD at 222 nm, the urea-induced transitions of 1 µM protein were measured by protein fluorescence at 340 nm after excitation at 280 nm at 15 °C. The fractional changes were obtained from two-state analyses of the data. The results from these analyses are shown in Table 1. The transitions were measured in 100 mM K phosphate (pH 7.0). (c) Comparison of the folding kinetics of (■,●,□,○) N2' and (■,□) 3T-N2'. The apparent rate constants λ are shown as a function of the urea concentration. The filled symbols represent refolding the open symbols unfolding kinetics. A chevron fitted to the experimental data on the basis of linear two-state model is shown by the continuous line.

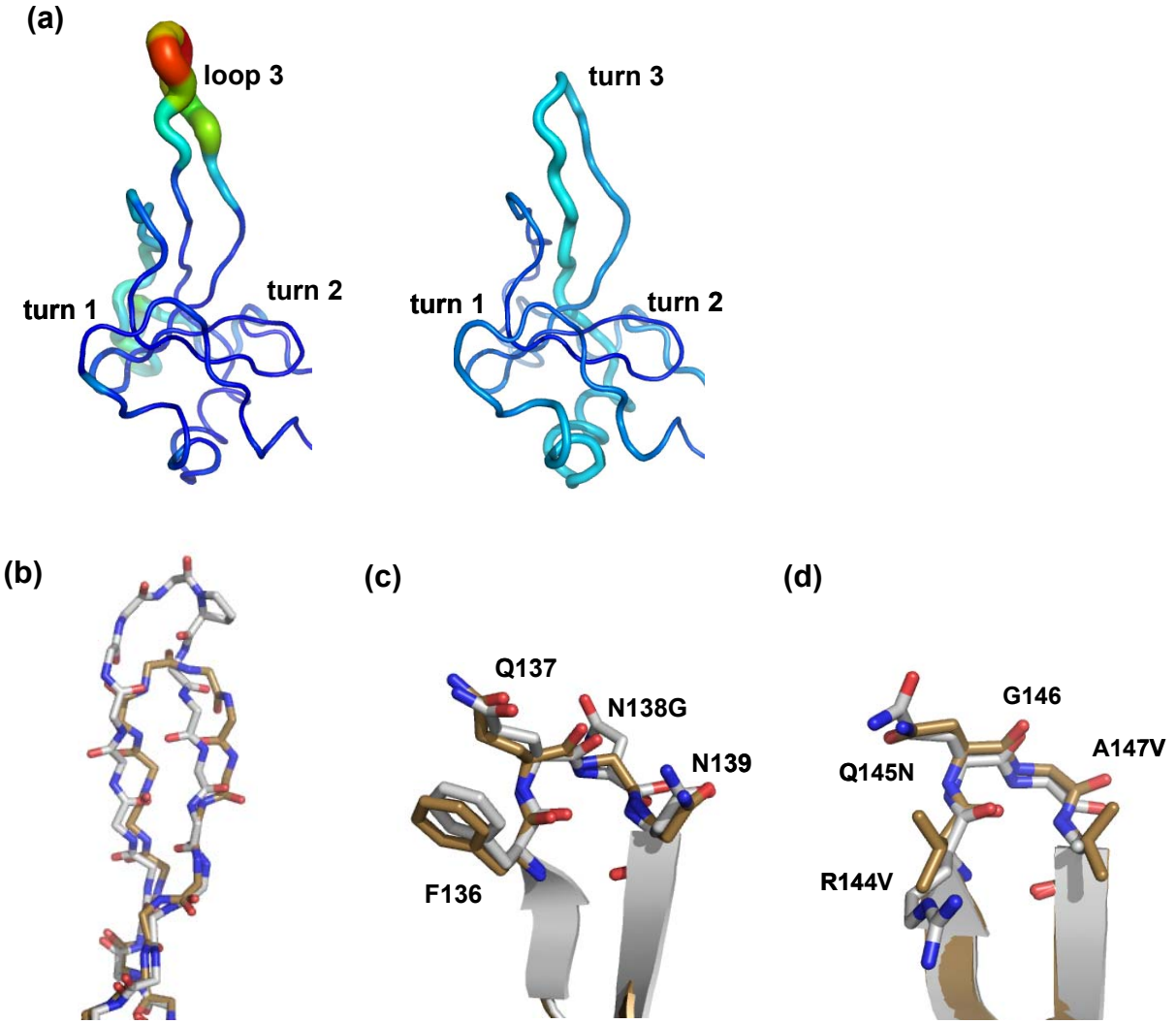


Figure 7: (a) Comparisons of the B-Factors for the crystal structures of wild-type G3P* (left) and 3T-G3P* (right). Turns 1, 2 and loop 3 are labeled. Comparison of the local structures of wild-type G3P* (gray) and 3T-G3P* (brown) in the vicinities of (b) the stabilized loop 3 (c) the stabilized turn 1 and (d) the stabilized turn 2. For this comparison, the 1G3P structure of wild-type G3P was used [56]. The figure were prepared by using Pymol.

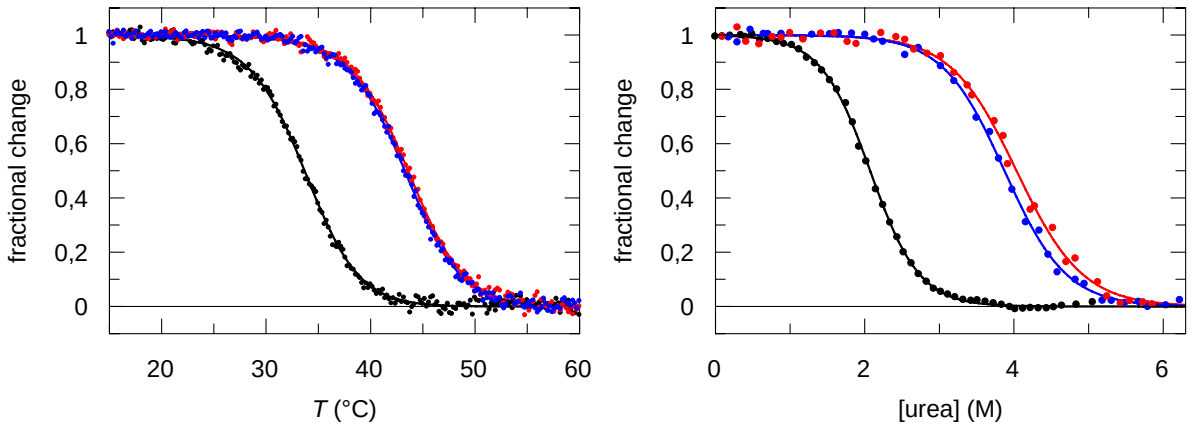


Figure 8. (a) Thermal and **(b)** urea-induced equilibrium unfolding transitions of the N2 domain (black) and the selected variants N2 T3-INGR (—) and N2 T3-IDGR (—). The thermal transitions of 4 µM protein were measured by CD at 222 nm, the urea-induced transitions of 1 µM protein were measured by protein fluorescence at 340 nm after excitation at 280 nm at 15 °C. The fractional changes were obtained from two-state analyses of the data. The results from these analyses are shown in Table 3. The transitions were measured in 100 mM K phosphate (pH 7.0).

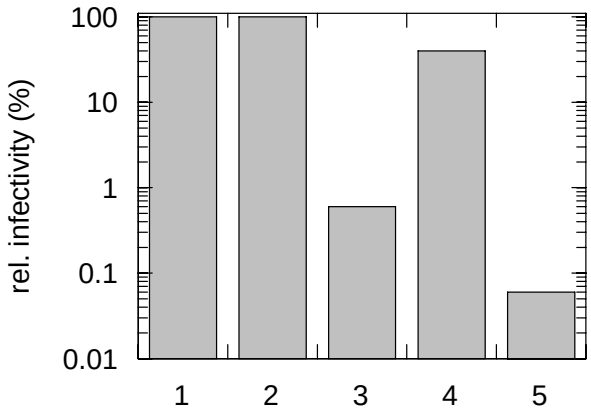


Figure 9: Relative infectivity of the phage variants with the most stable turn1, 2, 3 sequence. Phage with the wild-type G3P (1), with the G3P-T1-FQGN (2), with the G3P-T2-VNGV (3). with the G3P-T3-VNGK (4), with the G3P-T1-FQGN/T2-VNGV/T3-VNGK (5). *E. coli* TG1 cells were infected with samples of the phage variants with equal phage concentration. Infection was performed in the presence of an excess of E. coli TG1 cells. The infectivity of the wild type phage fd is set as 100%.

Supplementary data

Table S1: Backbone dihedral angles from different gene-3-protein structures for the L1 and L2 positions in the type I' turns 1, 2, 3

	turn1				turn2				turn3			
	L1		L2		L1		L2		L1		L2	
pdb-code	Φ	Ψ	Φ	Ψ	Φ	Ψ	Φ	Ψ	Φ	Ψ	Φ	Ψ
Typ I' turn	60	30	90	0	60	30	90	0	60	30	90	0
1G3P	56	40	67	10	55	42	95	-13				
2G3P(A)	49	48	63	20	53	49	94	-6				
2G3P(B)	50	47	61	29	55	44	92	-5				
3DGS(A)	62	35	80	-1	52	39	103	-13				
3DGS(B)	57	37	81	-3	47	52	88	-4				
3T-G3P(A)	51	31	75	6	60	26	92	-3				
3T-G3P(B)	51	36	73	12	45	45	77	10	51	40	105	-15

Dihedral angles were taken from the pdb-files. In crystal structures of the gene-3-protein 2 molecules were in the asymmetric unit, the molecule number is given in parentheses.

Teilarbeit G:

Weininger, U., **Jakob, R.P.**, Eckert, B., Schweimer, K., Schmid, F.X and Balbach, J.

A remote prolyl isomerization controls domain assembly via a hydrogen bonding network., *Proc. Natl. Acad. Sci. akzeptiert*

Classification: Biological Sciences, Biophysics

A remote prolyl isomerization controls domain assembly via a hydrogen bonding network

Ulrich Weininger[1], Roman P. Jakob[2], Barbara Eckert[2], Kristian Schweimer[3], Franz X. Schmid[2], Jochen Balbach[*,1]

[1] Institut für Physik, Biophysik, and Mitteldeutsches Zentrum für Struktur und Dynamik der Proteine (MZP), Martin-Luther-Universität Halle-Wittenberg, D-06120 Halle(Saale), Germany

[2] Laboratorium für Biochemie, Universität Bayreuth, D-95440 Bayreuth, Germany

[3] Lehrstuhl für Biopolymere, Universität Bayreuth, D-95440 Bayreuth, Germany

*Correspondence should be addressed to

Jochen Balbach	Tel.:	++49 345 55 28550
Institut für Physik, Biophysik	FAX:	++49 345 55 27161
Martin-Luther-Universität Halle-Wittenberg	e-mail:	jochen.balbach@physik.uni-halle.de
D-06120 Halle (Saale), Germany	e-mail:	fx.schmid@uni-bayreuth.de

Manuscript information: 21 text pages, 135 words (abstract), 36694 characters (text), 10566 characters (estimated for 4 figures)

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 and contains the stabilizing mutations T13I, T101I, Q129H, and D209Y; I, folding intermediate of G3P* with *trans* Pro213; GdmCl, guanidinium chloride; ΔG_D, Gibbs free energy of denaturation; TROSY, transverse relaxation optimized spectroscopy; HSQC, hetero single quantum coherence; HX, hydrogen exchange; NH, amide proton.

Abstract

Prolyl *cis*/*trans* isomerizations determine the rates of protein folding reactions and can serve as molecular switches and timers. In the gene-3-protein of filamentous phage, Pro213 *trans* → *cis* isomerization in a hinge region controls the assembly of the two domains N1 and N2 and, in reverse, the activation of the phage for infection. We elucidated the structural and energetic basis of this proline-limited domain assembly at the level of individual residues by real-time 2D NMR. A local cluster of inter-domain hydrogen bonds, remote from Pro213, is stabilized up to 3000fold by *trans* → *cis* isomerization. This network of hydrogen bonds mediates domain assembly and is connected with Pro213 by rigid backbone segments. Thus, proline *cis/trans* switching is propagated in a specific and directional fashion to change the protein structure and stability at a distant position.

Key words:
Protein folding, proline isomerization, molecular timer, real-time NMR, gene-3-protein, domain assembly

\body

Introduction

The *cis*/*trans* isomerizations of peptidyl-prolyl bonds in proteins are intrinsically slow processes. They determine the rates of protein folding reactions and are used as slow switches to regulate biological processes (1-8). In unstructured protein chains, the *trans* form is favored over *cis* and therefore proteins with *cis* prolyl bonds in the native state must undergo *trans* → *cis* isomerization during their folding (9, 10). In this case, folding typically starts with a particular proline still in the incorrect (*trans*) state, but when a certain extent of folding is reached, this *trans* proline acts as a barrier and blocks further folding (11-13).

The coupling between conformational folding and prolyl isomerization has been studied for several small single-domain proteins (14-19). In the folding of the gene-3-protein (G3P) of phage fd, prolyl isomerization determines the rate of the final domain assembly step. G3P consists of three domains. The carboxy terminal (CT) domain anchors G3P in the phage coat (20), and the two amino terminal domains N1 and N2 form a functional entity that protrudes from the phage surface and mediates the infection of *E. coli* cells (21). In the N1-N2 unit (Fig. 1a), N1 (residues 1-67) and N2 (residues 105-204) are linked by the hinge region, which forms numerous contacts with the N1 domain. The hinge region is formed by two non-contiguous chain regions (residues 89-104 between N1 and N2 and residues 205-220 after N2) and shows a well-ordered structure in the native protein.

The time course of folding of G3P extends from milliseconds to hours (Fig. 1b). It starts with an extremely rapid folding reaction in the N1 domain with a time constant τ = 9.4 ms, followed by two folding reactions in N2, which show τ values of 7 s and 42 s. In the final, very slow step (τ = 6200 s), the two domains assemble in a reaction that involves further folding in the hinge region and is limited in rate by the *trans* → *cis* isomerization of Pro213 (5, 22).

When G3P is fully folded, the phage is not infectious, because the binding site for its receptor is buried at the interface between the domains N1 and N2. The final, Pro213-limited folding step must be reverted to unlock the domains and activate the phage for infection (5, 6, 23-25). This slow process is thus of dual importance. It represents the final rate-limiting step of folding of G3P as well as the transition between the partially folded, functionally active state and the fully folded but functionally inactive state. Therefore, G3P belongs to the increasing number of examples reported recently, where the isomeric state of a proline modulates activity, ligand binding and physiological regulation (4, 8, 26-29).

Here we employed real-time 2D NMR spectroscopy (30) and amide hydrogen exchange experiments (31, 32) to analyze the structure and the stability of the folding intermediate of G3P at the level of individual residues. In the final domain docking reaction a chain of interconnecting hydrogen bonds gets stabilized. This provides a molecular link between the isomeric state of remote Pro213 and the stability of the inter-domain interface, which controls infectivity.

RESULTS

We used a G3P* variant of the two N-terminal domains with four substitutions (T13I, T101I, Q129H, and D209Y), which is more stable than wild-type G3P and thus better suited for refolding studies (33). Phage carrying this stabilized version of G3P infect *E. coli* cells, the infectivity is, however, tenfold reduced, because the locked, non-infectious form of G3P* is stabilized relative to the unlocked, infectious form (6). To study the folding mechanism of a protein by liquid state NMR spectroscopy at residue level, assignments of backbone amide resonance are required. G3P* consists of 226 residues, and shows a well-dispersed 2D ^{15}N-TROSY-HSQC spectrum. More than 90 % of the amide cross peaks of G3P* could be assigned (see supporting information (SI) Fig. 5) by standard triple resonance NMR experiments. These assignments together with the crystal structure of wild-type N1-N2 G3P (21) form the basis for the high resolution study of the structure and stability of the folding intermediate of G3P* with an incorrect *trans* Pro213.

The folding intermediate of G3P* shows non-native interactions in the N1 domain and in the hinge region

The final folding reactions of the N1 and N2 domain (Fig. 1b) are complete within 5 min, but domain assembly is very slow (6, 34). The partially folded intermediate I with an incorrect *trans* isomer at Pro213 and incompletely assembled domains therefore exists for an extended time (> 8 h) during the refolding of G3P*. To identify those regions in the structure of I that are native-like and those that are not yet native, we employed a kinetic 2D real-time NMR experiment (30, 31). In this experiment, refolding was started by a fourfold dilution of the unfolded protein (in 4.0 M GdmCl) with 50 mM potassium phosphate buffer, pH 7.0 at 25 °C. After the completion of the folding reactions of the individual domains N1 and N2 (5 min), we measured a single ^{15}N-TROSY-HSQC spectrum between 30 min and 510 min (in 1.0 M GdmCl at 25 °C), a time window that covers the kinetics of the slow, proline-limited domain assembly reaction (I→N). This “kinetic spectrum” thus contains resonances from two populations of G3P*, the folding intermediate I, which decreases in concentration, and the

folded protein N, which increases in concentration during the acquisition time of the kinetic spectrum (Fig. 2a). After the completion of refolding, a second NMR spectrum was recorded under identical conditions. It is referred to as the "reference spectrum" of the fully folded protein with a *cis* Pro213 (Fig. 2b). Finally, the kinetic spectrum was subtracted from the reference spectrum to obtain the "difference spectrum" (Fig. 2c).

Two sets of signals (31, 35) were analyzed: The signals of set 1 originate from amide NH that are already in a native environment in the intermediate I. They show identical chemical shifts and identical intensities in I and N, and thus they vanish when the kinetic spectrum is subtracted from the reference spectrum. In the reference spectrum they show a positive peak. Gln116 and Asn182 provide examples for such native-like residues in I (Fig. 2).

The signals of set 2 originate from those amides that are in a non-native environment in the intermediate and become native during the rate-limiting folding reaction, which takes place during the acquisition of the kinetic spectrum. At native chemical shift positions, they show a reduced intensity in the kinetic spectrum, because only N, but not I, contributes to these signals. Therefore, a reduced, but positive signal is observed in the difference spectrum. Examples are provided by Thr56 or Ser208 (Fig. 2).

In Figure 2d, the residues with native-like chemical shifts in the intermediate (set 1) are depicted in blue, those with a non-native chemical shift (set 2) in red. A detailed listing of individual residues is given in SI Table 1. Almost all of the assigned residues of the hinge belong to set 2, evidently because the hinge is not native-like structured in the folding intermediate with the non-native *trans* Pro213. In the N1 domain, the residues that are close to the hinge (shown in red in Fig. 2d) also do not show native chemical shifts, whereas those distant from the hinge subdomain are in a native-like environment in the intermediate already (shown in blue in Fig. 2d). In the N2 domain, almost all residues are in a native conformation in the folding intermediate except Asn138 and Asn139. They form contacts with N1, which apparently cannot be fully established when Pro213 in the hinge is still in the incorrect *trans* conformation. Together, this suggests that the incorrect *trans* isomer of Pro213 does not allow folding of the hinge to go to completion, and in turn, the hinge and the domains N1 and N2 cannot associate as in the fully folded state.

Folded G3P* unfolds as a cooperative unit under native conditions

Chemical shift analyses such as those in Figure 2 provide detailed, site-specific information about the structure but not the stability of the folding intermediate. Stability data can be derived from the protection of individual amide protons from exchange with the

aqueous solvent (36). First, the fully folded form of G3P* was studied by amide hydrogen exchange (amide HX) measurements at pH 7.0, 25 °C. The protonated (N^H) protein was transferred from 50 mM phosphate buffer in H_2O to the same buffer in D_2O, and the decrease in the intensities of the NH cross peaks were followed by a series of ^{15}N-TROSY-HSQC spectra as a function of time for 7000 min. We assumed that individual NH exchange by an EX2 mechanism (see SI Figure 6) in a monoexponential reaction. Protection factors (*P*) were calculated by using the intrinsic exchange rates of corresponding amides in peptides as a reference (37). They are shown in Figure 3a and in SI Table 1. Several NH of G3P* exchanged very slowly, and therefore the NMR cross peak intensity decreased less than 5% during the HX experiment. For these NH, a lower limit for *P* of $2 \cdot 10^6$ is assumed, as indicated in Figure 3a. This represents the highest protection factor of residues with well defined exchange kinetics. It is equivalent to $\Delta G_D \geq 36$ kJ/mol, which agrees with the ΔG_D value of 37 kJ/mol, obtained from denaturant-induced unfolding transitions measured by fluorescence spectroscopy (33). No protection was observed for the glycine-rich unordered linker between N1 and the hinge, which extends from residue 68 to residue 88.

