Structure and Dynamics of

Unfolded Polypeptide Chains

INAUGURALDISSERTATION

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Andreas Möglich

aus

Bad Homburg vor der Höhe, Deutschland

Basel, 2005

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von

Prof. Dr. Thomas Kiefhaber
Prof. Dr. Joachim Seelig

Basel, den 22.11.2005

Prof. Dr. Hans-Jakob Wirz

(Dekan)

Contents

1. Introduction

1.1 – Proteins and Protein Folding

Proteins constitute the major class of biomolecules living organisms are composed of. In comparison to other biomolecules like nucleic acids, lipids or carbohydrates, proteins fulfill the largest variety of different functions. Most enzymes are proteinaceous; amongst others the cytoskeleton, the immune system and cellular signaling networks are realized by proteins.

Proteins are linear polymers formed from twenty[I] different naturally occurring α-L-amino acids connected by amide (peptide) bonds. Under so-called physiological conditions (i. e. ambient temperature and pressure, low to moderate salt concentrations in aqueous solution) most natural proteins predominantly adopt a defined three-dimensional structure referred to as the native or folded state. Usually, attaining the native structure, in which the linearly connected amino acids are arranged in a defined manner in space, is an absolute prerequisite for protein function. When polypeptides are synthesized at the ribosome they are devoid of defined structure, therefore comprising an ensemble of different conformations. They are said to be in their unfolded or denatured state. Protein folding designates the conformational transition of a protein from its unfolded to its folded state. Proteins exist as an equilibrium mixture between their native and denatured states and continuously fold and unfold *in vivo*. Typical small single-domain proteins unfold and refold on average several times per minute[1]. Thus, protein folding reactions are of significance not only for nascent polypeptide chains but throughout the whole life span of a protein. In recent years the importance of correct folding was further underlined by the discovery of prion diseases, in which misfolded protein aggregates are the agents causative for disease[2].

1.2 – The Protein Folding Reaction

Early on it was realized that the denaturation of proteins is a reversible process[3,4]. The native structure of proteins is fully encoded in their amino acid sequence (primary structure)[5,6], thus allowing proteins to fold in the absence of any other factors. An important consequence of the

---

[I] In some organisms the standard twenty amino acids are supplemented by selenocysteine and/or pyrrolysine.

reversibility of protein folding is the possibility to study the folding reaction *in vitro*. However, especially some larger proteins are known to be prone to misfolding and aggregation. These proteins only fold efficiently in the presence of accessory proteins such as chaperones[7], disulfide bond[8] and peptidyl prolyl isomerases[9,10,11].

Despite the inherent complexity of the underlying processes, protein folding equilibria can be sufficiently described by the conformational interconversion of a few thermodynamic states. A thermodynamic state does not necessarily designate one specific conformation of the protein but rather an ensemble of different ones which equilibrate and interconvert fast on the timescale of the protein folding reaction. Usually, small proteins show cooperative two-state equilibrium unfolding[1]. Denaturation of these proteins can be fully described by a conformational equilibrium between the two thermodynamic states native (N) and unfolded (U). No intermediate structures are populated to any significant extent at equilibrium (eq. (1)).

$$U \rightleftharpoons N \tag{1}$$

The relative population of the individual states is governed by their different free energies (G^0) according to the Boltzmann distribution. For a two-state folding protein the equilibrium constant K of the reaction shown in eq. (1) is related to the free energy of folding, ΔG^0, i. e. the difference in free energy between N and U (eqs. (2) and (3)).

$$K = [\text{N}]/[\text{U}] \tag{2}$$

$$\Delta G^0 = -RT \cdot \ln K \tag{3}$$

A shift of the equilibrium is easily effected by changes in external conditions, leading to the stabilization of one state relative to the other. Commonly used perturbations are temperature and pressure or changes of solvent composition, like e. g. the addition of cosolutes.

Protein folding reactions can be sufficiently described by means of a unidimensional potential energy curve as illustrated in Fig. 1 for a two-state folding protein.
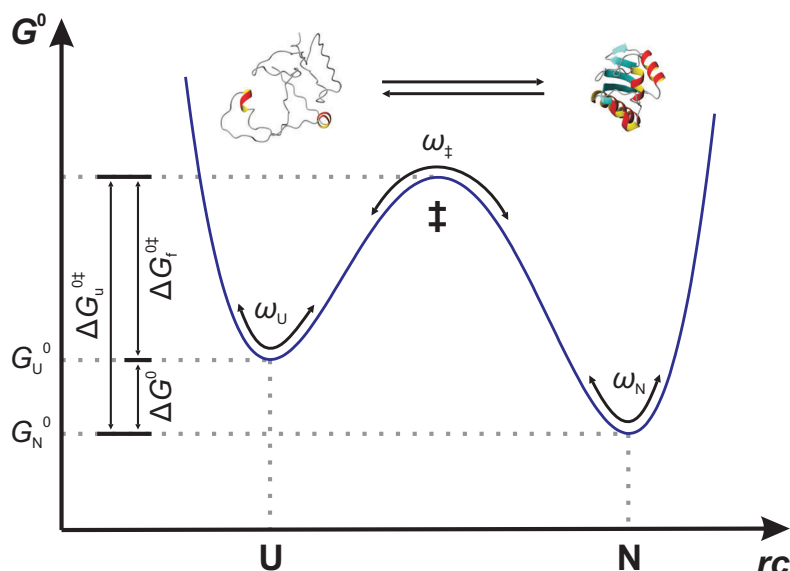
Fig. 1 – Energy surface for a two-state protein folding reaction. The *x*-axis is the reaction coordinate describing the progress of the folding reaction and the *y*-axis shows the free energy G^0. Native (N) and unfolded state (U) are separated by an energy barrier with the transition state (‡) on top. Protein stability is determined by the difference in free energy between N and U, ΔG^0, the folding kinetics by the activation free energies $\Delta G^{0\ddagger}_{f,u}$ and the motional frequencies ω at different positions of the energy curve (cp. chapter 1.3).

The abscissa in such diagrams is called the reaction coordinate and denotes the progress of the folding reaction as quantified by a certain experimental observable, such as e. g. the amount of solvent-accessible surface area. The *y*-axis signifies the potential (free) energies of the different states. Native and unfolded states represent local minima on the energy curve and are separated by an energy barrier. The barrier top is referred to as the transition state (‡).