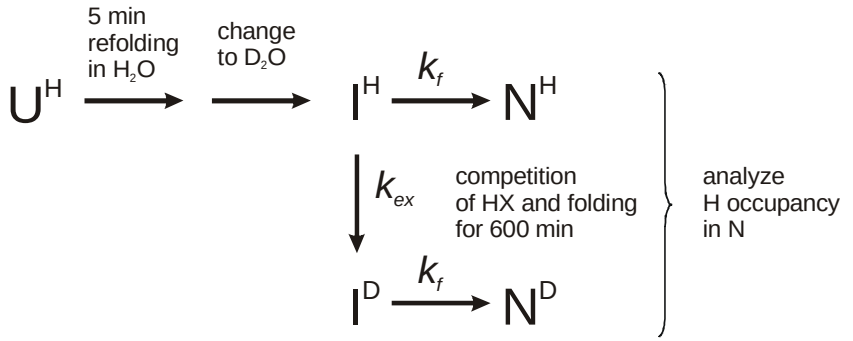
Protection factors of 10^6 or higher were observed for residues in all three structural units of G3P*: N1, N2, and the hinge region (Fig. 3a). This suggests that unfolding of G3P* under native conditions is a cooperative reaction, in which domain disassembly and domain unfolding are coupled reactions.

For the isolated N1 domain (Fig. 3c), the most strongly protected amide NH show protection factors of about $2 \cdot 10^5$. This is equivalent to an ΔG_D value of 30 kJ/mol, in good agreement with the ΔG_D value of 29.2 kJ/mol, as calculated from the GdmCl-induced unfolding transition of isolated N1 (33). The inter-domain interactions in G3P* mediated by the hinge thus increase the stability of the N1 domain at least tenfold. A particularly strong stabilization is provided for amide NH around β strand 1 of N1 (residues 10-25), which contacts the hinge in folded G3P*.

HX experiments could not be performed with the isolated N2 domain, because it aggregates at high concentration under the exchange conditions (38). Urea-induced unfolding transitions measured at low protein concentrations gave an equilibrium constant of folding for isolated N2 of 180 at 15 °C. This shows that N2 is only marginally stable and at least 10^4-fold stabilized when it is assembled with the hinge and the N1 domain in fully folded G3P*.

A competition experiment between folding and amide HX exchange reveals the local stability of the folding intermediate.

The conformational stabilities of the individual structural elements of G3P* in the folding intermediate were estimated from a competition experiment between refolding to the native state and amide HX (scheme 1) (31, 32). In this experiment, unfolded G3P* (U^H) was refolded for 5 min in H_2O buffer to populate the protonated intermediate state (I^H) with folded but unassembled domains and *trans* Pro213 in the hinge. Next, the H_2O-containing buffer was replaced by D_2O buffer without GdmCl to initiate the competition between the folding of the intermediate to the native state and its amide HX (scheme 1). After completion of the competition experiment (after 600 min), a ^{15}N-TROSY-HSQC was recorded to determine the intensities of the remaining NH cross peaks (N^H). In a reference experiment, native G3P* (N^H) was equilibrated for the same time in the same buffer in D_2O before measuring an identical ^{15}N-TROSY-HSQC experiment. This reference spectrum provided the full intensity of the NH cross peaks in the native protein.



[scheme 1]

To facilitate the interpretation, the analysis was restricted to those 48 amide NH that are highly protected in the native protein and thus retained the full cross peak intensity during the experiment (gray background in Fig. 3b). A decreased cross peak intensity in the competition experiment relative to the reference experiment shows that partial amide HX must have occurred in the folding intermediate at a particular residue. Assuming that the amide HX from the folding intermediate follows an EX2 mechanism, the protection factor *P* of individual NH can be calculated with equation 1(31). The rate of folding from the intermediate to the native state ($k_f = 1.6 \cdot 10^{-4}$ s^{-1}) is taken from our previous work (22), and the intrinsic exchange rates (k_{int}) were taken from ref. (37). *S* is the ratio of signal intensities as observed in the competition experiment (equivalent to N^H in scheme 1) and in the reference experiment (equivalent to $N^H + N^D$ in scheme 1).

$$P = \frac{k_{\text{int}} \cdot S}{k_{\text{f}} \cdot (1 - S)} \qquad (1)$$

The protection factors calculated for the folding intermediate by this procedure are shown in Figure 3b and in SI Table 1. They are rather uniform for N1 and N2 and typically, in the range of $\geq 3x10^5$, which is equivalent to a ΔG_D value of about 30 kJ/mol. In the C-terminal region of N1, the protection factors for the most highly protected NH (Fig. 3b) are similar in the folding intermediate and in the isolated N1 domain (Fig. 3c). In the N-terminal region, however, the protection factors are higher than for the isolated N1 domain, which indicates that this part is locally stabilized in the intermediate.

In the folding intermediate, the N2 domain is at least 10^3-fold stabilized compared to N2 in isolation (38), and only about tenfold less stable than in the fully folded protein. This correlates well with the finding that N2 shows native-like chemical shifts in the folding intermediate already (Fig. 2). In the hinge subdomain only six amide NH satisfied the criterion for our analysis, and they are not or only marginally protected in the folding intermediate (Fig. 3a, b). For residues that cluster in the β-strands of N1 that point towards the hinge and in the β-strands of the hinge itself up to 10^3-fold increases in the protection factors were observed (Fig. 4a).

DISCUSSION

Trans Pro213 arrests N1 and the hinge in partially folded conformations and maintains G3P in an unlocked form

The N1 and N2 domains of G3P* start to fold individually, but their assembly is very slow, because it is limited in rate by the *trans* → *cis* isomerization of Pro213 in the hinge between the domains (Fig. 1b). This isomerization retards folding more than 10^3-fold. When Pro213 is *cis*, the two domains become tightly locked, which, in turn, retards unfolding, also about 10^3-fold. Our real-time 2D NMR experiments in combination with amide HX revealed how Pro213 *trans* → *cis* isomerization changes the structure of G3P* and the local stability during the transition from the unlocked folding intermediate, which is the infectious state, to the locked, fully folded state.

The *trans* isomer of Pro213 in the unlocked folding intermediate affects the individual structural elements of G3P* differently. The N2 domain has virtually reached a conformation as in the fully folded protein already before the domain locking reaction (Fig. 2). However, in the hinge region and in parts of N1 facing the hinge, extended regions with non-native structure were detected. Folding of the hinge is thus coupled to Pro213 *trans* → *cis*

isomerization and the N1 domain cannot associate with the hinge in a native-like fashion when Pro213 is still *trans*. In folded G3P* (21), most inter-domain contacts are in fact localized between the hinge region and the N1 domain. Apparently, they are controlled by Pro213 and cannot be established before Pro213 *trans* → *cis* isomerization has occurred in the final step of folding. The loops Asp24-Asp28 in N1 and Phe136-Asn139 in N2 also change their structures during the slow domain locking reaction (Fig. 2d). These loops are remote from the hinge, but involved in the contacts between N1 and N2 in the folded protein. Evidently, the inter-domain contacts in this remote area are also controlled by Pro213 isomerization.

Pro213 isomerization couples the N1 and N2 domains to a single cooperative unit.

The equilibrium unfolding of G3P* consists of two successive transitions (6, 22). In the first transition, the domains disassemble and N2 unfolds, but N1 remains native-like. It unfolds in the second transition, which occurs at higher temperature or denaturant concentration as confirmed by experiments with the isolated domains. N1 shows the same transition midpoints as part of G3P* and in isolation. N2, however, is much less stable in isolation (22, 33, 38).

The analysis of the amide HX experiments provided a residue-specific insight into the local stability of the fully folded protein with *cis* Pro213. Highly protected amide NH with protection factors higher than 10^6 were found in all three parts of G3P* (Fig. 3a), suggesting that under native conditions N1, N2 and the hinge region form a cooperative unit. Apparently, domain assembly in the final folding reaction provides the major contribution to the stability of the entire protein and determines global unfolding under native conditions. This conclusion remains semi-quantitative, because several NH in N1 and N2 were so strongly protected that they exchanged to a small extent only in our HX experiments. In wild-type G3P, the coupling between the domains in probably weaker, because two of the stabilizing substitutions in G3P* reside in the hinge that holds the domains together.

The N2 domain strongly benefits from the cooperative interactions in folded G3P*. Its equilibrium constant of folding is about 10^4-fold higher compared to the isolated N2 domain (38). In the folded protein, the chain ends of the N2 domain are kept together by the antiparallel β-sheet structure in the hinge, and the corresponding favorable change in entropy accounts most likely for a major part of the strong stabilization of N2. This entropic effect on the N2 domain is present to a large extent already when the hinge is only partially structured, as in the folding intermediate with a *trans* Pro213.

Molecular trace for the Pro213-controlled domain docking in the final folding step

Figure 4a shows the locations of the 48 amide NH for which the protection factors could be determined in the intermediate state. Residues with similarly high protection factors in the intermediate and in the fully folded protein are colored blue. Here, the protection factor in I is already as high as in N or less than 10-fold smaller. These residues are mostly located in the N2 domain and in the part of N1 remote from the hinge. Their local stabilities are thus largely unaffected by the isomeric state of Pro213, which agrees with the finding that most of them show native-like chemical shifts in the intermediate already (Fig. 2d.)

Amide NH with a moderate (10 – 100-fold) increase in protection factor during the domain locking reaction are found in both N1 and N2 (magenta in Fig. 4a). Several of them are near the domain interface or in the core of N2, indicating that the already native-like folded structure of N2 is tightened during Pro213-mediated domain assembly. Similar stabilizations are found in N1 near the tip and the base of the β hairpin that contacts the N2 domain in the folded protein. The amide NH that are more than 100- or even more than 10^3-fold stabilized in response to the Pro213 *trans* → *cis* isomerization are all located in the hinge or in positions in the N1 domain that face the hinge.

Amide protons of a protein backbone are protected from exchange predominantly by H bonding with their acceptors, and increased protection factors indicate that the corresponding H bonds are strengthened. Figure 4b provides an alternative representation of the NH protection results. It highlights directly those hydrogen bonds that become strongly stabilized when Pro213 isomerizes from *trans* to *cis*. They form a chain of interactions that crosses the interface between N1 and the hinge twice. The chain initiates at Val215 in the hinge and bridges over into the N1 domain at Ser16. The protection of the backbone H bond between these two residues improves 650fold upon Pro213 *trans* → *cis* isomerization. H bonding then proceeds via Asn15 to Gly55 and on from Tyr54 to Val45. Its neighboring residues Cys46 and Val44 are engaged in backbone H bonds with Tyr100 and Tyr102, respectively, back in the hinge again. These three inter-domain H bonds also become strongly stabilized in N. The chain of events is completed by the strong stabilization of three H bonds between the two β strands within the hinge. Remarkably, the Cys46-Tyr100 H bonds between N1 and the hinge, as well as the Gly99-Ser208 H bonds in the hinge are stabilized more than 10^3-fold (SI Table 1). We propose that the strong stabilization of the cluster of hydrogen bonds between the residues 44-46 in N1 and the residues 99-102 and 206-208 of the hinge (Fig. 4b) provides the

structural and energetic basis of the domain locking reaction that is triggered by the *trans* → *cis* switching at Pro213.

The strongly stabilized backbone H bonds that are highlighted in Figure 4b trace a molecular path that uses the spatial specificity of H bonds and differences in local stabilities to propagate the signal that is triggered by the Pro213 switch. The cis/trans switching at Pro213 is thus propagated over a C_α-C_α distance between 13.7 Å (to Gly99) and 21.3 Å (to Gln206). At present, we do not know how the Pro213 isomerization signal itself is connected with the anchor points of the chain of H bonds at Val 215 and at Gly99/Ser208, because our technique probes the stabilization of backbone H bonds. It has, however, not escaped our attention that the chain regions between Pro213 and the anchor points of the H bonded clamp are strongly enriched in prolines and in β-branched residues. This increases the chain rigidity and thus helps in propagating the Pro213 isomerization signal to either side of the hinge.

During phage infection, the back reaction, Pro213 *cis* → *trans* isomerization, is used to expose the binding site for its receptor TolA. This activation probably uses a similar mechanism. It reverses the last step of folding and thus weakens the aforementioned cluster of hydrogen bonds between N1 and the hinge region to unlock the domains of G3P. However, the present NMR folding studies had to be performed with a stabilized variant of G3P that contains two substitutions in the hinge, and therefore conclusions regarding the molecular mechanism of page infection remain tentative. Still, we now begin to understand how changes in the isomeric state at a proline residue can be propagated in a specific and directional fashion to other regions during folding or in a folded protein.

Materials and Methods

Expression and purification

G3P* was expressed and purified as described previously (22), using M9 minimal medium containing ^{15}N-NH_4Cl or ^{15}N-NH_4Cl, ^{13}C-glucose and D_2O.

NMR experiments

All NMR-spectra were recorded in 50 mM sodium phosphate pH 7.0 (or pD 6.6) at 298K, containing 10% or 100% D_2O if not stated differently. Spectra were processed using NMRpipe (39) and analysed using NMRview (40). For the backbone resonance assignments, trHNCA, trHNCACB and trHN(CO)CACB spectra of a ^{15}N/^{13}C/^{2}H sample were acquired with a Bruker Avance 700 spectrometer equipped with a cryoprobe. The published backbone

assignments of the N1-domain (25) were confirmed by ^{15}N-NOESY-HSQC and ^{15}N-TOCSY-HSQC spectra with a Bruker Avance 600 spectrometer.

Amide HX of native G3P* was followed by measuring 58 ^{15}N-TROSY-HSQC spectra over five days with a Bruker Avance 800 spectrometer equipped with a cryoprobe. Amide HX of native N1 was followed by measuring 100 ^{15}N-HSQC spectra within three days with a Bruker Avance 600 spectrometer. Both reactions were started by dissolving lyophilized protein in D_2O buffer. Error bars in Figure 3a and 3c result from the fit of single exponential function to the NMR intensities. The EX2 exchange regime was verified for G3P* by repeating the experiment at pH8 (see SI Figure 6).

For real-time NMR experiments, refolding of ^{15}N G3P* was started by a fourfold dilution of the unfolded protein (in 4.0 M GdmCl, 50 mM sodium phosphate, pH 7.0) to a final GdmCl concentration of 1.0 M in the same buffer. After a dead time of 30 min, two identical ^{15}N-TROSY-HSQC experiments were acquired by a Bruker Avance 800 spectrometer equipped with a cryoprobe. Both spectra were measured for 8 h, the first (kinetic) spectrum during the folding reaction, and the second (reference) spectrum directly after the completion of refolding. Then the kinetic spectrum was subtracted from the reference spectrum to obtain the difference spectrum (31). The entire experiment was reproducibly repeated once.

For the competition experiment between refolding and amide HX, refolding of ^{15}N G3P was started by a fourfold dilution of the unfolded protein (in 4.0 M GdmCl, 50 mM sodium phosphate, pH 7.0) to a final GdmCl concentration of 1.0 M in the same buffer. All buffers contained 100% H_2O. After 5 min, when the refolding reactions of the individual domains were complete, the solvent was exchanged to a D_2O-containing buffer without GdmCl using a NAP10 gel filtration column (GE Healthcare, Uppsala, Sweden). Amide HX was allowed to proceed for 10 h at 298 K and then an 8h ^{15}N-TROSY-HSQC spectrum was measured at a Bruker Avance 900 spectrometer. For referencing, an identical amide HX experiment was performed starting with the same amount of folded G3P* in 1M GdmCl. After the amide HX experiments, the protein concentrations in the NMR tubes were determined by absorption and used for the comparison of the NMR intensities observed in the two spectra after the exchange experiments. The calculation of the protection factors for the intermediate I according to eq. 1 are based on end point analyses. With the current signal-to-noise ratio of the spectra, *S* values between 0.99 and 0.03 can be used for a quantitative analysis. The lower limit is additionally determined by the presence of residual protons in the solvent (about 1 %). Error bars for Figure 3c were calculated for ln(P) by error propagation, because only differences in orders of magnitude are discussed.

Acknowledgments

We thank Paul Rösch for NMR spectrometer time at 700 and 800 MHz, Hartmut Oschkinat and the FMP Berlin for NMR spectrometer time at 900 MHz, and Christian Löw for very helpful discussions. This research was supported by grants from the Deutsche Forschungsgemeinschaft (BA 1821/5-2 and SCHM 444/17-2) and the European Regional Development Fund (ERDF).

Author contributions

U.W., R.P.J., B.E., K.S. performed the research; U.W., F.X.S. and J.B. designed the research, analyzed data and wrote the manuscript

References

1. Yaffe MB, Schutkowski M, Shen M, Zhou XZ, Stukenberg PT et al. (1997) Sequence-Specific and Phosphorylation-Dependent Proline Isomerization: A Potential Mitotic Regulatory Mechanism. Science 278:1957-1960.
2. Mallis RJ, Brazin KN, Fulton DB, Andreotti AH (2002) Structural characterization of a proline-driven conformational switch within the Itk SH2 domain. Nat. Struct. Biol. 9:900-905.
3. Andreotti AH (2003) Native state proline isomerization: an intrinsic molecular switch. Biochemistry 42:9515-9524.
4. Fischer G, Aumüller T (2003) Regulation of peptide bond cis/trans isomerization by enzyme catalysis and its implication in physiological processes. Rev. Physiol. Biochem. Pharmacol. 148:105-150.
5. Eckert B, Martin A, Balbach J, Schmid FX (2005) Prolyl isomerization as a molecular timer in phage infection. Nat. Struct. Mol. Biol. 12:619-623.
6. Eckert B, Schmid FX (2007) A conformational unfolding reaction activates phage fd for the infection of Escherichia coli. J. Mol. Biol. 373:452-461.
7. Sarkar P, Reichman C, Saleh T, Birge RB, Kalodimos CG (2007) Proline cis-trans isomerization controls autoinhibition of a signaling protein. Mol. Cell. 25:413-426.
8. Lu KP, Finn G, Lee TH, Nicholson LK (2007) Prolyl cis-trans isomerization as a molecular timer. Nat. Chem. Biol. 3:619-629.
9. Brandts JF, Halvorson HR, 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.
10. Schmid FX, Baldwin RL (1978) Acid catalysis of the formation of the slow-folding species of RNase A: evidence that the reaction is proline isomerization. Proc. Natl. Acad. Sci. USA 75:4764-4768.
11. Cook KH, Schmid FX, Baldwin RL (1979) Role of proline isomerization in folding of ribonuclease A at low temperatures. Proc. Natl. Acad. Sci. USA 76:6157-6161.
12. Schmid FX (1982) Proline isomerization in unfolded ribonuclease A. The equilibrium between fast-folding and slow-folding species is independent of temperature. Eur. J. Biochem. 128:77-80.
13. Balbach J, Schmid FX (2000) in Mechanisms of Protein Folding, ed. Pain, RH (University Press, Oxford), pp. 212-237.
14. Goto Y, Hamaguchi K (1982) Unfolding and refolding of the constant fragment of the immunoglobulin light chain. J. Mol. Biol. 156:891-910.
15. Nall BT (1985) Proline isomerization and protein folding. Comments Mol. Cell. Biophys. 3:123-143.
16. Schreiber G, Fersht AR (1993) The refolding of *cis*- and *trans*-peptidylprolyl isomers of Barstar. Biochemistry 32:11195-11203.
17. Lilie H, Rudolph R, Buchner J (1995) Association of antibody chains at different stages of folding: Prolyl isomerization occurs after formation of quaternary structure. J. Mol. Biol. 248:190-201.
18. Pappenberger G, Bachmann A, Muller R, Aygun H, Engels JW et al. (2003) Kinetic mechanism and catalysis of a native-state prolyl isomerization reaction. J. Mol. Biol. 326:235-246.
19. Schmid FX (2005) in Handbook of protein folding, eds. Kiefhaber, T, Buchner, J (Wiley-VCH, Weinheim), pp. 916-945.

20. Boeke JD, 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.
21. Holliger P, Riechmann L, Williams RL (1999) Crystal structure of the two N-terminal domains of g3p from filamentous phage fd at 1.9 A: evidence for conformational lability. J. Mol. Biol. 288:649-657.
22. Martin A, Schmid FX (2003) The folding mechanism of a two-domain protein: folding kinetics and domain docking of the gene-3 protein of phage fd. J. Mol. Biol. 329:599-610.
23. Levengood SK, Beyer WF, Jr., Webster RE (1991) TolA: a membrane protein involved in colicin uptake contains an extended helical region. Proc. Natl. Acad. Sci. USA 88:5939-5943.
24. Click EM, Webster RE (1997) Filamentous phage infection: required interactions with the TolA protein. J. Bacteriol. 179:6464-6471.
25. Holliger P, Riechmann L (1997) A conserved infection pathway for filamentous bacteriophages is suggested by the structure of the membrane penetration domain of the minor coat protein g3p from phage fd. Structure 5:265-275.
26. Pletneva EV, Sundd M, Fulton DB, Andreotti AH (2006) Molecular details of Itk activation by prolyl isomerization and phospholigand binding: the NMR structure of the Itk SH2 domain bound to a phosphopeptide. J. Mol. Biol. 357:550-561.
27. OuYang B, Pochapsky SS, Dang M, Pochapsky TC (2008) A functional proline switch in cytochrome P450cam. Structure 16:916-923.
28. Min L, Fulton DB, Andreotti AH (2005) A case study of proline isomerization in cell signaling. Front. Biosci. 10:385-397.
29. Aumüller T, Fischer G (2008) Bioactivity of folding intermediates studied by the recovery of enzymatic activity during refolding. J. Mol. Biol. 376:1478-1492.
30. Balbach J, Forge V, Lau WS, van Nuland NAJ, Brew K et al. (1996) Protein folding monitored at individual residues during a two- dimensional NMR experiment. Science 274:1161-1163.
31. Steegborn C, Schneider-Hassloff H, Zeeb M, Balbach J (2000) Cooperativity of a Protein Folding Reaction Probed at Multiple Chain Positions by Real-Time 2D NMR Spectroscopy. Biochemistry 39:7910-7919.
32. Koide S, Dyson HJ, Wright PE (1993) Characterization of a folding intermediate of apoplastocyanin trapped by proline isomerization. Biochemistry 32:12299-12310.
33. Martin A, Schmid FX (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.
34. Martin A, Schmid FX (2003) A proline switch controls folding and domain interactions in the gene-3-protein of the filamentous phage fd. J. Mol. Biol. 331:1131-1140.
35. Balbach J, Steegborn C, Schindler T, Schmid FX (1999) A protein folding intermediate of ribonuclease T1 characterized at high resolution by 1D and 2D real-time NMR spectroscopy. J. Mol. Biol. 285:829-842.
36. Bai Y, Sosnick TR, Mayne L, Englander SW (1995) Protein folding intermediates: Native-state hydrogen exchange. Science 269:192-197.
37. Bai Y, Milne JS, Mayne L, Englander SW (1993) Primary Structure Effects on Peptide Group Hydrogen Exchange. Proteins: Struct. Funct. Gen. 17:75-86.
38. Jakob RP, Schmid FX (2008) Energetic coupling between native-state prolyl isomerization and conformational protein folding. J. Mol. Biol. 377:1560-1575.

39. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J et al. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR 6:277-293.
40. Johnson BA (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. Methods Mol. Biol. 278:313-52.

Figures

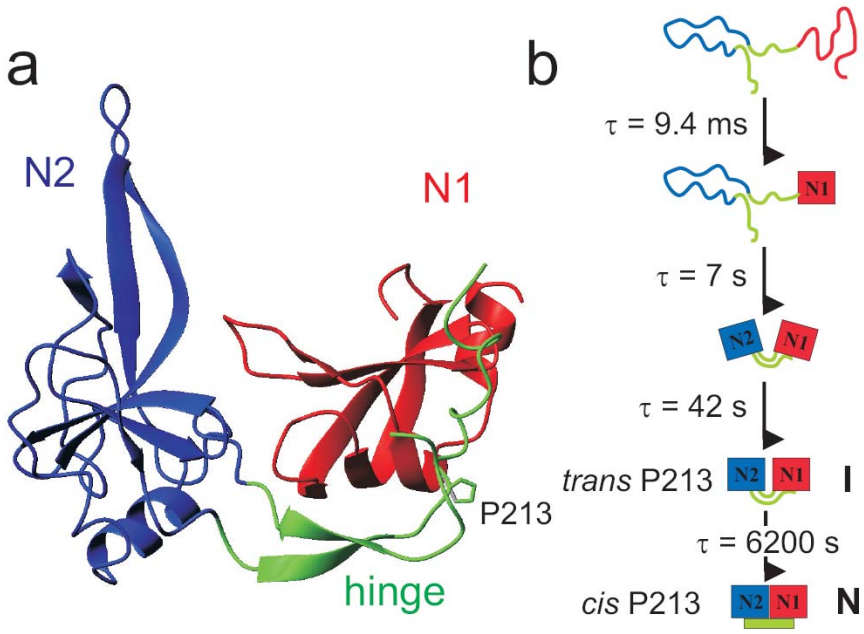


Figure 1 Structure and folding mechanism of the gene-3-protein. (a) Crystal structure of N1-N2 G3P (21). The N1 domain is colored red, the N2 domain blue, and the hinge region green. The proline 213 side chain is indicated. (b) Refolding mechanism of G3P* (22). The domains are colored as in (a). Time constants are given for refolding at 0.5M GdmCl, 25°C. The intermediate (*trans* P213) and the native (*cis* P213) state are labelled with I and N, respectively.

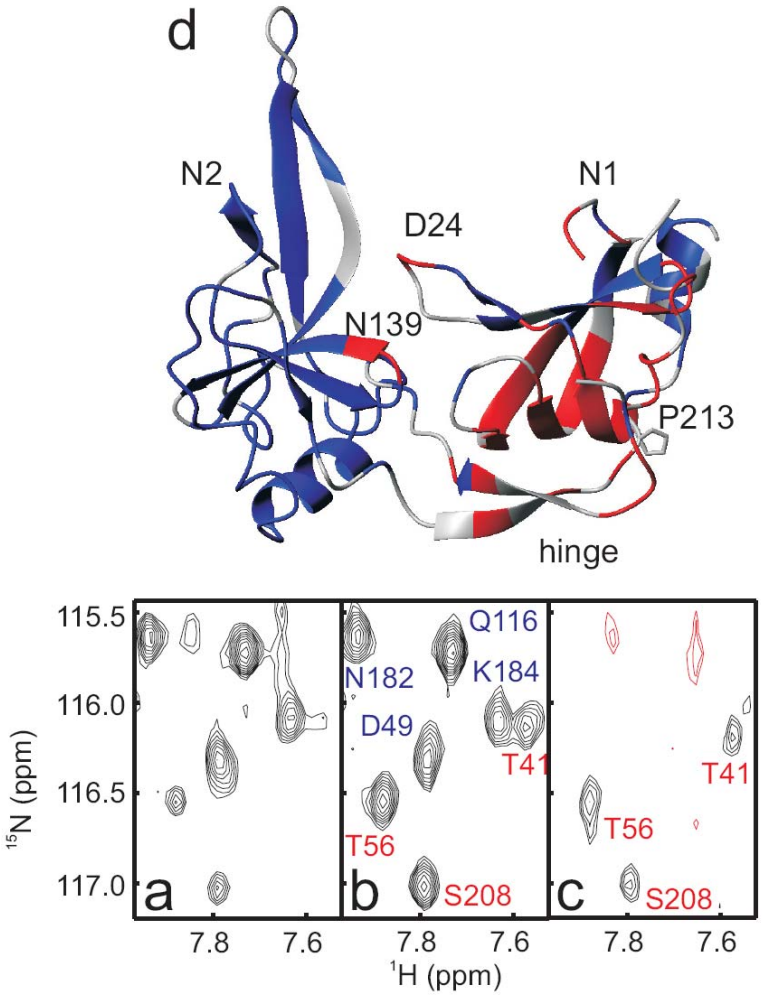


Figure 2 Structural properties of the folding intermediate with a *trans* Pro213 of G3P* derived from real-time NMR spectroscopy. Sections of (a) the kinetic, (b) the reference and (c) the difference ^{15}N-TROSY-HSQC spectrum are plotted. Black contoured cross peaks present in (c) identify amide NH in a non-native environment in I and are labeled red. Cross peaks that occur only in the reference, but not in the difference spectrum, identify amide NH in a native environment in I and are labelled blue in (b). Red cross peaks in (c) result from the long-lived but transiently accumulating intermediate state I. (d) Distribution of amide NH with a native-like (blue) or a non-native-like (red) chemical shift in the folding intermediate. The side chain of Pro213 is indicated.

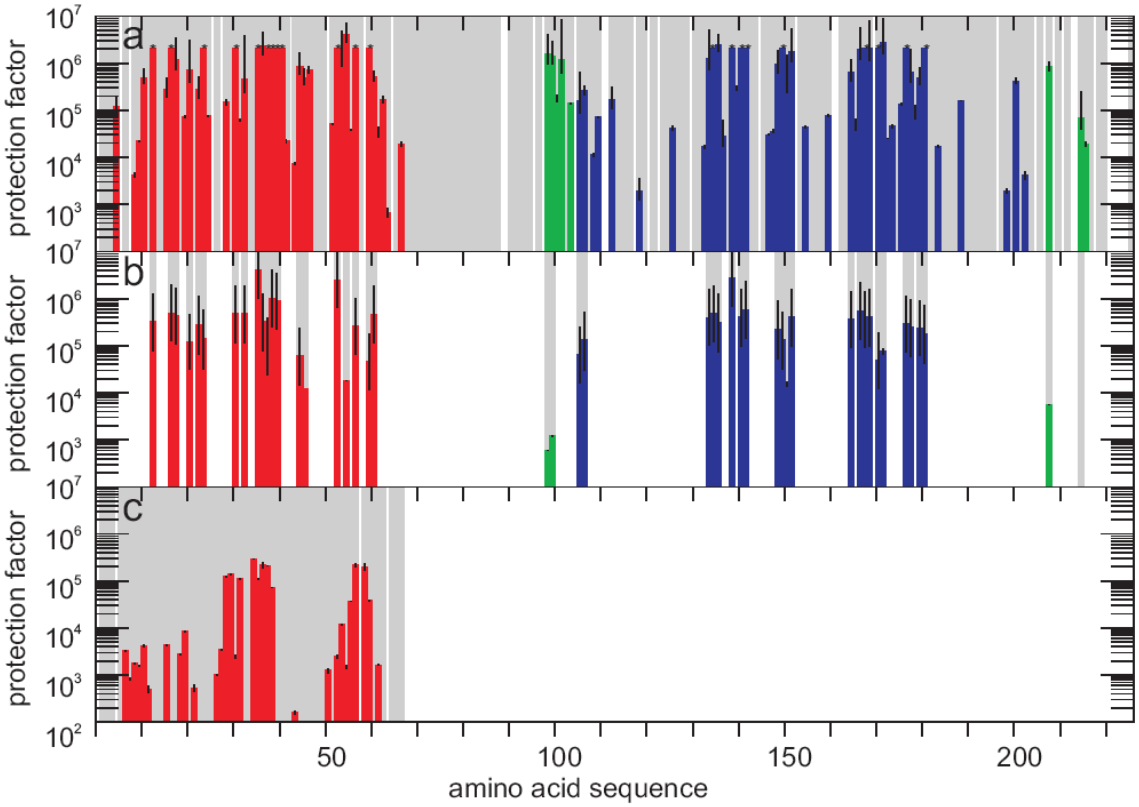


Figure 3 Amide NH protection factors of G3P* in various states. Protection factors are plotted as a function of the sequence position (a) in the native state, (b) in the folding intermediate, and (c) in the isolated N1 domain. Domains are colored as in Fig.1: (red) N1, (blue) N2, (green) hinge region. Amide NH that showed no exchange in (a) were assigned to a value of $2 \cdot 10^6$, which is the highest protection factor that could be determined under the experimental conditions. They are marked by asterisks without error bars. All residues that could be analyzed are marked by a grey background. In (a) and (c) these are all assigned non-prolyl residues. For the calculation of protection factors of the folding intermediate (b) only residues were analyzed that showed no exchange in the native state during the experiment. Black lines indicate error bars.

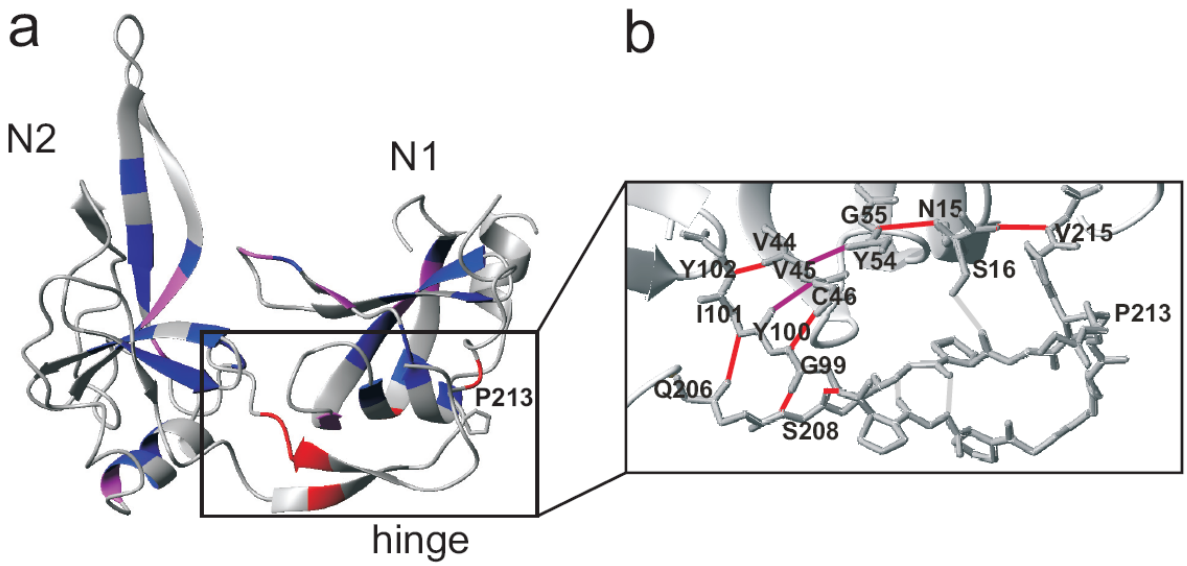


Figure 4 Stability of hydrogen bonds in N and I. (a) Increase in amide protection during the domain docking reaction of G3P*. (Blue) residues with an equal protection in the native protein and in the folding intermediate or a less than tenfold increase in protection. (Magenta) residues with a 10 - to 100-fold increase in protection upon folding. (Red) residues with a more than 100-fold increase in protection against exchange during the domain docking reaction. (b) H-bonding network in the hinge and between the hinge and N1. All residues are shown in backbone representation, for Ser16 and the Pro residues the side chains are shown as well. Residues involved in H-bonds and Pro213 are labelled. H-bonds that gain in NH protection in the I $\rightarrow$ N folding reaction are colored in red for >100-fold and in magenta for a 10-100-fold increase in protection. H-bonds that are not significantly protected in N are colored in grey.

Supporting Information (SI)

SI Table 1: Per residue information of NMR assignment of G3P*, chemical shift information of the intermediate state and protection factors.
For each residue the assignment of HN, N, CA and CB of native G3P* is given in ppm. Residues with native or non-native chemical environments in the intermediate are indicated (native chemical shift: yes or no). Protection factors of the native state (PF native), the intermediate state (PF int.) and the isolated N1 domain (PF iso N1) are given as well.

residue	number	HN	N	CA	CB	native in I	PF nativ	PF int.	PF iso N1
A	1	-	-	47.89	15.68				
E	2	8.32	119.31	51.47	26.73				
T	3	9.01	112.04	55.74	69.05	yes			
V	4	8.70	121.28	64.61	27.58	yes			
E	5	8.26	115.63	56.59	25.02	yes	$1.20•10^{5}$		
S	6			58.12	60.13				
C	7	8.18	123.64	55.57	34.24	yes			$3.24•10^{3}$
L	8			53.18	37.31				$8.43•10^{2}$
A	9	6.90	118.29	48.23	15.64	yes	$4.22•10^{3}$		$1.78•10^{3}$
K	10	7.27	120.77	54.54	29.80	yes	$2.19•10^{4}$		$1.58•10^{3}$
S	11	8.13	115.56	54.88	60.17		$4.88•10^{5}$		$4.19•10^{3}$
H	12	8.32	120.10	51.64	27.59	yes			$4.85•10^{2}$
I	13	8.87	123.01	57.43	37.48	no	$2.00•10^{6}$	$3.23•10^{5}$	
E	14	8.89	131.29	50.45	24.86	no			
N	15	7.97	119.59	48.59	37.90	no			
S	16	9.72	119.89	54.71	59.83	no	$2.74•10^{5}$		$4.30•10^{3}$
F	17	8.56	123.34	50.28	37.48	no	$2.00•10^{6}$	$4.86•10^{5}$	
T	18	9.36	113.43	53.52	66.66	no	$1.20•10^{6}$	$4.17•10^{5}$	
N	19	8.28	119.26	50.96	33.38	yes			$2.74•10^{3}$
V	20	7.32	116.41	57.61	29.12	no	$7.28•10^{4}$		$8.41•10^{3}$
W	21	8.89	123.44	53.18	28.95	yes	$7.03•10^{5}$	$1.18•10^{5}$	
K	22	8.37	121.14	49.42	31.00	yes			$5.17•10^{2}$
D	23	8.43	129.35	51.47	40.89	yes	$2.66•10^{5}$	$2.84•10^{5}$	
D	24	8.53	125.82	53.00	37.31	no	$2.00•10^{6}$	$1.44•10^{5}$	
K	25	7.90	115.19	53.18	29.29	no	$7.51•10^{4}$		
T	26			58.62	65.47				
L	27	7.52	121.21	48.06	38.16				$1.00•10^{3}$
D	28			49.43	38.67				$3.48•10^{3}$
R	29	8.79	116.00	49.42	26.05	yes	$1.46•10^{5}$		$1.26•10^{5}$
Y	30	8.34	122.26	52.15	38.84				$1.36•10^{5}$
A	31	7.93	115.64	48.23	20.25	yes	$2.00•10^{6}$	$4.71•10^{5}$	$2.45•10^{3}$
N	32	9.18	117.61	48.23	37.31	no	$6.10•10^{4}$		$1.11•10^{5}$
Y	33	9.20	119.44	54.71	35.95	no	$4.49•10^{5}$	$4.71•10^{5}$	
E	34	9.03	130.99	53.35	24.34	no			
G	35	8.54	131.86	41.75	-	no			$2.92•10^{5}$
C	36	8.11	120.29	49.76	35.60	yes	$2.00•10^{6}$	$4.00•10^{6}$	$1.12•10^{5}$
L	37	8.12	125.47	50.38	41.06	yes	$2.23•10^{6}$	$3.21•10^{5}$	$2.17•10^{5}$
W	38	9.35	126.37	51.98	29.12	yes	$2.00•10^{6}$	$1.01•10^{5}$	$2.11•10^{5}$
N	39	9.34	118.98	46.52	40.72	no	$2.00•10^{6}$	$1.01•10^{6}$	$7.28•10^{4}$
A	40	7.12	131.06	48.91	12.74	no	$2.00•10^{6}$	$8.77•10^{5}$	
T	41	7.52	116.22	56.25	68.87	no	$2.00•10^{6}$		
G	42	8.17	106.54	41.58	-	no	$2.21•10^{4}$		
V	43			60.00	26.90				
V	44	8.52	130.23	57.10	29.63	no	$7.30•10^{3}$		$1.61•10^{2}$
V	45	8.94	127.70	56.25	30.99	no	$8.47•10^{5}$	$6.04•10^{4}$	
C	46	8.60	122.76	51.30	44.13	no	$4.69•10^{5}$	$1.17•10^{4}$	
T	47	8.53	111.80	58.29	66.31	no	$7.05•10^{5}$		