A full understanding of protein folding equilibria requires detailed knowledge of the participating states (see below). As will be discussed in chapter 1.3, the comprehensive understanding of protein folding kinetics additionally calls for information about the shape and the nature of the potential energy curve and about the motional dynamics at different points of the reaction coordinate of the folding reaction. This thesis mainly deals with the characterization of the static and dynamic properties of unfolded polypeptides. Fast optical spectroscopic techniques (chapters 1.6 and 1.7) were employed to study the dynamics and dimensions of unstructured model peptides and their response towards addition of chemical denaturants (cp. chapter 1.4). Dynamics of unfolded peptide chains are crucial in the early stages of protein folding as will be discussed in chapter 1.5.

1.2.2 – The Native State of Proteins

Native proteins are stabilized by a multitude of noncovalent protein-protein and protein-solvent interactions, most of which are quite weak on their own[12,13]. While electrostatic interactions, hydrogen bonds and van-der-Waals forces contribute to protein stability, a major portion results from the so-called hydrophobic effect or hydrophobic free energy[14]. At ambient temperature the hydrophobic free energy is mainly entropic and is thought to result from the ordering of solvent molecules around the protein surface, which is accompanied by a loss in orientational freedom. Thus, the hydrophobic effect favours the burial of protein surface in order to minimize the protein-solvent interface. Recent work by the group of Baldwin[15] questioned the energetic role of hydrogen bonds. Taking into account the solvation of hydrogen bonds, these researchers[16,17,18,19] concluded that hydrogen bonds in proteins are strongly stabilizing as long as they are solvent-exposed and thus solvated. On the other hand, hydrogen bonds which are buried within the interior of proteins should contribute only weakly to the stability of the native state since they cannot be solvated. Multiple interactions between protein residues well separated in primary structure cause the overall compaction and defined fold of the native protein.

Tremendous insight into the nature of the native state came with the advent of protein X-ray crystallography and the determination of protein structures at atomic resolution[20,21]. Native proteins were found to adopt ordered and defined structures. Several recurring secondary structure elements (α-helices, β-pleated sheets, loops and turns) are arranged in a specific manner in space giving rise to the so-called tertiary structure of proteins[12].

Although native proteins display relatively defined three-dimensional structures, they are far from being static entities but rather comprise ensembles of several interconverting conformers. The temperature factors (Debye-Waller or B-factors) observed in protein crystals hint at varying mobilities in parts of the protein. Nuclear magnetic resonance (NMR) methods[22,23], such as relaxation measurements or hydrogen-deuterium exchange[24] monitored by NMR[25,26,27], are particularly suited to study motions within native proteins. Different kinds of dynamics are observed in native proteins[23,28,29], ranging from bond vibrations (femtosecond timescale) and isomerizations of amino acid sidechains (pico- to nanoseconds) to loop motions (nanoseconds) and local unfolding of protein domains (micro- to milliseconds).

Flexibility of native proteins is important for their function as is clearly seen for enzymes, e. g. for the prolyl isomerase cyclophilin[30,31] and in the case of the induced-fit mechanism of hexokinase[32,33]. Frauenfelder and coworkers[34] demonstrated that the photodissociation of CO

from myoglobin requires motions within the protein. Further evidence is provided by the work of Ansari et al.[35] who found that this photodissociation reaction is drastically slowed down in solutions of increased viscosity.

1.2.3 – The Unfolded State of Proteins

1.2.3.1 – Experimental Studies on Unfolded Proteins

Although the same types of interactions as in the native state also take effect in the denatured state of proteins, unfolded proteins are largely devoid of defined structure[36,37]. Under denaturing conditions intramolecular interactions between protein residues are outweighed by protein-solvent interactions. Consequently, unfolded proteins display large conformational heterogeneity and are not readily amenable to high-resolution structure determination. Accordingly, other techniques had to provide some insight into the properties of denatured proteins. Scattering methods (small-angle X-ray or neutron scattering)[38,39], pulsed-field gradient NMR[40] and in earlier days measurements of the intrinsic viscosity of proteins[41] were used to assess the average dimensions of unfolded proteins. The radius of gyration, R_g, was found to be a power function of protein size n (eq. (4))[38,39].

$$R_g = R_{g0} \cdot n^v \tag{4}$$

In two studies[38,39] values of 0.50 ± 0.02 and 0.60 ± 0.03 were found for the coefficient v. Both results are in agreement with expectations for polymers with no defined structure, referred to as random coils. While the lower value found for v implies that the excluded volume effect has a negligible influence on the dimensions of unfolded proteins, the higher value would mean that there is a considerable effect (cp. chapter 1.2.3.2).

On the other hand several findings at first glance contradicted the random-coil behaviour. NMR measurements revealed that in many unfolded proteins local ordered structure persists even under otherwise harshly denaturing conditions[42,43,44]. Evidence for residual structure in the unfolded state also comes from the work of Sánchez et al.[45] who observed changes in the position of the unfolded state of proteins along the reaction coordinate upon mutation of amino acid sidechains (cp. Fig. 1). Further, several studies illustrated that the backbone of unstructured peptides preferentially adopts conformations corresponding to the polyproline II helix[46,47,48,49,50,51]. These findings can still be reconciled with the random coil behaviour deducted from the scattering experiments. It is important to realize that both the residual

structure and the bias towards specific backbone conformations have consequences with mainly local range. In addition, these structures might be transient and only present in a small fraction of the ensemble of different conformations. In this way the global properties of the polypeptide could still correspond to that of random coils.

The dynamic properties of unfolded proteins were examined by different techniques[52], mainly fast relaxation methods. Recently, optical spectroscopic methods were introduced to study chain dynamics in unstructured peptides[53,54] which will be discussed in more detail in chapter 1.5. These studies revealed that unstructured polypeptides undergo conformational rearrangements on the nanosecond timescale. Formation of secondary structure is thought to start with a nucleation step in which the native conformation is adopted in parts of the structure element, e. g. an α-helical turn. In consecutive propagation reactions this nucleus extends to finally from the fully folded secondary structure element. Relaxation measurements on α-helix and β-hairpin formation were usually performed on equilibrium mixtures of molecules. As in these systems the secondary structure elements are most likely already nucleated at the start of the experiment, the observed relaxation kinetics report on the propagation steps and not on the nucleation process. For α-helices relaxation time constants of about 20 ns were obtained[55,56,57], β-hairpins were reported to form on the microsecond timescale[58,59]. In order to obtain information on the nucleation step, it must be ensured that at the start of the experiment the secondary structure is not yet nucleated. This can for example be achieved by cross-linking peptides with photolabile linkers. Ideally, in their linked form these peptides cannot adopt any secondary structure. Upon light irradiation the linker is rapidly cleaved and the peptide is free to adopt its secondary structure. Chen *et al.*[60] reported such studies on β-hairpins in which a cross-linked peptide was liberated by a short laser pulse. Photoacoustic measurements were used to follow the process and a time constant of only 40 ns was observed. Due to the unspecific nature of the monitored signal it remains unclear whether the observed kinetics really report on the nucleation process or another event. Compared to the reported kinetics of β-hairpin formation in the microsecond range[58,59], a time constant of about 40 ns for the nucleation process appears to be unexpectedly fast.