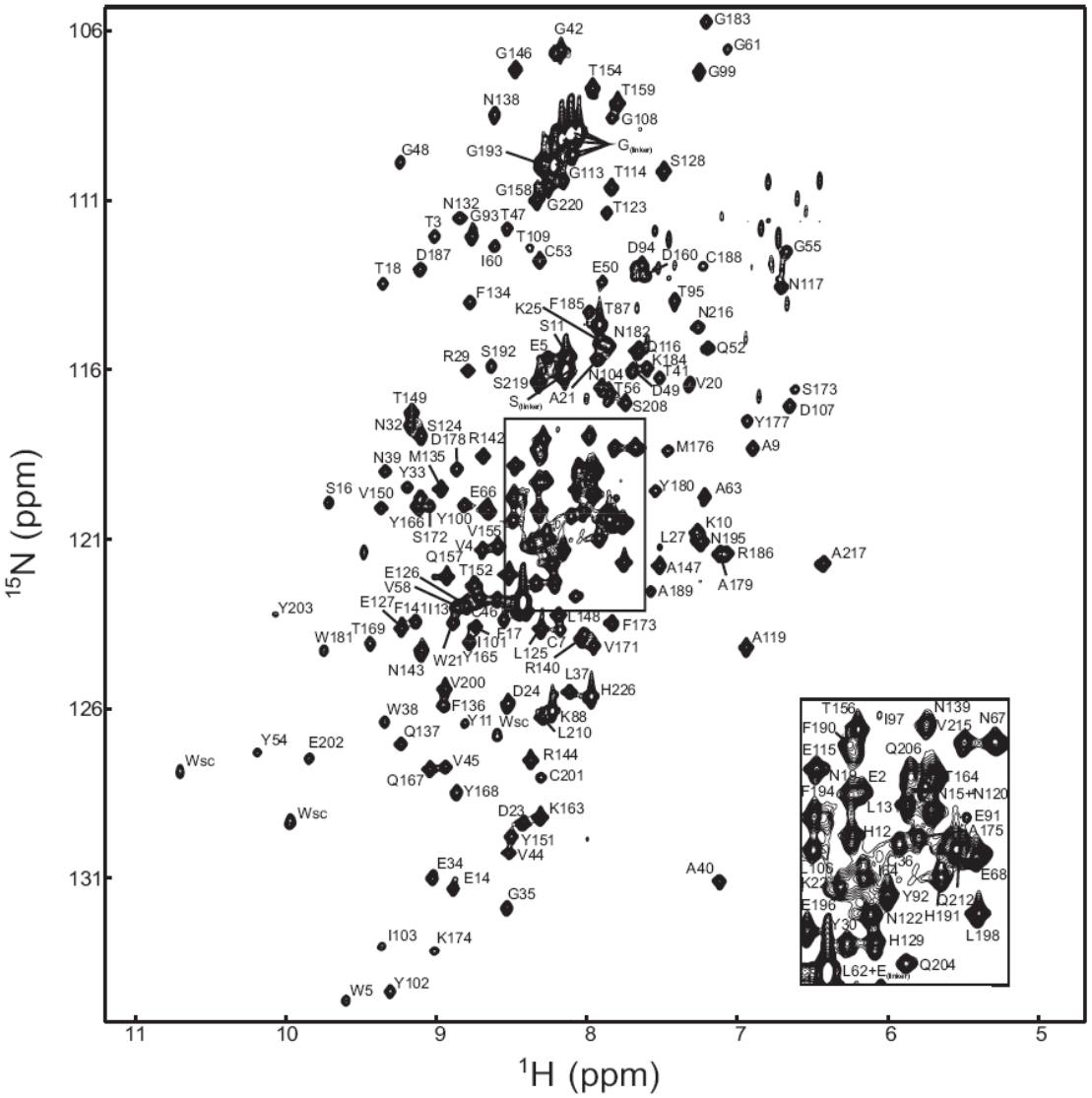
G	48	9.24	109.85	43.28	-				
D	49	7.70	116.00	49.42	36.63	yes			
E	50	7.90	113.37	54.20	23.32				
T	51			60.34	67.00				$1.24 \cdot 10^{3}$
Q	52	7.20	115.33	51.47	31.51	no	$5.02 \cdot 10^{4}$		
C	53	8.32	112.77	53.18	45.50	no	$2.00 \cdot 10^{6}$	$2.36 \cdot 10^{6}$	$2.47 \cdot 10^{3}$
Y	54	10.19	127.26	51.64	37.65	no	$1.46 \cdot 10^{6}$		$1.16 \cdot 10^{4}$
G	55	6.68	112.49	41.23	-		$4.01 \cdot 10^{6}$	$1.80 \cdot 10^{4}$	$1.47 \cdot 10^{3}$
T	56	7.85	116.68	58.98	63.24	no	$3.77 \cdot 10^{4}$		$3.69 \cdot 10^{4}$
W	57	9.60	134.58	54.20	25.53	no	$2.00 \cdot 10^{6}$	$2.59 \cdot 10^{5}$	$2.18 \cdot 10^{5}$
V	58	8.80	122.93	53.69	30.14	no			
P	59	-	-	58.13	26.39				$1.94 \cdot 10^{5}$
I	60	8.61	112.33	56.93	36.46	yes	$2.00 \cdot 10^{6}$	$4.50 \cdot 10^{4}$	$3.80 \cdot 10^{4}$
G	61	7.07	106.51	40.89	-	yes	$5.10 \cdot 10^{5}$	$4.58 \cdot 10^{5}$	
L	62	8.43	122.81	51.77	39.74		$3.23 \cdot 10^{4}$		$1.63 \cdot 10^{3}$
A	63	7.22	119.72	49.08	14.79	no	$1.67 \cdot 10^{5}$		
I	64	8.27	120.97	54.54	34.75	yes	$6.52 \cdot 10^{2}$		
P	65	-	-	60.34	28.27				
E	66	8.66	120.13	53.69	25.71	no			
N	67	7.68	118.27	49.42	34.24	no	$1.89 \cdot 10^{4}$		
E	68	7.76	120.53	53.01	26.39	no			
G	69	8.10 / 8.22	109.63 / 109.92	41.75	-				
G	70	8.06 / 8.11 / 8.17	109.01 / 109.00 / 109.16	41.58	-				
G	71	8.06 / 8.11 / 8.17	109.01 / 109.00 / 109.16	41.58	-				
S	72	8.15	116.04	54.88	60.34				
E	73	8.43	122.81	53.18	26.22				
G	74	8.10 / 8.22	109.63 / 109.92	41.75	-				
G	75	8.06 / 8.11 / 8.17	109.01 / 109.00 / 109.16	41.58	-				
G	76	8.06 / 8.11 / 8.17	109.01 / 109.00 / 109.16	41.58	-				
S	77	8.15	116.04	54.88	60.34				
E	78	8.43	122.81	53.18	26.22				
G	79	8.10 / 8.22	109.63 / 109.92	41.75	-				
G	80	8.06 / 8.11 / 8.17	109.01 / 109.00 / 109.16	41.58	-				
G	81	8.06 / 8.11 / 8.17	109.01 / 109.00 / 109.16	41.58	-				
S	82	8.15	116.04	54.88	60.34				
E	83	8.43	122.81	53.18	26.22				
G	84	8.10 / 8.22	109.63 / 109.92	41.75	-				
G	85	8.06 / 8.11 / 8.17	109.01 / 109.00 / 109.16	41.58	-				
G	86	8.06 / 8.11 / 8.17	109.01 / 109.00 / 109.16	41.58	-				
T	87	7.92	114.64	57.96	66.14				
K	88	8.23	126.03	50.45	28.78				
P	89	-	-						
P	90	-	-	59.32	28.05				
E	91	7.81	119.76	51.64	27.59				
Y	92	8.16	121.31	54.03	37.14	yes			
G	93	8.77	112.05	40.55	-	no			
D	94	7.64	112.90	48.74	38.16	no			

T	95	7.42	113.93	55.05	66.14	no			
P	96	-	-	54.28	26.21				
I	97	8.20	117.73	49.42	35.43				
P	98	-	-	58.81	28.78				
G	99	7.26	107.19	41.06	-	no	$1.54•10^{6}$	$6.00•10^{2}$	
Y	100	8.82	119.97	52.15	38.33	yes	$1.42•10^{6}$	$1.20•10^{3}$	
I	101	8.74	123.56	59.15	33.21	no	$1.78•10^{5}$	$6.10•10^{2}$	
Y	102	9.31	134.31	53.18	34.75		$1.15•10^{6}$	$1.79•10^{3}$	
I	103	9.37	132.99	55.74	34.41				
N	104	7.91	116.50	53.35	37.31	yes	$1.42•10^{5}$		
P	105	-	-	60.34	28.95				
L	106	8.49	120.41	49.08	36.29	yes	$1.53•10^{5}$	$6.41•10^{4}$	
D	107	6.66	117.04	51.47	36.97	yes	$2.58•10^{5}$	$1.33•10^{5}$	
G	108	7.84	108.55	43.28	-	yes			
T	109	8.38	112.38	60.00	70.98	yes	$1.14•10^{4}$		
Y	110	8.81	126.41	51.47	36.12	yes	$7.02•10^{4}$		
P	111	-	-						
P	112	-	-	58.45	27.76				
G	113	8.16	110.38	42.26	-	yes	$1.61•10^{5}$		
T	114	7.84	110.60	56.25	70.07	yes			
E	115	8.48	118.80	55.05	25.37	yes			
Q	116	7.66	115.41	54.41	24.68	yes			
N	117	6.71	113.53	45.33	33.39	yes			
P	118	-	-	59.15	27.93				
A	119	6.94	124.16	48.74	15.47	yes	$1.84•10^{3}$		
N	120	7.97	119.59	47.55	35.43	yes			
P	121	-	-	60.00	28.21				
N	122	8.24	121.68	48.06	34.92	yes			
P	123	-	-	58.98	28.95				
S	124	9.10	117.93	53.35	62.04	yes			
L	125	8.31	123.61	49.94	37.14	yes			
E	126	8.72	122.69	50.62	27.41	yes	$4.11•10^{4}$		
E	127	9.24	123.60	55.22	26.05	yes			
S	128	7.49	110.12	52.49	61.54	yes			
H	129	8.21	122.22	53.69	26.22	yes			
P	130	-	-	58.12	27.76				
L	131	8.07	119.51	51.64	38.67	yes			
N	132	8.85	111.50	51.64	34.24	yes			
T	133	7.87	111.34	58.64	67.85	yes	$1.64•10^{4}$		
F	134	8.78	113.99	50.28	38.33	yes	$1.26•10^{6}$	$3.87•10^{5}$	
M	135	8.97	119.49	49.25	30.83	yes	$2.00•10^{6}$	$4.76•10^{5}$	
F	136	8.96	125.88	53.35	37.48	yes	$2.42•10^{6}$	$3.14•10^{5}$	
Q	137	9.24	127.02	53.09	21.27	yes	$2.69•10^{4}$		
N	138	8.62	108.46	51.30	34.75	no			
N	139	7.99	117.92	48.23	37.82	no	$2.00•10^{6}$	$2.71•10^{6}$	
R	140	8.04	123.92	51.47	29.80	yes	$2.95•10^{5}$		
F	141	9.14	123.39	52.84	39.87	yes	$2.00•10^{6}$	$4.05•10^{5}$	
R	142	8.69	118.52	50.78	29.80	yes	$2.00•10^{6}$	$5.80•10^{5}$	
N	143	9.10	124.22	47.72	36.28	yes			
R	144	8.38	127.49	51.13	26.90	yes			
Q	145			52.94	23.49				
G	146	8.48	107.13	41.92	-	yes			
A	147	7.52	121.75	47.03	14.96	yes	$2.99•10^{4}$		
L	148	8.19	123.21	49.59	39.01	yes	$3.52•10^{4}$		
T	149	9.17	117.23	56.08	69.56	yes	$9.12•10^{5}$	$2.22•10^{5}$	
V	150	9.37	120.05	53.86	32.19	yes	$2.00•10^{6}$	$1.32•10^{5}$	
Y	151	8.50	129.74	52.84	34.75	yes	$4.10•10^{5}$	$1.50•10^{4}$	
T	152	8.75	122.34	55.91	63.41		$1.76•10^{6}$	$4.05•10^{5}$	
G	153			41.06	-				
T	154	7.96	107.67	56.25	68.19				
V	155	8.59	121.19	57.44	32.36	yes	$4.32•10^{4}$		

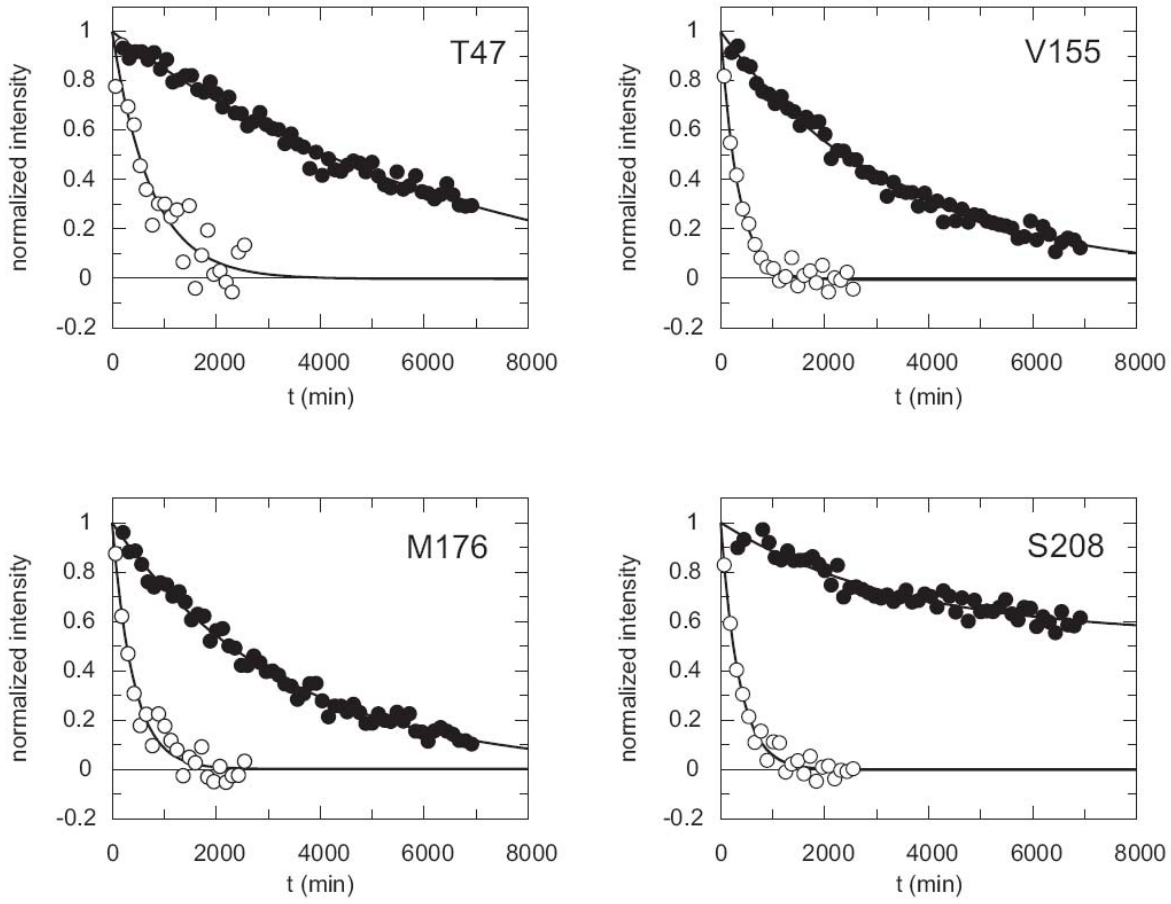
T	156	8.29	118.01	56.25	66.99	yes			
Q	157	8.93	122.08	51.64	29.46	yes			
G	158	8.33	110.96	40.55	-	yes			
T	159	7.80	108.12	57.78	66.83	yes			
D	160	7.62	113.23	52.67	33.39		$7.70 \cdot 10^{4}$		
P	161	-	-						
V	162			58.64	28.27				
K	163	8.31	129.17	51.47	31.17	yes			
T	164	7.96	118.98	58.98	65.97	yes			
Y	165	8.78	124.02	52.15	39.19	yes	$6.22 \cdot 10^{5}$	$3.57 \cdot 10^{5}$	
Y	166	9.12	119.99	53.86	34.75	yes	$4.70 \cdot 10^{4}$		
Q	167	9.05	127.76	52.15	27.76	yes	$1.97 \cdot 10^{6}$	$5.41 \cdot 10^{5}$	
Y	168	8.87	128.43	54.03	36.97	yes	$2.00 \cdot 10^{6}$	$3.57 \cdot 10^{5}$	
T	169	9.44	124.05	57.27	67.17	yes	$2.03 \cdot 10^{6}$	$4.05 \cdot 10^{5}$	
P	170	-	-	60.12	28.38				
V	171	7.96	124.11	60.12	28.88	yes	$2.00 \cdot 10^{6}$	$4.80 \cdot 10^{4}$	
S	172	9.05	120.02	51.98	62.05	yes	$2.74 \cdot 10^{6}$	$7.59 \cdot 10^{4}$	
S	173	6.62	116.56	51.30	61.71	yes	$2.48 \cdot 10^{4}$		
K	174	9.02	133.12	54.37	28.44	yes	$4.51 \cdot 10^{4}$		
A	175	8.02	120.16	51.30	14.79	yes			
M	176	7.46	118.35	53.35	27.59	yes	$1.32 \cdot 10^{5}$		
Y	177	6.94	117.47	60.17	32.70	yes	$2.00 \cdot 10^{6}$	$2.91 \cdot 10^{5}$	
D	178	8.87	118.89	54.20	36.63	yes	$6.42 \cdot 10^{5}$	$2.42 \cdot 10^{5}$	
A	179	7.08	121.40	51.30	13.25	yes	$8.69 \cdot 10^{4}$		
Y	180	7.55	119.55	57.96	34.58	yes	$4.88 \cdot 10^{5}$	$2.26 \cdot 10^{5}$	
W	181	9.75	124.26	55.05	24.68	yes	$2.00 \cdot 10^{6}$	$1.83 \cdot 10^{5}$	
N	182	7.86	115.27	49.94	35.09	yes			
G	183	7.21	105.71	42.60	-	yes			
K	184	7.60	115.93	52.32	26.22	yes	$1.68 \cdot 10^{4}$		
F	185	7.99	114.28	51.64	35.26	yes			
R	186	7.12	121.42	55.74	25.88	yes			
D	187	9.11	113.00	50.79	36.97	yes			
C	188	7.23	112.93	52.49	43.45	yes			
A	189	7.58	122.51	49.59	15.47	yes	$1.60 \cdot 10^{5}$		
F	190	8.33	118.32	53.52	34.92	yes			
H	191	7.92	120.94	53.35	26.05	yes			
S	192	8.64	115.87	53.69	61.71	yes			
G	193	8.29	109.88	40.72	-	yes			
F	194	8.49	119.74	53.86	34.75	yes			
N	195	7.25	121.04	49.08	36.46	yes			
E	196	8.52	122.02	53.18	26.56	yes			
D	197			50.28	37.99				
L	198	7.75	121.66	50.45	39.19	yes			
F	199	7.84	123.45	54.71	36.29	yes	$1.90 \cdot 10^{3}$		
V	200	8.94	125.39	59.32	28.27	yes			
C	201	8.31	128.00	54.20	38.67	yes	$4.13 \cdot 10^{5}$		
E	202	9.85	127.44	55.74	26.22				
Y	203	10.07	123.19	54.54	34.58		$4.02 \cdot 10^{3}$		
Q	204	8.08	122.66	50.79	25.03	yes			
G	205			41.58	-				
Q	206	8.06	118.95	52.15	26.56				
S	207			55.74	60.46				
S	208	7.75	116.97	52.15	59.66	no	$8.38 \cdot 10^{5}$	$5.55 \cdot 10^{3}$	
Y	209			55.74	34.07				
L	210	8.30	126.24	46.87	34.75	no			
P	211	-	-	58.98	28.10				
Q	212	7.85	120.39	51.64	27.58	no			
P	213	-	-						
P	214	-	-	58.67	27.07				
V	215	7.82	118.28	59.75	28.27	no	$6.90 \cdot 10^{4}$	$1.10 \cdot 10^{2}$	
N	216	7.26	114.71	47.72	34.24	no	$1.87 \cdot 10^{4}$		

A	217	6.43	121.68	47.21	14.62	no			
P	218	-	-	59.66	28.05				
S	219	8.32	116.34	54.71	60.51	no			
G	220	8.26	110.54	41.58	-	no			
H	221								
H	222								
H	223								
H	224								
H	225			52.27	26.20				
H	226	7.97	125.61	53.69	26.56	yes			

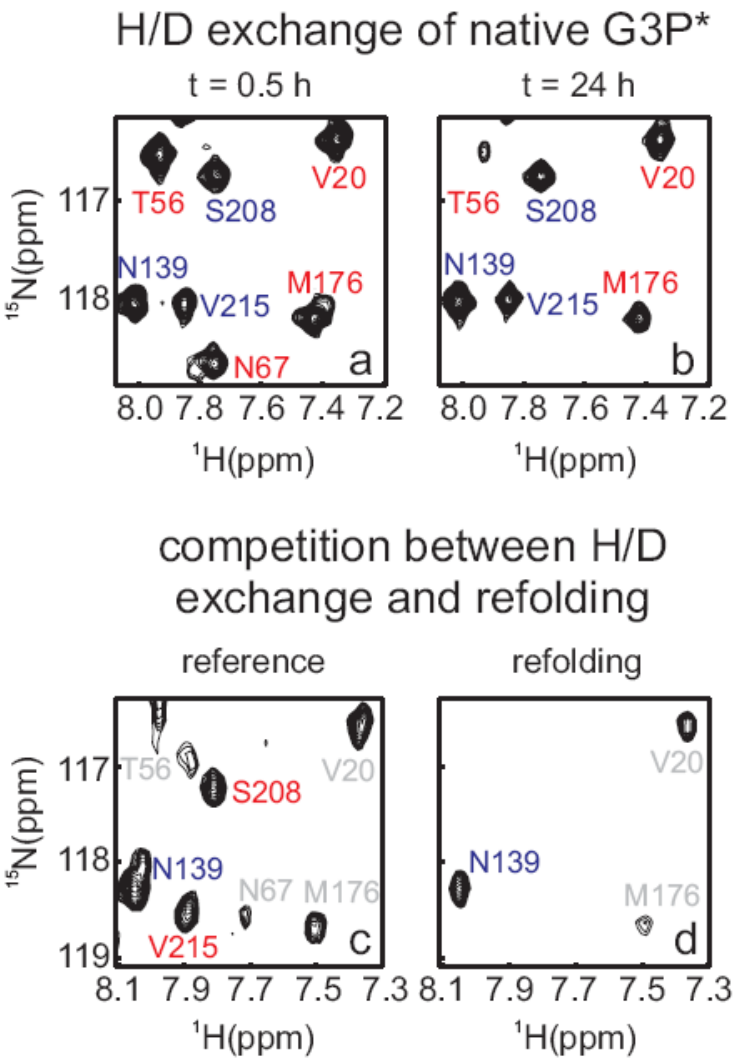
Supporting Information (SI)



SI Figure 5. ^{15}N-TROSY-HSQC spectrum of native G3P* at 25°C, pH 7.0. Amide cross peaks are labelled using the one letter code.