1.2.3.2 – Theoretical Models for the Unfolded State

The properties of unstructured molecules are extensively treated by polymer theory[61,62,63]. Results from experimental studies on unfolded proteins can be compared to the behaviour of unstructured polymer molecules predicted by theory.

In the simplest description a polymer is treated as an ideal or freely jointed chain which consists of n linearly connected segments of length l. The bond and torsion angles between consecutive segments can adopt any value with equal probability. The resulting length distribution of the end-to-end vectors, $p(r)$, is normally (Gaussian) distributed (eq. (5)).

$$p(r) = \sqrt{\frac{3}{2\pi nl^2}}^{3} \cdot 4\pi r^2 \cdot \exp\left\{-\left(\frac{3}{2nl^2}\right)r^2\right\} \tag{5}$$

As shown in eq. (6), the mean square end-to-end distance $<r^2>$ calculated from this distribution function linearly scales with the number of chain segments n.

$$\left\langle r^2 \right\rangle = nl^2 \tag{6}$$

In real polymer chains the assumptions of the freely jointed chain do not hold. Interactions between atoms of neighbouring segments restrict the accessible range of bond and torsion angles. As a result the polymer chain is less flexible and on average has larger dimensions than the freely jointed chain. Flory[61] introduced the characteristic ratio, C, as a measure for the average dimensions of a real polymer chain relative to a corresponding ideal chain. For a real polymer chain $<r^2>$ is given by eq. (7).

$$\left\langle r^2 \right\rangle = C_n nl^2 \tag{7}$$

Characteristic ratios are functions of chain length and approach limiting values C_∞ for a large number of chain segments n. C denotes the average number of consecutive chain segments which propagate in the same direction and thus provides a measure for the stiffness of a real polymer chain. For the limiting case of the freely jointed chain, C equals 1 which indicates that there is no correlation between the orientations of consecutive chain segments. A closely related concept is that of persistence length l_p[62]. The persistence length designates the average distance which a real polymer chain propagates in the direction of a given bond vector. In the limit of long chains, l_p is related to the limiting characteristic ratio C_∞ as shown in eq. (8).

$$l_p = (C_\infty + 1) \cdot \frac{l}{2} \tag{8}$$

For polypeptides limiting characteristic ratios C_∞ of about 2, 9 and 100 were calculated for chains consisting of the amino acids glycine, alanine and proline, respectively[61,62]. The characteristic ratios of the other amino acids closely correspond to that of alanine. Although this model is quite crude, its predictions were found to be in good agreement with

experimental observations, both for short (< 20 residues)[62,64,65,66] and long polypeptides (> 100)[62,67,68,69].

Another complexity in real polymers, which is not included in the simple polymer models discussed above, is the excluded volume effect. Real polymer chains are self-avoiding, i. e. different chain segments cannot intersect. As a consequence an additional increase in the average dimensions of the polymer is expected. $<r^2>$ should no longer scale linearly with n but as a power function of n.[61] For polymers with excluded volume, Edwards[70] derived different probability distribution functions for the length of the end-to-end distance vectors. A skewed Gaussian function as given in eq. (9) was found to sufficiently describe the conformation of unstructured peptides[71].

$$p(r) = c(a,b) \cdot 4\pi r^2 \cdot \exp\left\{-a(r-b)^2\right\} \tag{9}$$

The Gaussian distribution is offset from zero by b. The parameter a determines the width of the distribution and $c(a,b)$ is a normalization constant.

Flory[61] introduced the concept of the so-called Θ-conditions. Under these conditions (i. e. temperature, solvent composition etc.) the excluded volume effect is exactly balanced by intramolecular attractive forces in the polymer chain. Consequently, at these conditions the polymers behave as unperturbed chains and their average dimensions are given by eq. (7).

Another approach to characterize the unfolded state of proteins makes use of all-atom models of polypeptide chains. For the simulation of polypeptide conformations, values of the bond lengths and bond angles are usually derived from high-resolution X-ray structures (e. g. Engh and Huber[72]). Multiple peptide conformations are generated randomly or systematically and evaluated with respect to their potential energy. In the simplest case a hard-sphere potential is employed where atoms are treated as solid spheres which may not overlap. When any two atoms approach closer than their contact distance[73], the corresponding peptide conformation is considered sterically forbidden. Although the hard-sphere potential is a very simple representation of real atom interactions, it reproduces the core features of more elaborate potential functions, like e. g. Lennard-Jones potentials or quantum mechanical approximations[74]. A key advantage is that only pairwise repulsive interactions are considered which greatly reduces the computational effort needed for corresponding simulations. Despite its simplicity, the hard-sphere approximation has been successfully used in a number of studies. In their seminal work Pauling and Corey[75,76] evaluated sterically allowed and hydrogen-bonded conformations of peptide chains and were able to predict the three-dimensional structures of α-helices[77,78,79], β-pleated sheets[80] and the collagen triple helix[81],

almost ten years before the first X-ray structure of a protein[20] was determined at atomic resolution. Later, Ramachandran and others[82,83,84] employed the hard-sphere potential to determine the dihedral angle space sterically allowed for peptides (Ramachandran maps). Similar considerations were also the basis for Flory's treatment of polypeptide chains (see above)[61].

Recently, Rose and coworkers applied the hard-sphere model to assess the properties of unfolded polypeptides[85,86]. By exhaustively enumerating peptide chain conformations, they found that the backbone conformation of unfolded peptides is not only determined by interactions between neighbouring residues, which Flory[61] proposed in his isolated-pair hypothesis. Rather, mid- and long-range interactions render certain conformations sterically forbidden, thereby reducing the conformational space accessible to unfolded peptide chains[85,87]. In a later simulation study[88], which also included peptide solvation, it was found that unfolded peptide chains preferentially adopt the polyproline II helix conformation in accordance with experimental data[46,47]. Recent work by Fitzkee *et al.* showed how the experimental observations of random-coil behaviour and residual structure in unfolded proteins could be reconciled[89,90]. Even though considerable amount of defined structure could exist in large parts of the peptide chain, its global properties might still correspond to those of the random coil.