SI Figure 6. Representative amide proton exchange kinetics for four residues of native G3P* at 298K and pH 7.0 (filled circles) and pH 8.0 (open circles). Exchange rates were derived from fitting single exponential functions to the data (solid lines): T47 (N1 domain) $1.44 \cdot 10^{-4}$ min^{-1} (pH7) and $1.33 \cdot 10^{-3}$ min^{-1} (pH8); V155 (N2 domain) $2.99 \cdot 10^{-4}$ min^{-1} (pH7) and $2.96 \cdot 10^{-3}$ min^{-1} (pH8); M176 (N2 domain) $3.04 \cdot 10^{-4}$ min^{-1} (pH7) and $2.47 \cdot 10^{-3}$ min^{-1} (pH8); and S208 (hinge region) $2.29 \cdot 10^{-4}$ min^{-1} (pH7) and $2.83 \cdot 10^{-3}$ min^{-1} (pH8). For an EX2 mechanism, the exchange rate should increase tenfold when the pH is increased from 7.0 to 8.0, as observed.



SI Figure 7. Primary ^{15}N-TROSY-HSQC spectra measured during the H/D exchange experiments with G3P*. (a,b) Spectra measured during the exchange of native G3P*. (a) First spectrum recorded after initiation of the exchange by dissolving G3P* in D_2O. (b) Spectrum after one day of exchange. In (a) and (b) red labels indicate cross peaks from residues, which show exchange during this time and blue labels indicate residues showing no exchange. (c,d) Spectra measured during the competition between H/D exchange and folding. (c) Spectrum recorded in the reference experiment in which native G3P* was subjected to the competition between H/D exchange and refolding without prior unfolding of G3P*. (d) Corresponding spectrum measured during the competition between H/D exchange and refolding. In (c) and (d) red cross peaks indicate residues with low protection against exchange in the intermediate state I and blue cross peaks indicate residues that are highly protected in I. Gray residues that showed partial exchange in the reference experiment already, where not used in the analysis.

Teilarbeit H:

Weininger, U., **Jakob, R.P.**, Kovermann, M., Balbach, J. and Schmid, F.X. Structural and functional analysis of the parvulin domain of PpiD from *E. coli.*, *Manuskript in Vorbereitung*

Running title: The parvulin domain of the prolyl isomerase PpiD

The prolyl isomerase domain of PpiD from *Escherichia coli* shows a parvulin fold but is devoid of catalytic activity

Ulrich Weininger[1†], Roman P. Jakob[†2], Michael Kovermann[1],
Jochen Balbach[1] and Franz X. Schmid[2]

[1] Institut für Physik, Biophysik, and Mitteldeutsches Zentrum für Struktur und Dynamik der Proteine (MZP), Martin-Luther-Universität Halle-Wittenberg, D-06120 Halle(Saale), Germany

[2] Laboratorium für Biochemie, Universität Bayreuth, D-95440 Bayreuth, Germany

† U.W. and R.P.J. contributed equally to this work

please send correspondence regarding manuscript handling to

Franz X. Schmid
Biochemie
Universität Bayreuth
D-95440 Bayreuth, Germany

Tel: ++49 921 553660
Fax: ++49 921 553661
E-mail: fx.schmid@uni-bayreuth.de

Abbreviations used:

PPIase, peptidyl-prolyl *cis*/*trans* isomerase; CD, circular dichroism; T_M, midpoint of a thermal unfolding transition; ΔH_D, van´t Hoff enthalpy of denaturation at T_M; ΔG_D, Gibbs free energy of denaturation; m, cooperativity value of a denaturant-induced equilibrium unfolding transition; τ ($=\lambda^{-1}$), time constant of a folding reaction; k_u, microscopic rate constant and m_u ($= \partial \ln k_u / \partial$[urea]), kinetic m-value of unfolding; k_f, m_f, microscopic rate constant and kinetic m-value of refolding; NOESY, nuclear Overhauser enhancement and exchange spectroscopy; HSQC, hetero single quantum coherence; hNOE, ^{15}N hetero nuclear NOE; HX, hydrogen exchange; NH, amide proton; RCM-T1, reduced and carboxymethylated form of the S54G/P55N double mutant of ribonuclease T_1; N2, N2 domain of the gene 3 protein of phage fd. PpiD*, parvulin domain of PpiD

Abstract

PpiD is a periplasmic folding helper protein of *Escherichia coli*. It consists of an N-terminal helix that anchors PpiD in the inner membrane near the SecYEG translocon, followed by three periplasmic domains. The second domain (residues 265-355) shows homologies to parvulin-like prolyl isomerases. We produced this domain, investigated its catalytic prolyl isomerase function, solved its solution structure by NMR spectroscopy and characterized its interaction with a peptide substrate by a chemical shift analysis. The parvulin domain of PpiD is a well-folded protein with a high stability and a simple two-state folding mechanism. The analysis by NMR spectroscopy resulted in a well-determined structure with a parvulin fold, which resembles most closely the first parvulin domain of the SurA protein, which resides in the periplasm of *E. coli* as well. A previously reported prolyl isomerase activity of PpiD could not be reproduced, even when using a panel of five different peptides in a protease-free assay with improved sensitivity or assays with refolding proteins as substrates. The parvulin domain of PpiD interacts, however, with a proline-containing tetrapeptide, and the binding site, as identified by NMR resonance shift analysis, co-localized with the catalytic sites of other parvulins. Trigger factor in the cytosol and SurA and PpiD in the periplasm of *E. coli* show interesting common design principles. Trigger factor and PpiD are located at sites, where proteins exit from the ribosome or the SecYEG translocon, respectively. The chaperone regions of trigger factor, SurA, and PpiD show similar helical structure, and in trigger factor and in SurA a prolyl isomerase domain is inserted into an extended loop of the chaperone domain. PpiD lacks such a flexibly tethered active prolyl isomerase domain, but shares with SurA an inactive parvulin domain, which apparently forms an integral part of the chaperone entity in both proteins.

Keywords: parvulin, Pin1, prolyl isomerase, rational design, SurA, protein maturation, periplasm, SecYEG

Introduction

Proteins destined to the periplasm of bacteria are synthesized in the cytosol, and most of them are exported in an unfolded form through the Sec translocon [1]. Folding in the periplasm is assisted by folding helper proteins, such as protein disulfide isomerases, peptidyl-prolyl *cis-trans* isomerases, and chaperones [2; 3; 4]. Several of them combine prolyl isomerase domains with chaperone domains, notably FkpA, SurA, and PpiD[5; 6; 7; 8].

The SurA protein participates in the maturation of outer membrane proteins. It consists of an N-terminal chaperone domain followed by two prolyl isomerase domains of the parvulin type[9]. Studies with intact SurA and various truncated forms revealed that only the C-terminal parvulin domain (SurA-P2) is active as a prolyl isomerase [6]. *Sur*A mutants are viable but show defects in the folding of outer-membrane proteins [10; 11; 12; 13; 14]. *Ppi*D was discovered in a genetic screen as a multicopy suppressor in a *surA* deletion strain [15]. The simultaneous deletion of both *ppiD* and *surA* genes was initially reported to confer synthetic lethality [16], a finding that was recently disputed by Justice et al. [17]. Both proteins SurA and PpiD bind to peptidic substrates with SurA being more specific than PpiD [18].

In PpiD, an N-terminal transmembrane helix (residues 13-35) is followed by three domains that face the periplasm [15]. The transmembrane domain anchors PpiD in the inner membrane near the SecYEG translocon [19], the first (residues 36-264) and the third domain (residues 356-623) of PpiD are predicted to contain a high amount of α-helical structure, The second domain (residues 265-355) is a parvulin-like prolyl isomerase domain [18].

Parvulins are ubiquitous globular protein domains of about 100 residues that catalyze the *cis*/*trans* isomerization of prolyl peptide bonds[20; 21], which are often rate-limiting steps during protein folding [21; 22]. They fold into a four-stranded antiparallel β-sheet, surrounded by four α-helices (βα3βαβ2), which is called the parvulin-fold [20]. At present, structures are available for eight parvulins in the Protein Data Bank [9; 20; 23; 24; 25; 26; 27]. The different subtypes differ in the length and the composition of the loop between the strand β1 and the helix α1. Human Pin1 is the most prominent member of the parvulin family. It shows a strong preference for substrates in which proline is preceded by a negatively charged residue, in particular by phosphorylated serine or threonine [9; 27]. This specificity of Pin1 is mediated by several positively charged residues in the β1- α1 binding loop, which, in the SurA-P1 domain, contains mainly hydrophobic residues [9; 20].

Here we investigated structure, the stability, and the catalytic prolyl isomerase function of the parvulin domain of PpiD. Its solution structure was solved by NMR spectroscopy and its interaction with a peptide substrate by a chemical shift analysis. The parvulin domain of PpiD resembles most closely the first parvulin domain of SurA in its structure as well as in the lack of prolyl isomerase activity.

Results and Discussion

The parvulin domain of PpiD (265-355) is a well folded, monomeric and stable protein

We produced a recombinant version of the parvulin domain of PpiD that comprises the residues 265 to 355, and includes a hexa-His tag at the carboxyterminus. This isolated domain, denoted as PpiD*, is well folded and monomeric in solution. The circular dichroism spectrum shows a maximum near 195 nm, as expected for a protein with a high content of secondary structure (data not shown). The thermal unfolding transition is cooperative and shows a midpoint (T_M) at 57.1 °C and an unfolding enthalpy (ΔH_D) of 250 kJ·mol^{-1} (Figure 1A). It is independent of the protein concentration and about 80 % reversible when the unfolded protein is kept in the thermally unfolded form at 70 °C for less than 10 min.

PpiD* contains one Trp residue, and the fluorescence at 340 nm (after excitation at 280 nm) decreases by 70 % between 0 and 6 M urea. The urea-induced unfolding transition (Figure 1B) shows a midpoint at 3.6 M urea and an *m* value of 4.2 kJ·mol^{-1}·M^{-1}. The ΔG_D values extrapolated from urea-induced unfolding ($\Delta G_D^{25°C}$ = 15.3 ± 0.6 kJ·mol^{-1}) and heat-induced unfolding ($\Delta G_D^{25°C}$ = 15.7 ± 0.5 kJ·mol^{-1}) are very similar, which suggests that PpiD* is a well-folded protein.

The folding kinetics of PpiD* are well explained by a two-state mechanism

The equilibrium unfolding of PpiD* is well described by a two-state model. To examine the folding mechanism of this protein, stopped-flow fluorescence spectroscopy was used to measure the kinetics of unfolding and refolding. Both are rapid, monoexponential reactions at all denaturant concentrations. The logarithms of the microscopic rate constants of refolding and unfolding vary linearly with the concentration of urea, which results in a chevron-type profile for the denaturant dependence of the measured rate constant λ (Figure 1C). Linear extrapolation to 0 M urea gave rate constants of k_f = 241 s^{-1} for refolding and k_u = 1.0 s^{-1} for unfolding.

In its folding behavior, PpiD thus differs from the prototypic parvulin Par10. Par10 from *E. coli* is a small single domain protein of only 92 amino acids and the first parvulin to be discovered by Fischer and coworkers [28; 29]. Par10 shows two refolding reactions, a fast one that represents the direct, conformational folding and a dominant slow reaction that is limited in rate by the *trans*-to-*cis* isomerization of the Gly75-Pro76 bond [30]. In PpiD, Pro76 is replaced by Val , which apparently simplifies and accelerates folding.

Figure 1D shows the initial and final fluorescence values of PpiD* as obtained in the stopped-flow experiments. Identical final values were observed when unfolding and refolding was performed under the same conditions (between 3 and 5 M urea), The final values trace the equilibrium unfolding transition of PpiD* (Figure 1D), and the transition midpoint $[urea]_M$ determined from this plot (3.6 M) is identical with the value obtained from the equilibrium unfolding transition in Figure 1B. Moreover, the equilibrium constants calculated from the ratio of the microscopic rate constants (Figure 1C) agree well with those derived from the equilibrium unfolding transition in Figure 1B. These good agreements confirm that the folding of PpiD* is in fact a fully reversible two-state reaction.

The solution structure of PpiD* reveals a parvulin fold

The solution structure of PpiD* was determined by NMR spectroscopy. Resonances of ^{15}N labeled PpiD* were assigned by ^{15}N-NOESY-HSQC, ^{15}N-TOCSY-HSQC and ^{1}H-NOESY spectra. Using 2357 distance restraints and 162 TALOS [31] derived dihedral angle restraints an ensemble of 50 structures was calculated by ARIA 2.0 [32]. Experimental restraints and structure statistics for the ten lowest energy structures are summarized in Table 1. The ensemble is well defined between residues 5 and 94 (Figure 2A). According to the statistics and the quality assessment with PROCHECK-NMR [33] the NMR analysis yielded a family of structures of very high quality, with 95 % of the backbone dihedral angles residing in the most favored regions of the Ramachandran plot.

The PpiD* structure consists of four α-helices (residues 280-292, 296-302, 308-312 and 326-331) and four β-sheets (residues 269-277, 315-320, 339-344 and 349-358) revealing a parvulin fold [20; 24]. as shown in Figure 2B. In its overall structure, PpiD* closely resembles the first parvulin domain of SurA (with a backbone RMSD of 1.5 Å) and human Pin1, the best studied parvulin (Figure 2C). Within the elements of secondary structure, the backbone atoms of PpiD* and of human Pin1 (as taken from the x-ray structure 1pin.pdb) superimpose with a RMSD of 2.0

Å. The four β strands and helices 1-3 are virtually identical, only helix 4 is slightly re-oriented and shortened in PpiD*. The turns and loops connecting the secondary structure elements α2 with α3, α3 with β2, and α4 with β3 are also similar in PpiD* and Pin1, those connecting β1 with α1, α1 with α2, β2 with α4 and β3 with β4 are different. In particular, the extended loop between β1 and α1 and the N-terminal region of helix 1, as present in Pin1, are strongly shortened in PpiD* (Figure 2C). The long loop of 17 residues between β1 and α1 harbors the cluster of positively charged residues, which determine the specificity of Pin1 for phosphorylated Ser/Thr-Pro substrates. The chain regions of PpiD* with large variations between the individual structures of the NMR ensemble include the turn between β3 and β4, the loops between β2 and α4 (321-325) and between α4 and β3 (332-338), as well as helix 4.

PpiD is a rigid molecule with a defined structural core

The backbone dynamics of PpiD* on the ps-ns timescale was examined by recording ^{15}N R_1 (Figure 3A) and R_2 (Figure 3B) relaxation rates and hNOE-values (Figure 3C) [34], at two magnetic field strengths (B_0 = 14.1 T and B_0 = 18.8 T). Based on these relaxation data and an axially symmetric model (axes of inertia 1 : 0.98 : 0.84) for overall diffusion of the molecule we calculated order parameters S^2 (Figure 3D), internal correlation times (τ_e), additional exchange terms (R_{ex}) and an overall correlation time (τ_m) of motion of 7.3 ns [35] at 25 °C. Only 33 out of 91 amides required additional R_{ex} contributions to adequately describe the relaxation rates, 46 required τ_e terms. They do not cluster over the sequence. This indicates that ps-ns dynamic is described well by a simple model for local motions and no additional dynamic is taking place on this time range.

This analysis confirms that the turn between β3 and β4 (residues 345-348), the loops between β2 andα4 (residues 321-325) and between α4 and β3 (residues 332-338) and the chain termini show increased dynamics. This is indicated by a drop of the hNOE values (Figure 3C) below 0.7 (especially for turn β3/β4) and a reduced order parameter (S^2) below 0.8 (Figure 3D). The remaining regions of PpiD* are very rigid (S^2 0.8 or higher) on the ps-ns timescale. In its local dynamics, PpiD* resembles the parvulin domain of PrsA of *Staphylococcus aureus* [23]. Slower dynamics on a µs-ms timescale could not be observed by R_2-dispersion experiments [36; 37] (data not shown). The recorded R_{2eff} values show no dependence on the applied field strength by usage of CPMG pulse trains. This is true for all NH vectors.

The protection from H/D exchange [38]. with the solvent of individual amide NH of PpiD* was used to probe the local stability of folded PpiD* at pH 7.0 and 25 °C. 66 NH exchanged with the solvent within the dead time of the experiment, indicating that they show protection factors that are smaller than 100. As expected, all non-hydrogen-bonded regions, the chain termini, and the loops and turns belong to this class of fast exchanging NH. Rapid exchange was also observed for strand β2 and the exposed short helix α3. These structural elements apparently exchange via local unfolding (breathing) and thus do not belong to the cooperatively unfolding core of the molecule. Still, they show a well-ordered structure and do not differ from other elements of structure in ps-ns dynamics.

33 residues exchanged with measurable, slow kinetics. 9 of them showed protection factors between 100 and 1000 (colored light blue in Figure 4B), and 24 were protected 1000-10000fold (colored dark blue in Figure 4B). Most of them cluster around a value of 2000 (Figure 4a) leading to a ΔG_D of about 19 kJ/mol, which is 3-4 kJ/mol higher than the overall value obtained from the urea and temperature induced unfolding transitions (Figure 1).

Residues with high protection factors are found predominantly in the central β-strands β1 (residues 269-277) and β4 (residues 349-358), and in helix α1 (residues 280-292), which packs against the center of the β sheet. These elements of secondary structures apparently form a molecular core, which is disrupted only in the course of global unfolding. Several highly protected amide NH were also found in strand β3 (residues 338-340 and 343) and helix 2 (residues 298-301). Both are adjacent to the structural core of PpiD* defined by β1, β4 and α1. With protection factors < 500, helix 4 is less protected and, apparently, not linked with this structural core in terms of its exchange dynamics.

In summary, the elements of secondary structure of PpiD* are well defined and superimpose well with the corresponding structures in the homolog Pin1. The backbone of PpiD* is rigid on the ps-ns timescale. The peripheral chain regions can undergo local unfolding reactions, but the structural core of the molecule (formed predominantly by helix α1, the strands β1 and β4 and several adjacent areas) lose their structure only during global unfolding.

PpiD* does not catalyze prolyl isomerization in peptides or proteins

To investigate the prolyl isomerase activity of PpiD and its parvulin domain, we used an improved and highly sensitive protease-free assay. For this assay, tetrapeptides were used that

carry an aminobenzoyl (Abz) group at the aminoterminus and a para nitroanilide (pNA) group at the carboxyterminus, They have the general formula Abz-Ala-Xaa-Pro-Phe-pNA, where the position Xaa is occupied by various natural amino acids (Zoldák, G. *et al.*, unpublished). In these peptides, the extent of quenching of the Abz fluorescence by the pNA group is sensitive to the isomeric state of the prolyl bond.

The prolyl-isomerases SlyD and cyclophilin18 from *E. coli* (both at 0.5 µM) strongly accelerated the isomerization of the peptide with a Leu-Pro bond (Figure 5A). In contrast 10 µM PpiD* could not accelerate this reaction (Table S1). Several prolyl isomerases, in particular those from the FKBP family, show high substrate specificities with respect the amino acid at position Xaa before the proline [39]. To examine, whether PpiD* might have a peculiar substrate specificity as well, we employed additional assay peptides, in which the position before proline was occupied by a negatively charged (Glu), a positively charged (Lys), a small (Ala) or a large hydrophobic (Phe) residue. For all these peptides, a catalysis of prolyl isomerization could not be observed for PpiD*, even when its concentration was increased to 20 µM. These data suggest that PpiD* is in fact devoid of prolyl-isomerases activity.

We also examined whether PpiD* is able to catalyze prolyl isomerization in the course of protein folding and measured its effect on the proline-limited folding reactions of two proteins, the reduced and carboxymethylated form of the S54G/P55N variant of ribonuclease T_1 (RCM-T1) [40] and the N2 domain of the gene-3-protein of the phage fd [41]. Again, a catalysis of these reactions could not be detected (Figure 5B). Full length PpiD was equally inactive in the peptide and the protein folding assays (Figure 5B).

PpiD* has the same peptide binding site as other parvulins

The lack of prolyl isomerase activity might originate from impaired binding of PpiD* to its substrate. To examine whether PpiD* is able to bind to proline-containing peptides, we titrated ^{15}N labeled PpiD* with Suc-Ala-Leu-Pro-Phe-pNA, a highly soluble peptide substrate for prolyl isomerases Adding this peptide led to shifts stronger than 0.035 ppm for the NH resonances of 12 residues (Figure 6 a) at a 4:1 excess of the peptide (2.29 mM peptide to 0.59 mM PpiD*). In all cases, the resonances changed their positions continuously, indicating that binding and dissociation occurred in the fast exchange limit of the NMR chemical shift timescale (about 100 ms). Saturation could not be reached in these titration experiments, indicating that the

dissociation constant of the complex between the peptide and PpiD* is in the range of 1 mM. Similar values were observed previously for other parvulins [23] and for cyclophilin18 [26;42;26; 43].