1.3 – Protein Folding Kinetics

Apart from the question of how the native structure of proteins is encoded in their amino acid sequence, another major aspect of protein folding is the study of the kinetics of this process. How fast is protein folding, does it occur in one concerted step or are there intermediates? What is the dependence of the folding process on external conditions? Are there competing side reactions like e. g. aggregation?

Based on simplistic calculations, Levinthal[91,92] argued that even a peptide chain of only moderate size would need an unrealistically long time to reach its native state if it did so by randomly sampling all sterically allowed conformations (Levinthal's paradox). A small energetic bias towards the native state, however, could greatly reduce the conformational space accessible to the polypeptide chain and thus considerably speed up the folding process[93].

Different theories were proposed to explain how this might proceed[94]. In one model the unfolded peptide chain is thought to undergo a rapid hydrophobic collapse to a so-called molten globule structure which eventually rearranges to the native state[95,96]. The diffusion-collision or framework model[97,98] proposes that secondary structure elements form early during folding and then meet and associate through diffusional events to form the tertiary structure. Both models imply that folding possibly occurs through intermediate states. Indeed, many apparent two-state folding proteins were found to fold through high-energy intermediates[99,100]. Theoretical considerations showed that intermediates can speed up protein folding as long as they are not populated to any large extent during the reaction[101]. In contrast, the nucleation model[102] predicts that folding proceeds without the accumulation of intermediates. In this model, a few contacts between parts of the protein are ascribed a crucial role early in the folding process. Starting from these contact sites the formation of native structure is nucleated. Results from simulations using coarse-grained lattice models are in agreement with this theory and led to the concept of the so-called folding funnel landscapes[103].

The folding kinetics of most proteins are successfully treated in terms of classical reaction kinetics originally developed for simple chemical reactions. Different states of the folding reaction are separated by sufficiently large energy barriers ($\sim 5\,kT$, cp. Fig. 1) and, accordingly, folding kinetics of monomeric proteins can be described by exponential functions. A protein folding reaction involving n monomeric states gives rise to n-1 exponential functions in the corresponding kinetics. Not in all cases are all of these phases observable, as for example in the case of high-energy intermediates which are not populated during the reaction. In the simplest case of a two-state folding protein (Fig. 1) native and unfolded state interchange with the microscopic rate constants, k_f and k_u (eq. (10)).

$$\mathrm{N} \underset{k_f}{\overset{k_u}{\rightleftharpoons}} \mathrm{U}$$
(10)

After perturbation from equilibrium a single-exponential relaxation process with rate constant λ can be observed (eq. (11)).

$$\lambda = k_f + k_u$$
(11)

Analysis of reaction rate constants according to transition state (Eyring)[104] theory (eq. (12)) allows to infer properties of the potential energy curve of the folding reaction. The magnitude

of a microscopic rate constant k is a function of the so-called preexponential factor (k_0) and the free energy barrier for this reaction ($\Delta G^{0\ddagger}$).

$$k = k_0 \cdot \exp\left(-\Delta G^{0\ddagger}/RT\right) \tag{12}$$

k_0 represents the maximum rate constant in the absence of free energy barriers and is specific for each reaction. It imposes a maximum velocity for a given reaction.

An alternative reaction rate theory was formulated by Kramers[105]. Here, reactions are considered as thermally activated diffusional processes along a potential energy surface. For reactions in solution, such as protein folding, the so-called 'high friction limit' holds[105,106]. In this limiting regime the rate constant for interconversion from species U to N, k, is given by eq. (13) (cp. Fig. 1).

$$k = \frac{\omega_U \omega_\ddagger}{2\pi\gamma} \cdot \exp\left(-E_b/RT\right) \tag{13}$$

E_b denotes the height of the energy barrier between U and N, γ is a friction factor, and ω_U and $\omega_\ddagger$ are the frequencies of motion in ground state U and in the transition state $\ddagger$ on top of the energy barrier, respectively. The frequencies ω refer to motions along the potential energy curve and depend on its shape and properties. It is not exactly clear how these motions relate to real motions and dynamics of atoms and molecules, but clearly they must be connected. Therefore, a comprehensive understanding of protein folding kinetics also has to include information about the dynamics at different stages of the folding process.

Kinetic analysis allows to characterize the potential energy curve of the protein folding reaction and its transition states. By definition, transition states represent maxima of potential energy and are therefore unstable and never populated to any significant extent (Fig. 1). Thus, their structure and dynamics cannot be studied directly but have to be inferred indirectly. Rate-equilibrium free energy relationships which were pioneered by Leffler[107] proved to be particularly powerful in that respect. Changes in equilibrium free energy of a reaction (ΔG^0) induced by a perturbation ∂x are correlated with the corresponding changes in the free energy of activation ($\Delta G^{0\ddagger}$).

$$\alpha_x = \frac{\partial \Delta G^{0\ddagger}/\partial x}{\partial \Delta G^0/\partial x} \tag{14}$$

By choosing suitable perturbations ∂x, the transition states of protein folding reactions can be characterized with regard to several different properties[108,109]. Commonly used perturbations include pressure[110] and temperature changes[99,111,112,113], amino acid mutations[114,115] and

changes in solvent composition such as the addition of chemical denaturants[36,37] (see chapter 1.4). The quotient α quantifies to what extent a reaction has occurred in the transition state with respect to the examined property.

1.4 – Chemical Denaturants

Chemical denaturants, most prominently urea and guanidinium chloride (GdmCl), have long been known to stabilize the unfolded state of proteins relative to the native one and to thereby cause protein unfolding[116,117,118,119,120]. In contrast to thermal denaturation, unfolding by urea and GdmCl is usually fully reversible[36,37]. Further, rapid dilution of denaturants in fast-mixing devices facilitates the study of protein folding kinetics with millisecond (stopped-flow)[121] or sub-millisecond time-resolution (continuous-flow)[122,123]. Therefore, chemical denaturants have become the method of choice to unfold proteins. Although the exact molecular mechanism of denaturant action has not been elucidated yet, it is clear that denaturant molecules interact with solvent-exposed parts of the protein[124]. Since unfolded proteins generally have larger solvent-accessible surfaces than native proteins, more interaction sites for denaturants are accessible in the unfolded state. Denaturants thus favour the exposure of protein surface and shift the equilibrium from native towards unfolded protein.

Mainly two models were proposed to account for the mechanism of unfolding by denaturants. The transfer model by Tanford[37,120,125] states that denaturants increase the solubility of certain parts of the protein in the solvent and thus cause their exposure. Free energies of transfer (δg_{tr}) from solutions of varying denaturant concentration [D] to water were determined for simple model compounds such as amino acids and dipeptides[126,127,128,129,130]. In Tanford's model the stability of a protein, ΔG^0, is determined by its transfer free energy from a denaturant solution of concentration [D] to water. The total transfer free energy of a protein can be calculated as the sum of the transfer free energies of its individual groups. Based on the model compound data (δg_{tr}), ΔG^0 is then given by eq. (15).

$$\Delta G^0 = \Delta G^0\left(\mathrm{H_2O}\right) + \sum_{i=1}^{N} \alpha_i \, \delta g_{tr,i} \tag{15}$$

The summation runs over all N groups of the protein; $\Delta G^0(\mathrm{H_2O})$ is the protein stability in the absence of denaturant and the parameters α_i designate the change in the degree of solvent exposure of group i upon unfolding of the protein. In practice, the transfer model suffers from the fact that the coefficients α_i are usually unknown and can only be estimated.

In the binding model[125,131,132,133] denaturant molecules are thought to bind to parts of the protein. Denaturation of the protein would result from a smaller number of binding sites in the native (N) compared to the unfolded state (U). Different binding mechanisms can be assumed, but most often the simplest case of identical and noninteracting binding sites for denaturant molecules is employed. The free energy of folding, ΔG^0, is determined by the difference in the number of denaturant binding sites between U and N, Δn, and the binding constant K (eq. (16)).

$$\Delta G^0 = \Delta G^0(H_2O) + \Delta n \cdot RT \ln(1 + K[D]) \tag{16}$$

Schellman pointed out that the interaction of denaturants with proteins is very weak and that it occurs on the molar concentration range[132,134,135,136]. Over this concentration interval the activity of water significantly changes which is taken into account in the 'site exchange' formalism proposed by Schellman (eq. (17)).

$$\Delta G^0 = \Delta G^0(H_2O) + \Delta n \cdot RT \ln\left(1 + (K_{ex} - 1)X_D\right) \tag{17}$$

Usually, eq. (17) is given on the mole fraction scale of denaturant X_D, rather than on the molarity scale $[D]$. Accordingly, K_{ex} is a dimensionless constant for denaturant binding on the mole fraction scale. It is important to realize that the transfer and the binding model are not in contradiction and that they describe the experimental data on protein folding equally well.

A third empirical model, the linear extrapolation method, was introduced by Greene and Pace[137] and is now the one commonly used for the analysis of protein stability and kinetics. The free energy of protein folding (ΔG^0) was found to linearly increase with denaturant concentration ($[D]$)[138,139] according to eq. (18).

$$\Delta G^0 = \Delta G^0(H_2O) + m_{eq} \cdot [D] \tag{18}$$

Strong support for the applicability of this model comes from the work of Santoro and Bolen[140] who studied protein unfolding induced by different denaturants and found that the extrapolated protein stabilities agree. The so-called equilibrium m-value (m_{eq}) quantifies the sensitivity of protein stability towards denaturant addition. It was shown to correlate with protein size and the change in solvent-accessible surface area upon unfolding[141]. On average, m-values for guanidinium chloride were found to be 2.3 times larger than those observed for urea[141].

Relationships similar to eq. (18) were also found for the activation free energies $\Delta G_f^{0\ddagger}$ and $\Delta G_u^{0\ddagger}$ of the folding and unfolding rate constants of proteins (eq. (19))[37,138].

$$\Delta G_{f,u}^{0\ddagger} = \Delta G_{f,u}^{0\ddagger}\left(\mathrm{H_2O}\right) + m_{f,u} \cdot \left[D\right] \qquad (19)$$

Comparison of m_{eq} with the kinetic m-values m_f and m_u allows the characterization of the transition state of a given reaction with respect to its solvent-accessible surface area[141] (see chapter 1.3, eq. (14)).

$$\alpha_D = \frac{\partial \Delta G_f^{0\ddagger}/\partial[D]}{\partial \Delta G^0/\partial[D]} = \frac{m_f}{m_{eq}} \qquad (20)$$

The linear dependence of the activation free energies on denaturant concentration gives rise to the characteristic V-shape of chevron plots in which the logarithm of the relaxation rate constant is plotted against denaturant concentration (cp eqs. (11) and (12)).

1.5 – Peptide Loop Closure Dynamics

Native proteins are characterized by their defined three-dimensional structure which is stabilized by many interactions between different parts of the protein (see chapter 1.2.2). In contrast, only few or no defined interactions exist in the unfolded state (cp. chapter 1.2.3). The protein folding process amounts to acquiring the native interactions. Therefore, intramolecular contact formation respectively loop closure between two sites on a polypeptide chain should represent an elementary step in the protein folding reaction and should precede the formation of any higher order structure. Loop closure reactions are expected to be particularly important in the unfolded state where no or little defined structure yet exists.

Intramolecular contact formation of polypeptides has been studied by a number of different techniques[63]. Haas and coworkers[142] used fluorescence resonance energy transfer (FRET) from naphthalene to dansyl to determine end-to-end probability distributions and diffusion constants in short peptides. From these data a first estimate for the velocity of end-to-end contact formation was derived[53]. Time constants between 3 and 9 ns were calculated for the loop closure reaction of peptides with 4–8 peptide bonds between the FRET chromophores. Since these FRET measurements do not directly report on the contact formation reaction, the derived contact rates depend on the exact model used for their calculation. Later studies imply that the above results overestimate the real contact formation velocities[54,63] (see below).

Using time-resolved laser spectroscopy, Hagen *et al.*[143] studied the binding of the heme prosthetic group of cytochrome *c* by a methionine residue, which are separated by approximately 50 residues in the primary structure. The observed kinetics showed time

constants of 35–40 μs at 5.6 M GdmCl. A problem in these measurements is that the rebinding of heme by methionine is not diffusion- but reaction-controlled. Later work by Gray and coworkers on cytochrome *c* using electron transfer from optically excited zinc-porphyrine to a ruthenium-complex[144] yielded significantly faster contact formation rate constants.