Figure 6B shows the side chains of the 12 residues for which the NH chemical shifts changed by more than 0.035 ppm upon peptide binding (R271, Y272, I274, S309, I324, D326, E327, K329, S346, V347, F349, L350). They all are located along the rim of a groove in the PpiD* structure, which, in the homologs human Par14[20], Pin1At from *Arabidopsis thaliana* [23; 25] and PrsA from *Bacillus subtilis* [23] forms the substrate binding site. The residues that are assumed to participate in substrate binding and catalysis in Pin1 (H59, C113, L121, M130, F134, T152, H157) are also shown in Figure 6B. The side chains of I274, D306, M316, and L335 of PpiD* of PpiD* in fact occupy similar positions as the side chains of the corresponding residues H59, C113, L121 and H157 of Pin1. The chemical nature of these side chains is, however, different, which probably is correlated with the different substrate specificities of the two proteins and the lack of prolyl isomerase of PpiD*

Prolyl isomerase function of PpiD* is not reconstituted by mutations of the active site

In its architecture, the binding site for proline-containing peptides in PpiD* resembles the catalytic prolyl isomerase site of Pin1, but PpiD* is apparently inactive [20; 25]. To examine whether a possibly latent prolyl isomerase function in PpiD* could be activated by mutations towards the Pin1 sequence at the seven sites that have been suggested to be involved in the prolyl isomerase activity of Pin1 (Fig 6B)[27], we replaced the corresponding amino acids of PpiD* with those present in Pin 1, first individually in seven single mutants (I274H, D306C, M316L, P325M, L328F, S345T, L350H) and then in various combinations.

To examine the consequences of these substitutions for the stability of PpiD*, we measured thermal unfolding transitions for all variants (Fig S1A). Two substitutions (M316L, S345T) were found to be stabilizing, three were mildly destabilizing (P325M, L328F, D306C) and two (I274H and L350H) were strongly destabilizing. They decreased the melting temperature T_M by 8.6 and 14.5 degrees, respectively, indicating that these two residues are important for protein stability. This is consistent with results obtained for other parvulins. In PrsA from *B. subtilis* the mutation H122A (corresponding to I274 in PpiD*) leads to an unfolded molecule [23], and in Pin1, mutations at position 157 (corresponding to position L350 in PpiD*) are also strongly destabilizing [44].

The potential prolyl isomerase activities of the seven variants with single substitutions were assayed by using five different tetrapeptide substrates and two proline-limited protein folding reactions. Catalytic activities could not be detected, even when PpiD* concentrations as high as 10 μM were employed (Table S1).

Next, the seven single substitutions were combined in various ways, leading to eight variants with two to seven substitutions (Table S1). All variants were completely folded at 15 °C, where the prolyl isomerase assays were performed and showed circular dichroism spectra similar to the wild-type protein (data not shown). The effect on the stability of the amino acid replacements were almost additive (Figure S1B) and the seven-fold variant showed a T_M value of 40.9 °C, which is 16.4 degrees lower than the T_M of wild-type PpiD*.

All PpiD* variants, including the sevenfold variant, were inactive in our prolyl isomerase assays (Figure S2). Grafting all residues that have been implicated to form the active site of Pin1 onto the conserved scaffold of PpiD* is apparently not sufficient to create a prolyl isomerase activity or to activate a latent activity.

Discussion

The parvulin domain of PpiD (PpiD*) is a well-folded protein with a high stability and a simple two-state folding mechanism. The analysis by NMR spectroscopy resulted in a well determined structure with a parvulin fold, which shows the highest similarity to the first parvulin domain of SurA.

Dartilongue and Raina reported very high prolyl isomerase activities of up to $3 \cdot 10^9$ $M^{-1}s^{-1}$ for PpiD in the classical protease-coupled assay[15]. We were unable to confirm this prolyl isomerase activity for full-length PpiD or the isolated parvulin domain when using a panel of five different peptides in a protease-free direct assay with improved sensitivity and in a protein folding assay with refolding proteins as substrates. The origin for this discrepancy remains unclear. All attempts to activate PpiD* by transplanting seven key residues of the catalytically active homolog Pin1, individually or in various combinations, to PpiD* remained unsuccessful. It is thus unlikely that PpiD* possesses a latent prolyl isomerase function that can be activated by variations in the substrate or mutations in PpiD*. PpiD* interacts, however, with a standard tetrapeptide as used for the assay, and the binding site, as identified by NMR resonance shift analysis, co-localized with the catalytic sites of other parvulins.

The natural function of PpiD is not well characterized. It is ubiquitous in gram-negative bacteria and the residues of the substrate binding site of the parvulin domain are highly conserved (Figure S3). This domain is probably important for the biological function of PpiD. Dartigalongue & Raina (1998) used a screen based on the ability to induce the σ^E response as well as hypersensitivity to novobiocin[15] to identify residues that are critical for PpiD function *in vivo*. Four of the isolated point mutations mapped to the parvulin domain. The NMR analysis of PpiD* provides a structural explanation for the effects of the inactivating mutations G312R und G313R. Both are located in the loop that connects helix α3 with strand β2. These positions are sterically demanding and residues other than glycine probably interfere with the proper folding of PpiD*. The other two positions identified, G347 and I350, are located at the peptide binding site, suggesting that substrate binding to the parvulin domain is important for the function of PpiD.

Originally, PpiD was identified as a multicopy suppressor in a *surA* deletion strain [15], and, in fact, the parvulin domain of PipD shows the closest structural similarity to the first parvulin domain of SurA. SurA consists of three globular domains and a carboxyterminal helix [9]. The N-terminal domain, which is also α-helical, the first parvulin domain and the C-terminal helix form a compact structural entity, which is thought to function as a chaperone. The second parvulin domain protrudes from this entity, to which it is tethered by two long and presumably flexible linkers [9]. It is active as a prolyl isomerase, but apparently dispensable for the function of SurA *in vivo* [6; 9] . The first parvulin domain, which is structurally similar to the parvulin domain of PpiD, is devoid of prolyl isomerase activity as well, both in intact SurA and as an isolated domain. [6] As a part of the chaperone entity, it seems to cooperate with the N-terminal domain of SurA. It was suggested that this domain acts as a binding module for unfolded proteins and confers specificity for unfolded outer-membrane proteins. [6; 12; 45; 46]

The similarity between PpiD and SurA extends also to the helical aminoterminal domains of the two proteins. The 86-191 region of PpiD shares 31 % sequence identity and 50 % similarity with the 65-168 region of SurA, which comprises the major part of the chaperone entity of this protein (residues 21-168) [9]. The 54-85 region of PpiD is less similar to SurA, but it is also predicted to form two short β-sheets and a long α-helix as the corresponding 21-64 region of SurA. The long helix α1 is involved in substrate binding in SurA [9] and possibly also in PpiD. These similarities in sequence and structure suggest that the aminoterminal regions of the two proteins share an identical fold. Together with the also highly conserved inactive parvulin domain, they probably form homologous chaperone entities with similar architecture and function.

The binding interface of the PpiD parvulin domain is nonpolar, and similar to the corresponding binding site of the inactive parvulin domain of SurA. This explains the finding that these domains bind to similar peptide substrates and that binding is sensitive to the presence of Triton X-100. Using a variety of peptide substrate, Stymest and Klappa found, that the interaction with PpiD usually requires longer peptide chains compared with SurA, and that PpiD seems to be less specific than SurA[18]. This difference might reflect the different biological functions of the two proteins. PpiD is anchored in the inner membrane, near the SecYEG translocon[19] and can be crosslinked to translocated proteins after their release into the periplasm. PpiD is thus possibly a periplasmic gatekeeper at the exit site of the Sec translocon and involved in the folding of many exported proteins. SurA is a soluble periplasmic protein and presumably participates in the maturation of proteins destined to the outer membrane[15].

The folding helper proteins trigger factor in the cytosol and SurA and PpiD in the periplasm of *E. coli* show interesting common design principles. Trigger factor is targeted to the exit site for newly synthesized protein chains by a ribosome-binding domain, PpiD is located next to the exit site of the SecYEG translocon by a membrane-anchor helix. The chaperone domains of trigger factor, SurA and, presumably, also PpiD share a common helical fold, and in trigger factor and in SurA a prolyl isomerase domain is inserted into a long loop of the chaperone domain. This prolyl isomerase module is an FKBP domain in trigger factor but a parvulin domain in SurA. PpiD lacks such a flexibly tethered active prolyl isomerase domain, but it shares with SurA an inactive parvulin domain, which apparently forms an integral part of the chaperone entity in both proteins. The functional correlations based on these structural homologies are still not well understood, in particular because individual deletions of the corresponding genes have only marginal consequences for the viability of *E. coli*.

Materials and Methods

Expression and purification of variants of PpiD*

For the expression of the parvulin domain of PpiD* (residues 265-355 of mature PpiD, plus $(Ala_2His)_6$) the gene fragments were PCR-amplified from *E. coli* XL1B. 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) ΔSlyD (gift from B. Eckert). Site-directed mutagenesis of PpiD was performed by Quik-Change (Stratagene).

After lysis of the cells in 50 mM Tris/HCl, 50 mM NaCl, pH 8.0, 40 mM imidazol with a microfluidizer and centrifugation, all PpiD*-variants were found in the supernatant. The proteins were purified by immobilized metal-affinity chromatography on a Ni-NTA column (elution with 250 mM imidazole) and then subjected to size-exclusion chromatography in 100 mM K phosphate, pH 7.0 on a Superdex HiLoad column (GE Healthcare). The protein-containing fractions were pooled and concentrated in Amicon Ultra units (Millipore). Yields were about 5-30 mg/l. Isotopically labeled ^{15}N-NMR-samples were produced using M9 minimal medium containing $^{15}NH_4Cl$ as the source of ^{15}N and supplemented with vitamin mix. The isolated N2 domain variants and RCM-T1 were expressed and purified as described [47; 48; 49].

Equilibrium unfolding transitions

Samples of PpiD* (1.0 μM) were incubated for 1 h at 25 °C in 100 mM K phosphate, pH 7.0 and varying concentrations of urea. The fluorescence of the samples was measured in 1-cm cuvettes at 340 nm (10 nm band width) after excitation at 280 nm (5 nm band width) in a Hitachi F4010 fluorescence spectrometer. The experimental data were analyzed according to a two-state model by assuming that ΔG_D as well as the fluorescence emissions of the folded and the unfolded form depend linearly on the urea concentration. A nonlinear least-squares fit with proportional weighting of the experimental data was used to obtain ΔG_D as a function of the urea concentration [50].

The heat induced unfolding transitions were measured in a Jasco J-600A spectropolarimeter equipped with a PTC 348 WI peltier element at a protein concentration of 4 μM in 100 mM K phosphate, 1 mM EDTA, pH 7.0 at a heating rate of 1 °C/min. The transitions were monitored by the increase of the CD signal at 222 nm with 1 nm band width and 10 mm path length. The

experimental data were analyzed on the basis of the two-state approximation [51], with a heat capacity change ΔC_p of 6000 J mol^{-1} K^{-1} (calculated as described in ref. [52]).

Folding experiments

A DX.17MV stopped flow spectrometer from Applied Photophysics (Leatherhead, UK) was employed to follow the urea-induced unfolding and refolding kinetics. All experiments were performed in 100 mM K phosphate, pH 7.0, at 25 °C The native or the unfolded (in 8.8 M urea) protein was diluted 11-fold with urea solutions of varying concentrations. The kinetics were followed by the change in fluorescence above 320 nm after excitation at 280 nm (10 nm bandwidth) in an observation cell with 2 mm path length. A 0.5-cm cell with acetone was placed between the observation chamber and the photomultiplier to absorb scattered light from the excitation beam. The kinetics were measured at least eight times under identical conditions and averaged.

Prolyl isomerase activity assays

The prolyl isomerase activities of the PpiD* variants were measured by using tetrapeptides with the general formula Abz-Ala-Xaa-Pro-Phe-pNA. They show an aminobenzoyl (Abz) group at the aminoterminus and a 4-nitroanilide (pNA) group at the carboxyterminus, Position Xaa before proline was occupied by Glu, Lys, Ala, or Phe. For the assay, the peptide substrates Abz-Ala-Xaa-Pro-Phe-pNa (3 mM) were dissolved in trifluoroethanol containing 0.5 M LiCl. Under these conditions, about 50 % of the peptide molecules are in the *cis* conformation. Upon 600-fold dilution into aqueous buffer the *cis* content decreases to about 10 %. The kinetics of the decrease in *cis* content was measured by the change in fluorescence at 416 nm after excitation at 316 nm in a Jasco FP-6500 fluorescence spectrophotometer. The assays were carried out in 50 mM Hepes/NaOH, (pH 7.8) at 15 °C. Under these conditions, the *cis*-to-*trans* isomerization of the prolyl bond was a mono-exponential process (Zoldák, G. *et al.*, unpublished). The folding experiments with RCM-T1 and the variants of N2 as substrates were performed as described [48; 53]

NMR measurements and NMR structure calculations

All spectra were measured in 100 mM potassium phosphate, pH 7.0 at 25 °C, including 10% (v/v) D_2O, processed with NMRPipe [54] and analyzed with NMRView [55]. For resonance assignment and structure calculation 3D ^{15}N-NOESY-HSQC (120 ms mixing time), 3D ^{15}N-

TOCSY-HSQC (80 ms mixing time) and 2D $^{1}H^{1}H$-NOESY (120 ms mixing time) spectra of a 5 mM ^{15}N PipD* sample were acquired at a Bruker AvanceIII 800 NMR spectrometer equipped with an inverse triple-resonance cryoprobe. A complete assignment of all ^{15}N resonances and ^{1}H without Lysine H^{ζ}, Methionine H^{ε} and residues of the histidine taq could be achieved. An final ensemble of 50 structures was calculated with ARIA 2.0 [32] using 2357 NOESY derived ambiguous distance restraints and 162 TALOS [31] derived dihedral angle restraints (Table 1). The final 10 lowest energy structures were water refined, their stereochemistry was analyzed with PROCHECK [33] while structure ensembles were aligned according to their secondary structure elements and analyzed using MOLMOL [56]. Finale energies, r.m.s. deviations from the averaged structure and from ideal geometry as well as the Ramachandran analysis are given in Table 1. The backbone superposition of the NMR structure of PpiD* and human Pin1 (1pin.pdb) was performed on the basis of the lowest energy NMR structure og PpiD*.

Dynamics, H/D exchange and substrate binding

Spectra for the analysis of the nanosecond-to-picosecond dynamics of PpiD*, for following the H/D exchange and for substrate binding were acquired at a Bruker Avance III 600 NMR spectrometer equipped with a triple-resonance probe. Measurements of the dynamics were complemented by spectra acquired at the Bruker AvanceIII 800. R_1 rate constants, R_2 rate constants and ^{1}H-^{15}N heteronuclear NOEs [34] were analyzed using MODELFREE [57] and TENSOR 2.0 [58]. We have chosen an axially symmetric model for overall diffusion of the molecule (principal axes of inertia are 1 : 0.98 : 0.84). H/D exchange was started by exchanging the aqueous buffer against D_2O buffer (pD 6.6, pH-meter reading) using a PD MiniTrap G-25 colum (GE Healthcare) and followed by a series of 30 ^{15}N-HSQC spectra. Protection factors were calculated assuming an EX2 exchange mechanism and using exchange rates from reference peptides [38]. Suc-Ala-Leu-Pro-Phe-pNA was titrated in 10 steps from an 8 mM stock solution to 820 µM ^{15}N labeled PpiD* up to a 4:1 excess (2.29 mM peptide, 0.59 mM PpiD*). Changes in the chemical shifts were followed by ^{15}N-HSQC spectra and quantified by calculating the averaged weighted chemical shift $\Delta\delta_{MW}$(ppm) [59]. An cutoff value of $\Delta\delta_{MW}$ (ppm) = 0.035 was used for structural interpretations.

Acknowledgements

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

References

1. Driessen, A. J. & Nouwen, N. (2008). Protein translocation across the bacterial cytoplasmic membrane. *Annu Rev Biochem* **77**, 643-67.
2. Rietsch, A. & Beckwith, J. (1998). The genetics of disulfide bond metabolism. *Annu Rev Genet* **32**, 163-84.
3. Ito, K. & Inaba, K. (2008). The disulfide bond formation (Dsb) system. *Curr Opin Struct Biol* **18**, 450-8.
4. Kleinschmidt, J. H. (2003). Membrane protein folding on the example of outer membrane protein A of Escherichia coli. *Cell Mol Life Sci* **60**, 1547-58.
5. Behrens, S. (2002). Periplasmic chaperones--new structural and functional insights. *Structure* **10**, 1469-71.
6. Behrens, S., Maier, R., de Cock, H., Schmid, F. X. & Gross, C. A. (2001). The SurA periplasmic PPIase lacking its parvulin domains functions in vivo and has chaperone activity. *Embo J* **20**, 285-94.
7. Ramm, K. & Pluckthun, A. (2001). High enzymatic activity and chaperone function are mechanistically related features of the dimeric E. coli peptidyl-prolyl-isomerase FkpA. *J Mol Biol* **310**, 485-98.
8. Ramm, K. & Pluckthun, A. (2000). The periplasmic Escherichia coli peptidylprolyl cis,trans-isomerase FkpA. II. Isomerase-independent chaperone activity in vitro. *J Biol Chem* **275**, 17106-13.
9. Bitto, E. & McKay, D. B. (2002). Crystallographic structure of SurA, a molecular chaperone that facilitates folding of outer membrane porins. *Structure* **10**, 1489-98.
10. Rouviere, P. E. & Gross, C. A. (1996). SurA, a periplasmic protein with peptidyl-prolyl isomerase activity, participates in the assembly of outer membrane porins. *Gene Develop.* **10**, 3170-3182.
11. Lazar, S. W. & Kolter, R. (1996). SurA assists the folding of Escherichia coli outer membrane proteins. *Journal of Bacteriology* **178**, 1770-1773.
12. Bitto, E. & McKay, D. B. (2004). Binding of phage-display-selected peptides to the periplasmic chaperone protein SurA mimics binding of unfolded outer membrane proteins. *FEBS Lett* **568**, 94-8.
13. Lazar, S. W., Almiron, M., Tormo, A. & Kolter, R. (1998). Role of the Escherichia coli SurA protein in stationary-phase survival. *J Bacteriol* **180**, 5704-11.
14. Tormo, A., Almiron, M. & Kolter, R. (1990). surA, an Escherichia coli gene essential for survival in stationary phase. *J Bacteriol* **172**, 4339-47.
15. Dartigalongue, C. & Raina, S. (1998). A new heat-shock gene, ppiD, encodes a peptidyl-prolyl isomerase required for folding of outer membrane proteins in Escherichia coli. *Embo J* **17**, 3968-80.
16. Rizzitello, A. E., Harper, J. R. & Silhavy, T. J. (2001). Genetic evidence for parallel pathways of chaperone activity in the periplasm of Escherichia coli. *J Bacteriol* **183**, 6794-800.
17. Justice, S. S., Hunstad, D. A., Harper, J. R., Duguay, A. R., Pinkner, J. S., Bann, J., Frieden, C., Silhavy, T. J. & Hultgren, S. J. (2005). Periplasmic peptidyl prolyl cis-trans isomerases are not essential for viability, but SurA is required for pilus biogenesis in Escherichia coli. *J Bacteriol* **187**, 7680-6.
18. Stymest, K. H. & Klappa, P. (2008). The periplasmic peptidyl prolyl cis-trans isomerases PpiD and SurA have partially overlapping substrate specificities. *Febs J* **275**, 3470-9.

19. Antonoaea, R., Furst, M., Nishiyama, K. & Muller, M. (2008). The periplasmic chaperone PpiD interacts with secretory proteins exiting from the SecYEG translocon. *Biochemistry* **47**, 5649-56.
20. Sekerina, E., Rahfeld, J. U., Muller, J., Fanghanel, J., Rascher, C., Fischer, G. & Bayer, P. (2000). NMR solution structure of hPar14 reveals similarity to the peptidyl prolyl cis/trans isomerase domain of the mitotic regulator hPin1 but indicates a different functionality of the protein. *J Mol Biol* **301**, 1003-17.
21. Fischer, G., Tradler, T. & Zarnt, T. (1998). The mode of action of peptidyl prolyl cis/trans isomerases in vivo: binding vs. catalysis. *FEBS Lett* **426**, 17-20.
22. Balbach, J. & Schmid, F. X. (2000). Proline isomerization and its catalysis in protein folding. In *Mechanisms of Protein Folding* 2nd edit. (Pain, R. H., ed.), pp. 212-237. University Press, Oxford.
23. Heikkinen, O., Seppala, R., Tossavainen, H., Heikkinen, S., Koskela, H., Permi, P. & Kilpelainen, I. (2009). Solution structure of the parvulin-type PPIase domain of Staphylococcus aureus PrsA - Implications for the catalytic mechanism of parvulins. *BMC Struct Biol* **9**, 17.
24. Kuhlewein, A., Voll, G., Hernandez Alvarez, B., Kessler, H., Fischer, G., Rahfeld, J. U. & Gemmecker, G. (2004). Solution structure of Escherichia coli Par10: The prototypic member of the Parvulin family of peptidyl-prolyl cis/trans isomerases. *Protein Sci* **13**, 2378-87.
25. Landrieu, I., Wieruszeski, J. M., Wintjens, R., Inze, D. & Lippens, G. (2002). Solution structure of the single-domain prolyl cis/trans isomerase PIN1At from Arabidopsis thaliana. *J Mol Biol* **320**, 321-32.
26. Eisenmesser, E. Z., Millet, O., Labeikovsky, W., Korzhnev, D. M., Wolf-Watz, M., Bosco, D. A., Skalicky, J. J., Kay, L. E. & Kern, D. (2005). Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **438**, 117-21.
27. Ranganathan, R., Lu, K. P., Hunter, T. & Noel, J. P. (1997). Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent. *Cell* **89**, 875-86.
28. Rahfeld, J.-U., Schierhorn, A., Mann, K.-H. & Fischer, G. (1994). A novel peptidyl-prolyl cis/trans isomerase from Escherichia coli. *FEBS Letters* **343**, 65-69.
29. Rahfeld, J. U., Rucknagel, K. P., Schelbert, B., Ludwig, B., Hacker, J., Mann, K. & Fischer, G. (1994). Confirmation of the existence of a third family among peptidyl- prolyl cis/trans isomerases - Amino acid sequence and recombinant production of parvulin. *FEBS Lett.* **352**, 180-184.
30. Scholz, C., Rahfeld, J., Fischer, G. & Schmid, F. X. (1997). Catalysis of protein folding by parvulin. *J Mol Biol* **273**, 752-762.
31. Cornilescu, G., Delaglio, F. & Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* **13**, 289-302.
32. Linge, J. P., Habeck, M., Rieping, W. & Nilges, M. (2003). ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics* **19**, 315-6.
33. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477-486.
34. Dayie, K. & Wagner, G. (1994). Relaxation-rate measurements for 15N-1H groups with pulsed-field gradients and preservation of coherence pathways. *J. Magn. Reson. A* **111**, 121-126.