The first experimental systems to measure absolute rates of contact formation were introduced by Kiefhaber and coworkers[53,54]. Bieri *et al.*[53] employed triplet-triplet energy transfer (TTET) between thioxanthone and naphthalene attached to unstructured model peptides to study loop closure dynamics in ethanol/H$_2$O mixtures. Single-exponential kinetics for contact formation with time constants between 20 and 100 ns were observed for glycine-serine based peptides containing up to nine residues between the spectroscopic labels. In a later study by Krieger *et al.*[54], thioxanthone was replaced by xanthone, thus enabling studies in aqueous solution. Several peptides with different length and sequence were studied and all displayed single-exponential kinetics for contact formation (Fig. 2).
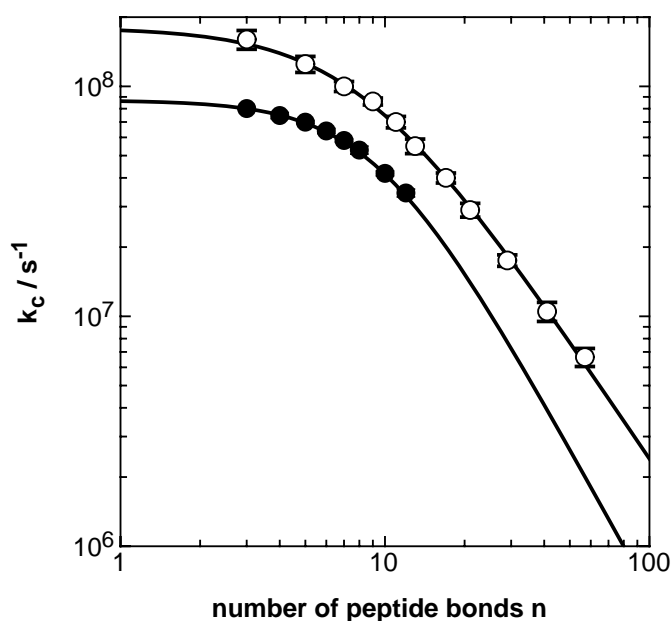
Fig. 2 – Rate constants of end-to-end contact formation in glycine-serine (○) and serine (●) peptides as measured by intramolecular TTET from xanthone to naphthalene. The solid lines represent fits to eq. (21). Data were taken from Krieger *et al.*[54]

The more flexible glycine-serine (GS$_x$) peptides showed smaller time constants ($\tau=1/k_c$) for end-to-end contact formation than serine peptides (S$_x$) of corresponding length. For both peptide series the dependence of the rate constant for end-to-end contact formation, k_c, on the number n of peptide bonds between the TTET labels could be described by eq. (21).

$$k_c = 1/\left(1/k_a + 1/\left(k_b \cdot n^m\right)\right) \tag{21}$$

In the limit of short chains k_c was almost independent of chain length, approaching the limiting value k_a, corresponding to time constants of 5 and 12 ns for the GS_x and S_x peptides. On the other hand, for the longer peptides k_c decreases as a power function of n with a stronger dependence for the S_x ($m = -2.1 \pm 0.3$) than for the GS_x peptides ($m = -1.7 \pm 0.1$). For short chains loop closure dynamics appear to be limited by chain stiffness while in longer chains the probability of forming end-to-end contact is limiting. This view was corroborated in a later study on the temperature and viscosity dependence of loop closure reactions[145]. Krieger et al.[54] further showed that the amino acid sequence affects intramolecular contact formation in peptides. Fastest reactions were observed for glycine-containing peptides while single amino acid substitutions led to a slowdown of contact formation of up to 2.5-fold. Two relaxation processes could be observed in proline-containing peptides and were ascribed to populations of molecules with the peptide bond preceding proline in the *cis* resp. *trans* conformation. Finally, it was reported that the cosolvents urea, GdmCl and ethanol slow down end-to-end contact formation[54]. Subsequently, intrachain contact formation was also studied in heterogeneous peptides with amino acid sequences derived from naturally occurring proteins[145,146]. Their behaviour largely corresponded to that of the model peptides studied before.

Stimulated by the initial studies of Bieri et al.[53], in the last years a number of other techniques were introduced to study intrachain contact formation in peptides (cp. Fierz and Kiefhaber[63]). Eaton and coworkers used quenching of the tryptophan triplet state by thiols or disulfides to measure loop closure reactions in model peptides[147,148,149] and proteins[150,151]. In the group of Nau[152,153,154] fluorescence quenching of DBO, a small organic moiety, by tryptophan was employed. On the single-molecule level Neuweiler et al.[155] used fluorescence quenching of the dye MR121 by tryptophan. Another approach is based on electron transfer from optically excited zinc-porphyrine to a ruthenium-complex[144]. With the exception of the latter work by Gray and coworkers[144], the contact formation rate constants reported in these studies are systematically lower than the results obtained by TTET from xanthone to naphthalene[63]. These discrepancies cannot be fully ascribed to the use of different peptides and proteins in these studies. Rather, it appears that in these systems the different reactions which report on contact formation are not diffusion-controlled as it is the case for TTET from xanthone to naphthalene. Accordingly, in these reaction-controlled systems the observable relaxation times are not solely limited by the frequency of encounter between the spectroscopic labels (see chapter 1.6). Reliable determination of absolute contact formation rates strictly requires a

diffusion-controlled system. In contrast, results obtained from reaction-controlled systems depend on the model used to analyze the experimental data.

Chain dynamics and loop closure reactions are also treated by polymer theory. Early theoretical work by Jacobsen and Stockmayer[156] showed that for an ideal polymer the probability of forming end-to-end contact, p_c, should scale with the number of chain segments n according to eq. (22).

$$p_c \propto n^{-3/2} \tag{22}$$

The first passage time theory developed by Szabo, Szabo and Schulten[157] treats the kinetics of end-to-end contact formation (SSS-theory). Loop closure between two sites on a polymer chain is considered as a diffusional process on a potential energy surface. Within the framework of the theory, contact formation can be well approximated by a single-exponential process, provided that on average only a small fraction of all molecules forms contact simultaneously and that interconversion among molecules which do not form contact is fast (eq. (23)).

$$\Sigma(t) \approx \Sigma(t)_{approx.} = \exp(-t/\tau) \tag{23}$$

Here, $\Sigma(t)$ denotes the probability of a molecule not yet having formed contact at time t. The time constant τ is a function of the end-to-end distance probability distribution, $p(x)$, and end-to-end diffusion constant, D, of the polymer according to eq. (24).

$$\tau = \frac{1}{D} \int_{r_b}^{\infty} \frac{1}{p(x)} \left(\int_x^{\infty} p(y)\,dy \right)^2 dx \bigg/ \int_{r_b}^{\infty} p(x)\,dx \tag{24}$$

The so-called reactive boundary r_b signifies the distance at and below which a reaction between the two sites of the polymer occurs. For an ideal chain the SSS-theory predicts that k_c (=$1/\tau$) should scale with $n^{-3/2}$ which is in agreement with the theory of Jacobsen and Stockmayer[156]. The predictions of the SSS-theory match with the experimental observations for long peptides (see above). Contact formation was shown to be a single-exponential relaxation process. Further, for long peptides k_c scales as a power function of n albeit stronger than predicted.

1.6 – Triplet-Triplet Energy Transfer

Triplet-triplet energy transfer (TTET)[158] denotes the transfer of excited state energy from a triplet donor molecule A to a triplet acceptor molecule B.

$$^3A + {}^1B \xrightarrow{k_{TTET}} {}^1A + {}^3B \tag{25}$$

The superscripts in eq. (25) indicate the electronic state of the molecules, 1 for singlet and 3 for triplet state. Transfer is thought to proceed via a Dexter mechanism[159] involving the exchange of two electrons between donor and acceptor. Fig. 3 shows a Jabłoński diagram describing this process.
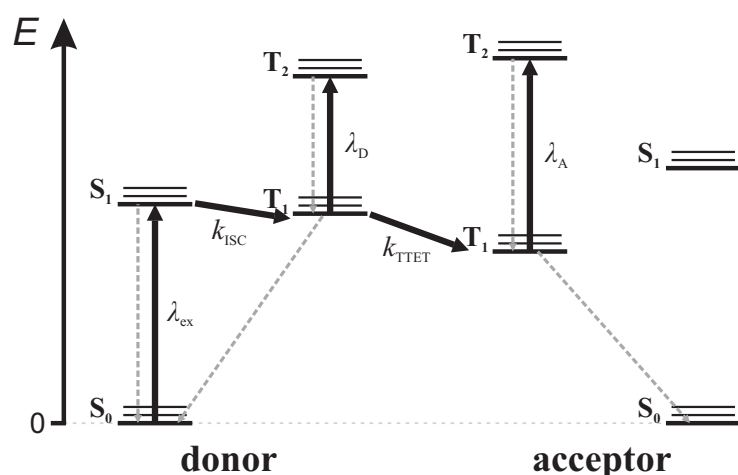

Fig. 3 – Jabłoński diagram of triplet-triplet energy transfer. S_0 and S_1 denote singlet electronic states, T_1 and T_2 triplet states. Donor chromophores are optically excited at λ_{ex} and undergo fast intersystem crossing (k_{ISC}) to the triplet state T_1. Upon van-der-Waals contact formation, triplet excited state energy is transferred to the acceptor (k_{TTET}). Triplet states can be monitored due to their intense long-wavelength absorption bands (λ_D and λ_A). Broken gray arrows represent internal conversion processes. The relative energy levels of the electronic states correspond to the values measured for xanthone[160,161] (donor) and 1-methyl-naphthalene[160,162] (acceptor).

The rate constant for energy transfer, k_{TTET}, shows an exponential dependence on the distance r separating donor and acceptor (eq. (26))[158,159].

$$k_{TTET} = A \cdot \exp\left(-\frac{2r}{L}\right) \tag{26}$$

Here, L is the average van-der-Waals radius of donor and acceptor. Due to the strong distance dependence of the transfer rate constant, efficient triplet-triplet energy transfer essentially requires van-der-Waals contact between donor and acceptor. Therefore, in solution the first

step in a TTET reaction between two molecules A and B has to be their diffusional[163,164,165] encounter according to eq. (27).

$$^1A + {}^1B \underset{k_0^A}{\overset{k_{ex}}{\rightleftharpoons}} {}^3A + {}^1B \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} {}^3A\cdot{}^1B \underset{k_{-TTET}}{\overset{k_{TTET}}{\rightleftharpoons}} {}^1A\cdot{}^3B \underset{k_{on}}{\overset{k_{off}}{\rightleftharpoons}} {}^1A + {}^3B \overset{k_0^B}{\longrightarrow} {}^1A + {}^1B \tag{27}$$
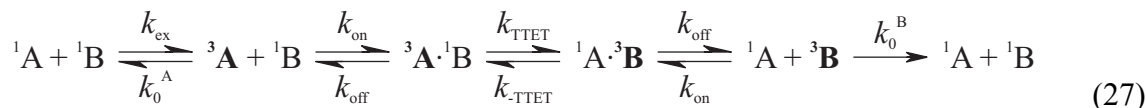
k_{on} and k_{off} denote the rate constants for the association and dissociation of the encounter complex. Formation of the triplet state (k_{ex}) should be fast, since it determines the maximum time-resolution of TTET experiments. Triplet states are usually long-lived since the reversion to their singlet ground states is spin-forbidden and thus slow (k_0^A and k_0^B). If the triplet energy of the donor is higher than that of the acceptor by more than a few kJ/mol[158], then k_{TTET} will be much larger than the rate constant for the reverse reaction (k_{-TTET}) and the transfer reaction will be irreversible.

When triplet donor and acceptor are attached to one molecule, intramolecular contact formation between the sites of attachment can be studied according to eq. (28) (cp. chapter 1.5).[63]

$$O \underset{k_{-c}}{\overset{k_c}{\rightleftharpoons}} C \overset{k_{TTET}}{\longrightarrow} C^* \tag{28}$$

In eq. (28), O denotes open conformations of the molecule in which no contact between triplet donor and acceptor exists. Intrachain diffusion (k_c) leads to the formation of intramolecular contact between the triplet labels (C conformations) and to consecutive triplet-triplet energy transfer (C^*). The reaction scheme depicted in eq. (28) gives rise to two observable rate constants $\lambda_{1,2}$ which are functions of the microscopic rate constants.

$$\lambda_{1,2} = \frac{B \pm \sqrt{B^2 - 4C}}{2} \tag{29}$$

$$B = k_c + k_{-c} + k_{TTET} \qquad C = k_c \cdot k_{TTET} \tag{30}$$

If on average only a small fraction of molecules forms contact at a given time (i. e. $k_c \ll k_{-c}$), the observed kinetics will be single-exponential and the relaxation rate constant will equal the smaller eigenvalue λ_1 (eq. (29)).