35. Lipari, G. & Szabo, A. (1982). Model-free approach to the interpretation of nuclear magentic resonance relaxation in macromolecules .2. Analysis of experimental results. *J. Am. Chem. Soc.* **104**, 4559-4570.
36. Loria, J. P., Rance, M. & Palmer, A. G., 3rd. (1999). A TROSY CPMG sequence for characterizing chemical exchange in large proteins. *J Biomol NMR* **15**, 151-5.
37. Skrynnikov, N. R., Mulder, F. A., Hon, B., Dahlquist, F. W. & Kay, L. E. (2001). Probing slow time scale dynamics at methyl-containing side chains in proteins by relaxation dispersion NMR measurements: application to methionine residues in a cavity mutant of T4 lysozyme. *J Am Chem Soc* **123**, 4556-66.
38. Bai, Y. W., Milne, J. S., Mayne, L. & Englander, S. W. (1993). Primary Structure Effects on Peptide Group Hydrogen Exchange. *Proteins: Structure Function and Genetics* **17**, 75-86.
39. Stein, R. L. (1993). Mechanism of enzymatic and nonenzymatic prolyl cis-trans isomerization. *Adv Protein Chem* **44**, 1-24.
40. Mücke, M. & Schmid, F. X. (1994). Folding mechanism of ribonuclease T1 in the absence of the disulfide bonds. *Biochemistry* **33**, 14608-14619.
41. Knappe, T. A., Eckert, B., Schaarschmidt, P., Scholz, C. & Schmid, F. X. (2007). Insertion of a chaperone domain converts FKBP12 into a powerful catalyst of protein folding. *J Mol Biol* **368**, 1458-68.
42. Kern, D., Kern, G., Scherer, G., Fischer, G. & Drakenberg, T. (1995). Kinetic analysis of cyclophilin-catalyzed prolyl cis/trans isomerization by dynamic NMR spectroscopy. *Biochemistry* **34**, 13594-602.
43. Eisenmesser, E. Z., Bosco, D. A., Akke, M. & Kern, D. (2002). Enzyme dynamics during catalysis. *Science* **295**, 1520-3.
44. 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.
45. Webb, H. M., Ruddock, L. W., Marchant, R. J., Jonas, K. & Klappa, P. (2001). Interaction of the periplasmic peptidylprolyl cis-trans isomerase SurA with model peptides. The N-terminal region of SurA is essential and sufficient for peptide binding. *J Biol Chem* **276**, 45622-7.
46. Xu, X., Wang, S., Hu, Y. X. & McKay, D. B. (2007). The periplasmic bacterial molecular chaperone SurA adapts its structure to bind peptides in different conformations to assert a sequence preference for aromatic residues. *J Mol Biol* **373**, 367-81.
47. Jakob, R. P. & Schmid, F. X. (2009). Molecular determinants of a native-state prolyl isomerization. *J Mol Biol* **387**, 1017-31.
48. Jakob, R. P. & Schmid, F. X. (2008). Energetic coupling between native-state prolyl isomerization and conformational protein folding. *J Mol Biol* **377**, 1560-75.
49. 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.
50. Santoro, M. M. & Bolen, D. W. (1988). Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry* **27**, 8063-8.
51. Mayr, L. M. & Schmid, F. X. (1993). Kinetic models for unfolding and refolding of ribonuclease T1 with substitution of cis-proline 39 by alanine. *J Mol Biol* **231**, 913-26.
52. Privalov, P. L. & Gill, S. J. (1988). Stability of protein structure and hydrophobic interaction. *Adv Protein Chem* **39**, 191-234.

53. Schmid, F. X., Mayr, L. M., Mucke, M. & Schonbrunner, E. R. (1993). Prolyl isomerases: role in protein folding. *Adv Protein Chem* **44**, 25-66.
54. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277-93.
55. Johnson, B. A. (2004). Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol. Biol.* **278**, 313-52.
56. Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* **14**, 51-55.
57. Palmer, A. G., III, Rance, M. & Wright, P. E. (1991). Intramolecular motions of a zinc finger DNA-binding domain from xfin characterized by proton detected natural abundance C-13 heteronuclear NMR spectroscopy. *J. Am. Chem. Soc.* **113**, 4371-4380.
58. Daragan, V. A., Ilyina, E. E., Fields, C. G., Fields, G. B. & Mayo, K. H. (1997). Backbone and side-chain dynamics of residues in a partially folded beta-sheet peptide from platelet factor-4. *Protein Sci* **6**, 355-63.
59. Grzesiek, S., Stahl, S. J., Wingfield, P. T. & Bax, A. (1996). The CD4 determinant for downregulation by HIV-1 Nef directly binds to Nef. Mapping of the Nef binding surface by NMR. *Biochemistry* **35**, 10256-61.
60. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-8.
61. Thompson, J. E., Venegas, F. D. & Raines, R. T. (1994). Energetics of catalysis by ribonucleases: fate of the 2',3'-cyclic phosphodiester intermediate. *Biochemistry* **33**, 7408-14.

Figures

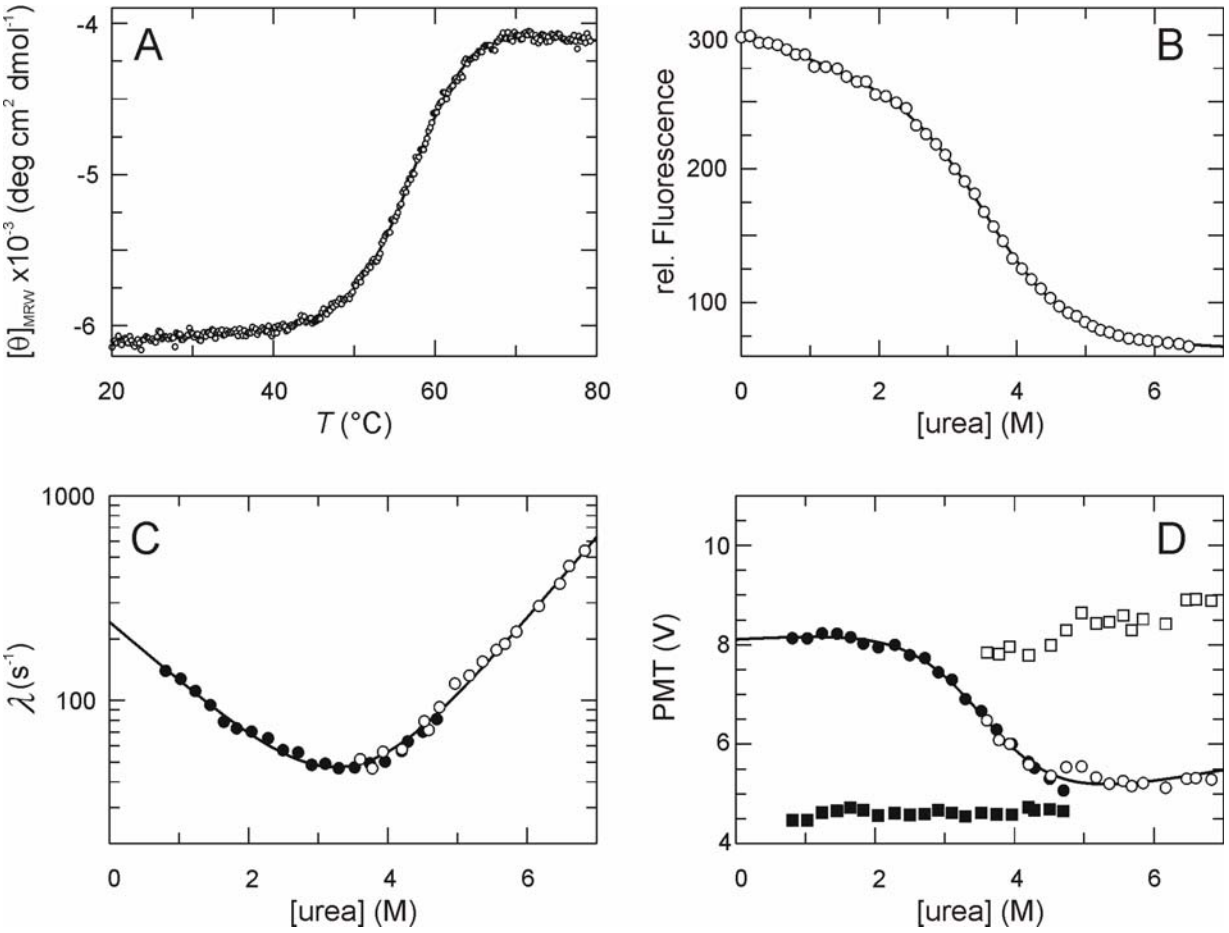


Figure 1: Stability and folding kinetics of PpiD*

(A) Thermal transition of 4 µM protein measured by CD at 222 nm. The two-state analysis (continuous line) results in a T_M = 57.1 ± 0.1 °C and ΔH_D = 253 ± 10 kJ mol^{-1}. (B) Urea-induced unfolding transition of 1 µM protein at 25 °C measured by protein fluorescence at 340 nm after excitation at 280 nm in 100 mM K phosphate (pH 7.0). The two-state analysis (continuous line) gives values of $\Delta G_D^{25°C}$ = 15.3 ± 0.6 kJ·mol^{-1}, m = 4.2 ± 0.3 kJ·mol^{-1}·M^{-1}.and [urea]$_M$ = 3.6 M. (C) Refolding kinetics (filled symbols) and unfolding kinetics (open symbols). The apparent rate constants λ are shown as a function of the urea concentration. A chevron fitted to the experimental data on the basis of a linear two-state model is shown by the continuous line. The results of the analysis are: k_{nu} = 0.99 ± 0.09 s^{-1}, k_{un} = 241 ± 14 s^{-1}, m_{nu} = 0.92 ± 0.02 M^{-1}, m_{un} = − 0.67 ± 0.04 M^{-1}, $\Delta G_D^{25°C}$ = 14 ± 0.4 kJ mol^{-1}, m = 4.0 ± 0.1 kJ mol^{-1}·M^{-1}, [urea]$_M$ = 3.5 M. (D) Initial (■,□) and final (●,○) values of the unfolding (open symbols) and refolding (closed symbols) kinetics as a function of the urea concentration. The continuous line indicates the fit of a transition curve to the final values, giving the following parameters: $\Delta G_D^{25°C}$ = 16.3 ± 2.8 kJ mol^{-1}, m = 4.5 ± 0.7 kJ mol^{-1} M^{-1}, [urea]$_M$ = 3.6 M. The folding kinetics were measured after stopped-flow mixing by the change in fluorescence above 320 nm (excitation at 280 nm) in 100 mM K phosphate (pH 7.0) at 25 °C.

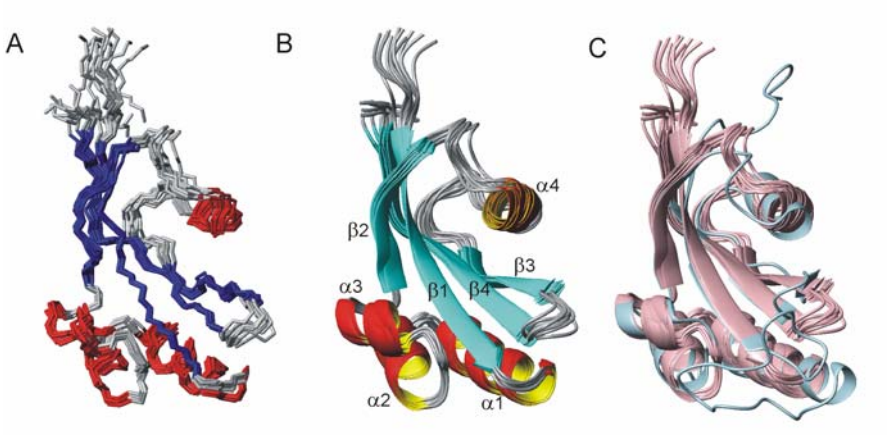


Figure 2: Solution structure of PpiD*
Ensemble of the best ten NMR structures superimposed (A) on the backbone heavy atoms in ordered regions of the protein and (B) ribbon representation in the same orientation for residues 265 - 360. (A) and (B). β strands are colored in blue, α helices in red and coil region in grey. The α helices and β strands are labeled in (B) sequentially along the primary sequence from N- to C-terminus. (C) Ribbon representations of the backbone superposition of the NMR structure of PpiD* colored pink and the crystal structure of human Pin1 (1pin.pdb) colored in cyan. The figures were prepared with MolMol[56].

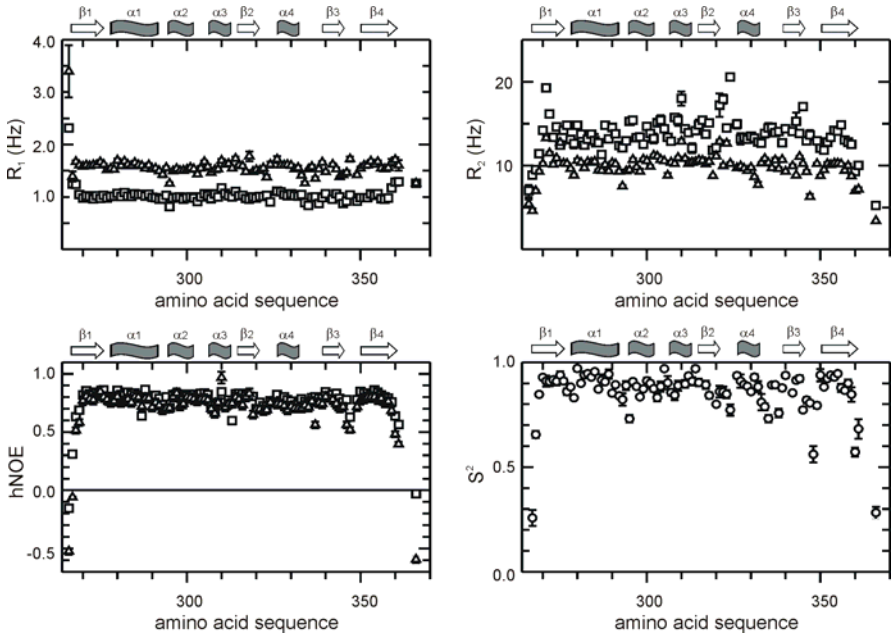


Figure 3: Picoseconds to nanoseconds dynamics of PpiD*
(A) R_1 rates, (B), R_2 rates, (C), ^{15}N hetero nuclear NOEs and order parameter S^2 are plotted as a function of the amino acid sequence. Parameters shown as squares in (A-C) were derived at 800 MHz, those shown as triangles at 600 MHz proton resonance frequency. The order parameters S^2 (D) were calculated using both datasets in MODELFREE [57]. Secondary structure elements are indicated at the top of each panel.

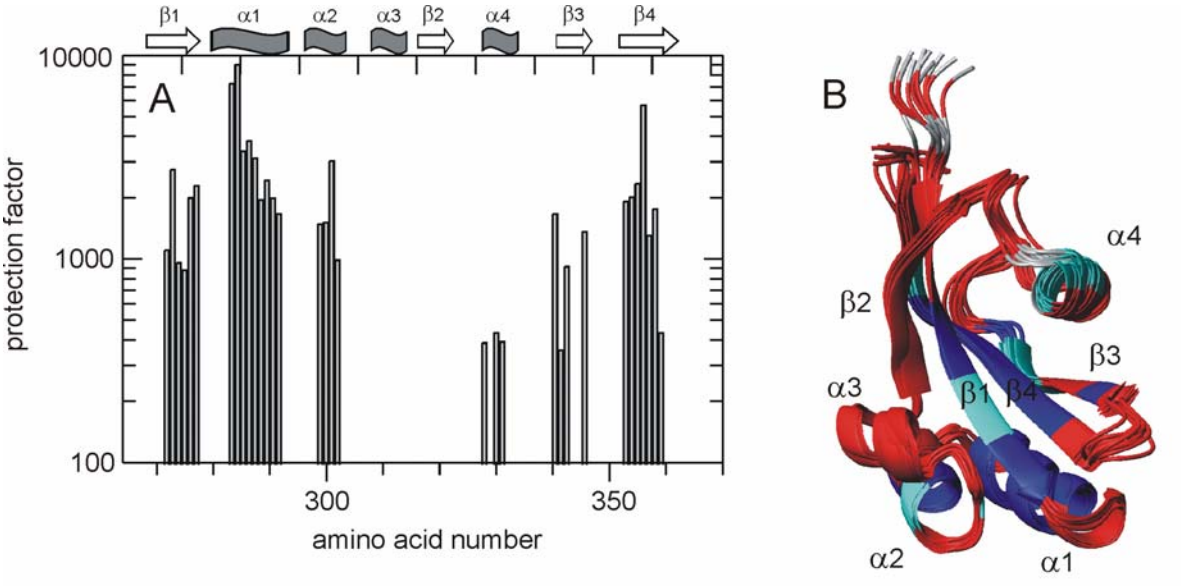


Figure 4: Protection PpiD* from amid NH exchange
(A) The protection factors derived from H/D exchange experiments in 100 mM K phosphate (pH 7.0) at 25 °C are plotted as a function of the amino acid sequence. The secondary structure elements are indicated at the top of the figure. (B) Protection factors higher than 1000 are mapped in blue on the backbone structure, protection factors between 100 and 1000 in cyan, and residues exchanging in the dead time of the experiment are colored in red. Residues lacking an amide proton are colored in grey.

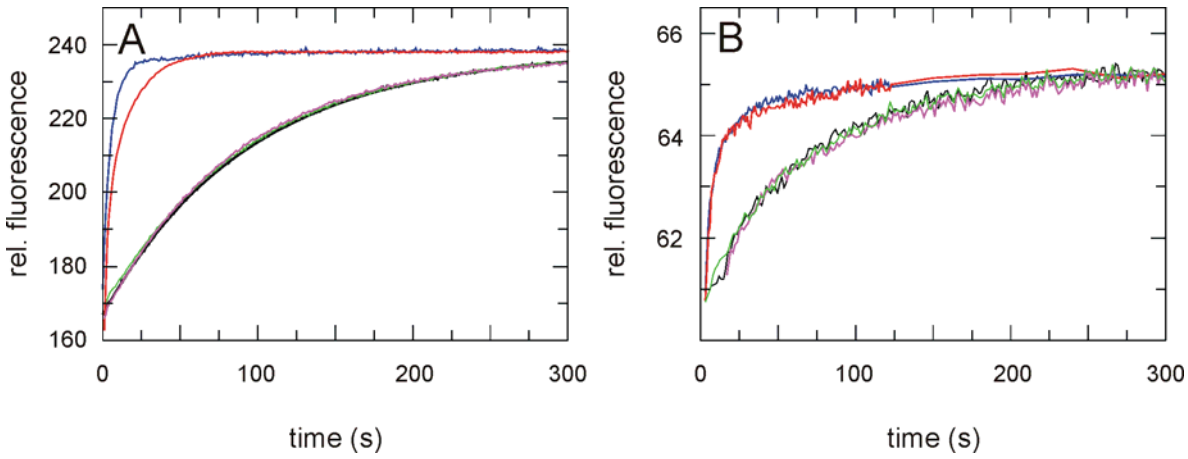


Figure 5: Assays for the prolyl isomerase activity of PpiD
(A) *Cis*/*trans* isomerization of Abz-Ala-Leu-Pro-Phe-pNa followed by fluorescence at 416 nm (5 nm bandwidth) after excitation at 316 nm (3 nm bandwidth). (B) Slow refolding of the N2 domain of the gene-3-protein of phage fd, followed by the change in fluorescence at 340 nm, after excitation at 280 nm. All kinetics were measured at 15 °C in 0.1 M K phosphate, pH 7.0. Kinetics without PPIase are shown in black, in the presence of 10 µM PpiD* in green, in the presence of full length PpiD in magenta, in the presence of 0.5 µM SlyD in blue and in the presence of 0.5 µM cyclophilin in red.

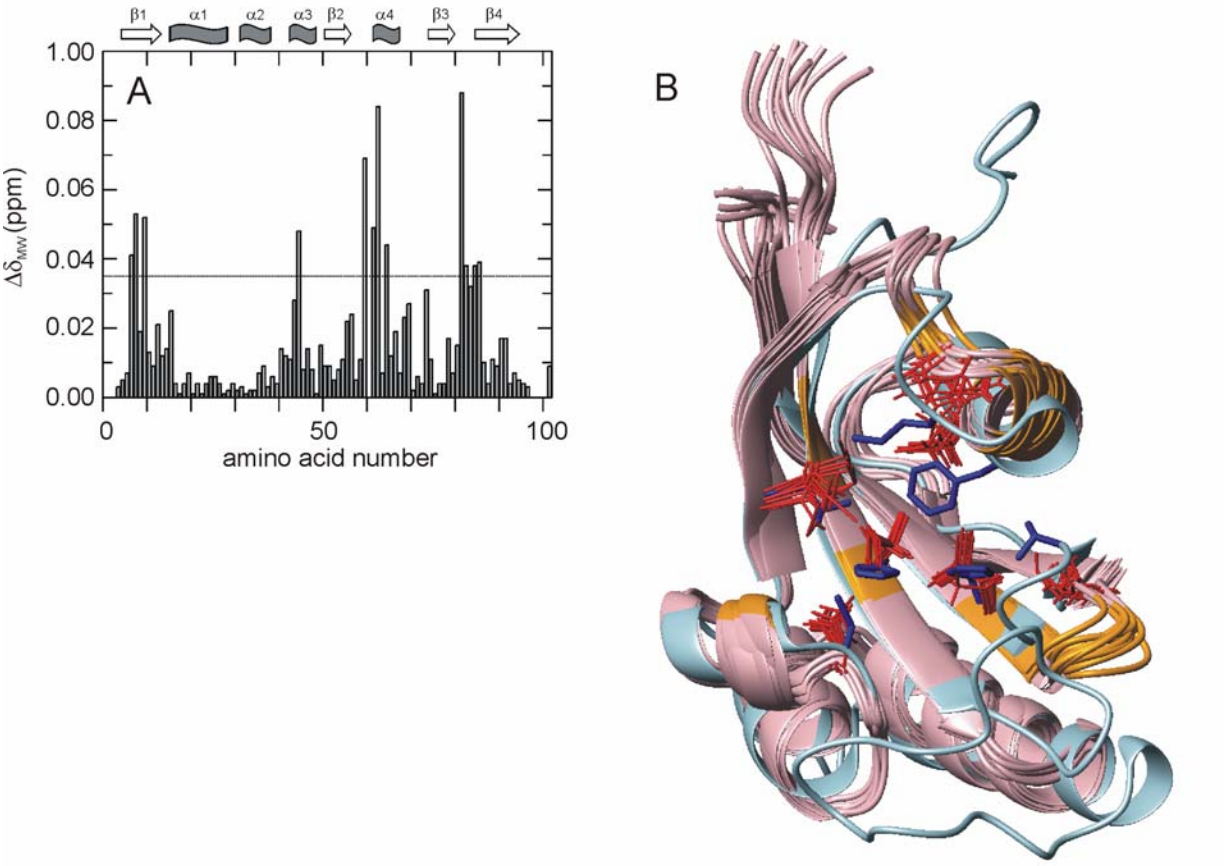


Figure 6: Peptide binding site of PpiD*

(A) The change in the averaged weighted chemical shift, caused by titration with Suc-Ala-Leu-Pro-Phe-pNA) at a ratio of 4:1 (2.29 mM peptide, 0.59 mM PpiD*) at 25 °C, 0.1 M potassium phosphate and pH 7.0 is plotted against the amino acid sequence. Secondary structure elements are indicated on top. (B) Overlay of the solution structure of PpiD* (in pink) and the crystal structure of human Pin1 (1pin.pdb, in cyan) in ribbon presentation. The side chains of putative active site residues of PpiD (I274, D306, M316, P325, L328, S345, L350) are shown in red, the side chains of the active site residues of Pin1 (H59, C113, L121, M130, F134, T152, H157) in blue. The backbone of residues of PpiD* with a $\Delta\delta_{MW}$ larger than 0.035 ppm (R271, Y272, I274, S309, I324, D326, E327, K329, S346, V347, F349, L350) are colored in orange. The same threshold of 0.035 ppm is indicated by the horizontal line in panel (A).

.

Tables

Table 1: Statistics of the structure calculation

experimental restraints	
NOEs	2357
dihedrals	164
NOE violations (Å)	
> 0.5	0.00 ± 0.00
> 0.3	0.00 ± 0.00
> 0.1	14.80 ± 2.95
energies (kcal/mol)	
E_{tot}	160.1 ± 5.8
E_{bond}	5.2 ± 0.3
E_{angle}	68.1 ± 2.4
$E_{improper}$	18.7 ± 1.9
E_{NOE}	28.4 ± 2.2
E_{cdih}	39.6 ± 4.5
rms deviation (Å)	
backbone (res. 5-94)	0.77 ± 0.20
heavy atom (res. 5-94)	1.29 ± 0.19
NOE	0.0155 ± 0.0006
bond	0.00181 ± 0.00005
rms deviation (deg)	
angle	0.394 ± 0.007
improper	0.390 ± 0.019
Ramachandran analysis[a] residue 5-94 (%)	
Most favoured	95.1
Additionally allowed	4.6
Generously allowed	0.3
disallowed	0.0

[a]analysis by PROCHECK [33]

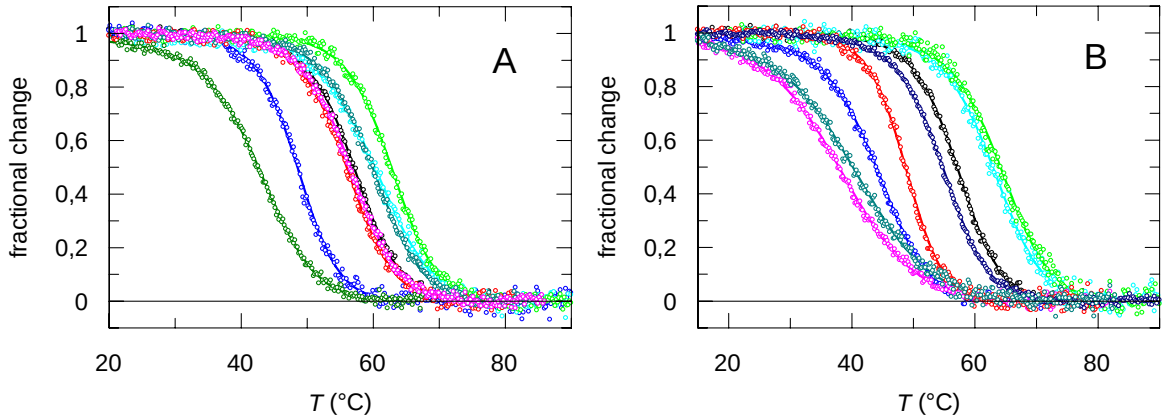


Figure S1: Thermally induced equilibrium unfolding transitions of PpiD* (○) and the (A) PpiD* single mutant variants I273H (○), D305C (○), M315L (○), P324M (○), L327F (○), S344T (○) and L349H (○) and (B) the PpiD* variants with multiple replacements: S344T/L349H (○), D305C/M315L (○), D305C/M315L/S344T (○), I273H/D305C/M315L/S344T/L349H (○), I273H/D305C/M315L/S344T (○), I273H/D305C/M315L/P324M/L327F/S344T (○) and I273H/D305C/M315L/P324M/L327F/S344T/L349H (○). The thermal transitions of 4 µM protein were measured in 100 mM K phosphate (pH 7.0) by CD at 222 nm. The fractional changes were obtained from two-state analyses of the data. The corresponding stability data are given in Table S1.

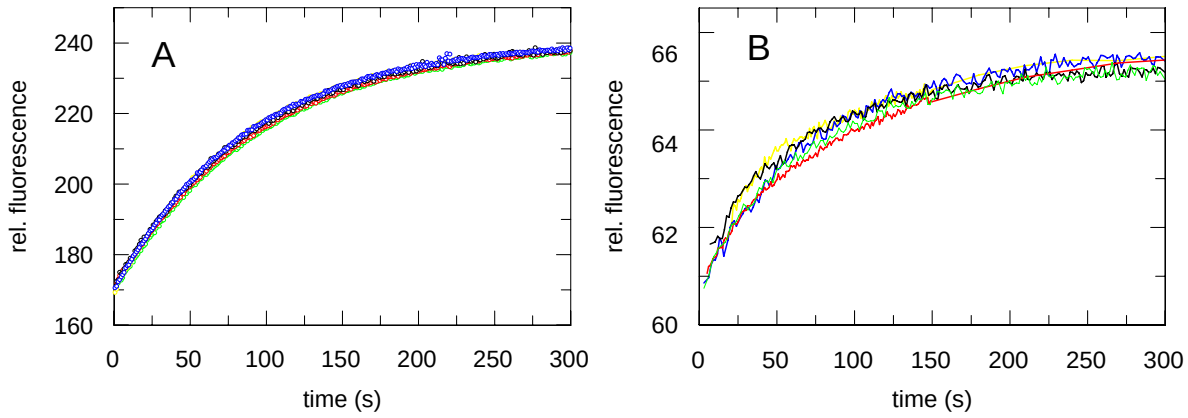
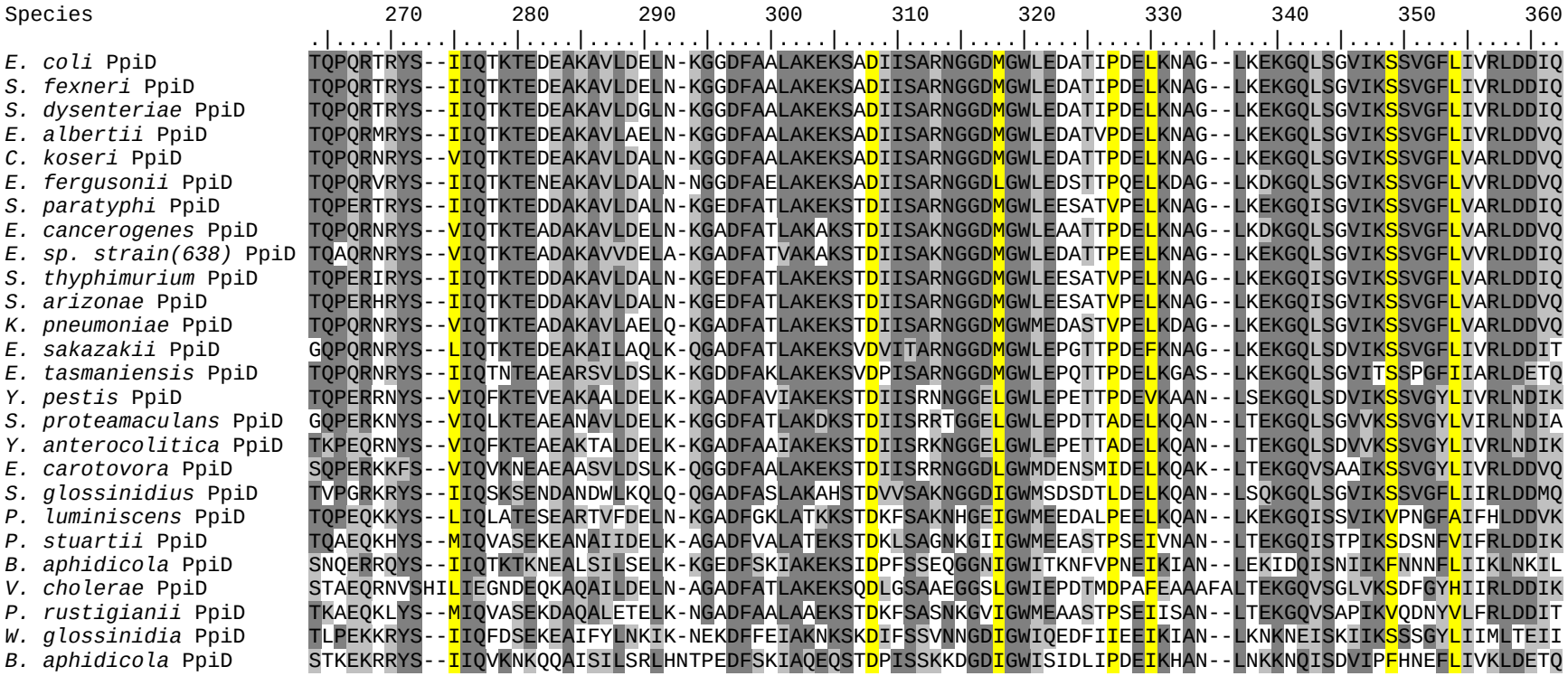


Figure S2:
(A) *Cis*/*trans* isomerization of Abz-Ala-Leu-Pro-Phe-pNa followed by fluorescence at 416 nm (5 nm bandwidth) after excitation at 316 nm (3 nm bandwidth). (B) Slow refolding of the N2 domain of the gene-3-protein of phage fd, followed by the change in fluorescence at 340 nm, after excitation at 280 nm. All kinetics were measured at 15 °C in 0.1 M K phosphate, pH 7.0. Kinetics in the absence of PpiD* are shown in black, those in the presence of 10 µM PpiD D305C/M315L in green, in the presence of PpiD* D305C/M315L/S344T/L349H in blue, in the presence of PpiD* I273H/D305C/M315L/S344T/L349H in blue and in the presence of PpiD* I273H/D305C/M315L/P324M/L327F//S344T/L349H in red.

Figure S3: Sequence alignment of the parvulin domains of PpiD from different species



Sequence alignment of parvulin domain of PpiD variants as performed by CLUSTAL W [60; 61]. Shown are the PpiD orthologues from *E. coli* (SwissProt accession number Q8XE55), from *S. fexneri* (SwissProt accession number Q83M53), from *S. dysenteriae* (SwissProt accession number B3X7A8), from *E. albertii* (SwissProt accession number B1EKP6), and from *C. koseri* (SwissProt accession number A8AK12), *E. fergusonii* (SwissProt accession number B7LMD8), from *S. paratyphi* (SwissProt accession number Q5PFN8), from *E. cancerogenes* (SwissProt accession number B6F8W2), from *E. sp. strain(638)* (SwissProt accession number A4W7B2), from *S. thyphimurium* (SwissProt accession number Q8ZRB8), *S. arizonae* (SwissProt accession number A9MM19), from *K. pneumoniae* (SwissProt accession number A6T5I4), from *E. sakazakii* (SwissProt accession number A7MFI4), from *E. tasmaniensis* (SwissProt accession number B2VHT6), from *Y. pestis* (SwissProt accession number Q7CK29), *S. proteamaculans* (SwissProt accession number A8GAR3), from *Y. anterocolitica* (SwissProt accession number A1JNM7), from *E. carotovora* (SwissProt accession number Q6D823), from *S. glossinidius* (SwissProt accession number Q2NV75), from *P. luminiscens* (SwissProt accession number Q7N0L7), from *P. stuartii* (SwissProt accession number O83369), *B. aphidicola* (SwissProt accession number P0A9K9), from *V. cholerae* (SwissProt accession number B2Q708), from *P. rustigianii* (SwissProt accession number Q8K987), from *W. glossinidia* (SwissProt accession number Q9KQT0) and from *B. aphidicola* (SwissProt accession number B8D807). Amino acids that form the prolyl isomerase site in other parvulins are highlighted in yellow. Identical and conserved amino acids are highlighted in gray of graded intensity.

Table S1. Stability and prolyl isomerase activities of the PpiD* variants

	thermal unfolding		prolyl isomerase assay						
			Abz-Ala-Xaa-Pro-Phe-pNA					refolding experiments	
PpiD variant	T_M (°C)	ΔH_D (kJ mol^{-1})	Ala	Arg	Glu	Leu	Tyr	N2	RCM-T1
wild-type protein	57.1	253	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$
I273H	48.5	271	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$
D305C	60.2	208	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$
M315L	62.8	269	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$
P324M	56.8	239	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$
L327F	56.0	250	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$
S344T	59.9	239	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$
L349H	42.6	187	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$
D305C/M315L	62.9	228	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$
S344T/L349H	44.0	191	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$
D305C/M315L/S344T	64.1	228	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$
D305C/M315L/S344T/L349H	48.7	275	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$
I273H/D305C/M315L/S344T	45.9	152	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$
I273H/D305C/M315L/S344T/L349H	38.4	134	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$
I273H/D305C/M315L/P324M/L327F//S344T	54.9	235	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$
I273H/D305C/M315L/P324M/L327F//S344T/ L349H	40.9	129	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$	$<0.2\cdot10^3$

The stability parameters result from two-state analyses of the thermal unfolding transitions (columns 2–3) For all proteins the melting temperatures (T_M), the van´t Hoff enthalpies of denaturation at T_M (ΔH_D) are given. All stability measurements were performed in 100 mM potassium phosphate, 1 mM EDTA (pH 7.0). The precisions of T_M and ΔH_D are ±0.1 °C and ±10 kJ mol^{-1}. The peptide assay was performed as described in Materials and Methods. Peptides with Ala, Arg, Glu, Leu, or Tyr at position Xaa were employed. The values in columns 4-10 represent k_{cat}/K_M values with the dimension s^{-1}M^{-1}. They all were lower than the threshold of the experiment ($0.2\cdot10^3$ s^{-1}M^{-1}).

Danksagung

Die vorliegenden Arbeit von Januar 2005 bis Juli 2009 am Lehrstuhl Biochemie der Universität Bayreuth unter der Leitung von Herrn Professor Dr. Franz X. Schmid angefertigt. Bei ihm möchte ich mich für sein großes Engagement und für die hervorragende wissenschaftliche Betreuung bedanken. Durch seine zahlreichen Ratschläge, Ideen und gründlichen Analysen hatte er maßgeblichen Anteil an dem Entstehen dieser Arbeit. Darüber hinaus möchte ich mich bei ihm auch für viel Freiheit in der Wahl meiner Forschungsprojekte bedanken, was es mir ermöglichte, viele verschiedene Themengebiete zu ergründen. Er war ein Chef wie man ihn sich wünscht.

Meinen Laborkolleginnen und -kollegen Barbara Eckert, Anne-Juliane Geitner, Insa Kather, Stefan Heiner Lorenz, Christian Löw, Christine Magg, Andreas Martin, Raimund Maier, Stephanie Thoms, Ulrich Weininger, Michael Wunderlich, Markus Zeeb, Gabriel Zoldák gilt mein Dank für die angenehme Laboratmosphäre und die Hilfe bei den kleinen und großen Problemen des Laboralltags. Es hat mir Spaß gemacht mit Euch zusammenzuarbeiten. Ein besonderer Dank gilt Maximilian von Delbrück, Anne-Juliane Geitner, Stefanie Hofmann und Bettina Zierer für ihre fleißige, hervorragende Mitarbeit an meinen Projekten im Rahmen ihrer Bachelor- und Diplomarbeiten.

Bei den NMR-Spektroskopiekern Ulrich Weininger und Prof. Dr. Jochen Balbach bedanke ich mich für die exzellente Zusammenarbeit in vielen gemeinsamen Projekten. Für die Unterstützung zur Bestimmung der Kristallstrukturen von verschiedenen Proteinen möchte ich mich bei Prof. Dr. Holger Dobbek, Dr. Jae-Hun Jeoung, Stefan Knauer und Dr. Berta Martins bedanken.

Für die Bereitstellung von Proteinproben danke ich Dr. Christian Scholz und Prof. Dr. Josef Goris.

Den Fonds der Chemischen Industrie und der GlaxoSmithKline-Stiftung danke ich für die finanzielle Unterstützung während meiner Doktorarbeit.

Meiner Frau Sandra, Johanna Koch und Stefan Lorenz möchte ich für das Korrekturlesen dieser Arbeit danken.

Herzlicher Dank gilt auch meinen Eltern, für ihre Unterstützung während meines Studiums und meiner Doktorarbeit.

Der größte Dank gilt meiner lieben Ehefrau und größten Kritikerin, Sandra. Sie war häufig die erste, mit der ich meine Ideen zu Projekten erörterte und ihr Rat und auch ihre Kritik waren mir immer sehr wichtig. Darüber hinaus war sie maßgeblich an meinen guten wissenschaftlichen Vorträgen beteiligt, die sie in Schrift, Sprache und Layout perfektionierte. In der Zeit mit ihr habe ich immer wieder Kraft für meine Arbeit getankt. Vielen Dank, dass es ein Leben mit dir neben der Wissenschaft gibt.

Hiermit versichere ich, die vorliegende Arbeit selbständig angefertigt und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet zu haben. Ferner erkläre ich, dass ich nicht anderweitig mit oder ohne Erfolg versucht habe, diese Dissertation einzureichen. Ich habe keine gleichartige Doktorprüfung an einer anderen Hochschule endgültig nicht bestanden.

_____

Roman Jakob Bayreuth, den 30.06.2009