Absolute rate constants of contact formation can only be measured if the transfer reaction is sufficiently faster than chain diffusion ($k_{TTET} \gg k_c, k_{-c}$). In this case the observable reaction is diffusion-controlled and λ_1 equals k_c (eq. (31)).

$$\lambda_1 = k_c \tag{31}$$

Experimentally, it was found that the reaction occurs in a diffusion-controlled manner, when the energy transfer reaction is exergonic, i. e. the triplet energy of the donor molecule is higher than that of the acceptor molecule[158]. If on the other hand the transfer reaction occurs on the same timescale as chain diffusion, the measured rate constant for contact formation contains contributions from the transfer process itself and no longer reports on the absolute collision frequency of the triplet labels. In the limit of very slow TTET ($k_{\text{TTET}} \ll k_{\text{c}}, k_{\text{-c}}$) the observable kinetics are fully reaction-controlled and λ_1 is given by eq. (32).

$$\lambda_1 = \frac{k_c}{k_c + k_{-c}} \cdot k_{TTET} \approx \frac{k_c}{k_{-c}} \cdot k_{TTET} \qquad (32)$$

Under these conditions the observable kinetics can provide information about the equilibrium between closed and open chain conformations (k_c/k_{-c}). It should be noted that these considerations closely correspond to the theory of hydrogen-deuterium exchange[24,27].

The triplet transfer reaction from xanthone to naphthalene employed in the present work was shown to be diffusion-controlled[145]. Formation of the donor triplet state occurs sufficiently fast ($\sim 2\ \text{ps}$)[166] and with a high quantum yield ($\sim 99\ \%$). Triplet transfer to naphthalene irreversibly proceeds with a time constant of approximately $1\ \text{ps}$[166]. The lifetime of the xanthone triplet state is on the order of tens of microseconds. Xanthone and naphthalene triplet states strongly absorb light at 590 and 420 nm, respectively, which allows to monitor the transfer process. Taken together, triplet-triplet energy transfer from xanthone to naphthalene is well suited for the study of peptide dynamics (cp. chapter 1.5).

1.7 – Fluorescence Resonance Energy Transfer

Fluorescence resonance energy transfer (FRET) designates the radiationless transfer of excited-state energy between two fluorophores commonly termed donor and acceptor. Förster developed the theoretical background and the formalism describing this process[167]. Energy transfer occurs through space and is mediated by the resonant oscillation of transition dipoles in the two chromophores. As the resonance condition has to be met, FRET only occurs between pairs of chromophores with spectral overlap between the donor emission (fluorescence) and the acceptor absorbance spectra. Fig. 4 shows absorption and emission spectra of the FRET chromophores naphthyl and dansyl also used in the present work.
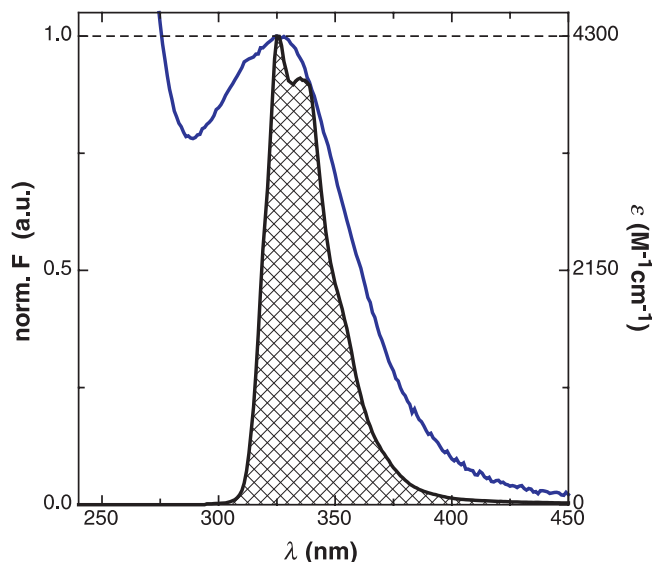
Fig. 4 – Normalized fluorescence emission spectrum of a naphthyl-labelled peptide (—) and absorption spectrum of a dansyl-labelled peptide (—). The FRET overlap integral (cp. eq. (34)) is indicated as the cross-hatched area.

The rate of energy transfer, k_T, depends on the properties of the chromophores and on the distance r separating them according to eq. (33).

$$k_T = \frac{1}{\tau_D}\left(\frac{R_0}{r}\right)^6 \tag{33}$$

Here, τ_D denotes the intrinsic fluorescence lifetime of the donor fluorophore and R_0 the characteristic Förster distance which is a function of the properties of the given pair of fluorophores (eq. (34)).

$$R_0^6 = \frac{9000\,\ln 10\;\kappa^2\Phi_D}{128\pi^5 Nn^4}\int_0^\infty F_D(\lambda)\varepsilon_A(\lambda)\lambda^4 d\lambda \tag{34}$$

In eq. (34) N is Avogadro's number, n the refractive index of the medium, κ^2 an orientational term which equals 2/3 for rapidly and isotropically reorienting fluorophores, Φ_D the fluorescence quantum yield of the donor, $F_D(\lambda)$ the normalized emission of the donor and $\varepsilon_A(\lambda)$ the extinction coefficient of the acceptor at wavelength λ. The dependence of the energy transfer rate on the inverse sixth power of the interchromophore separation can be utilized to accurately determine distances on a molecular scale. To this end the efficiency of energy transfer, E, is usually calculated by comparing the fluorescence properties of the donor chromophore in the absence and in the presence of the acceptor (eq. (35)).

$$E = \frac{F_D - F_{DA}}{F_D} = \frac{\tau_D - \tau_{DA}}{\tau_D} = \frac{k_T}{1/\tau_D + k_T} \tag{35}$$

FRET efficiencies can be determined by steady-state fluorescence measurements where the fluorescence intensities in the absence (F_D) and the presence (F_{DA}) of the acceptor chromophore are determined. Typically, a more accurate determination of E is afforded by comparing the corresponding fluorescence lifetimes, τ_D and τ_{DA}, determined by time-resolved FRET (trFRET) measurements. Combining eqs. (33–35) shows that transfer efficiencies can be directly correlated with the distance r separating the two fluorophores according to eq. (36).

$$E = \frac{R_0^6}{R_0^6 + r^6} \tag{36}$$

As illustrated in Fig. 5, FRET measurements are mainly sensitive to distance changes around the characteristic distance R_0 at which the transfer efficiency amounts to 50 %. The applicability of the above theory has first been shown experimentally by Stryer and Haugland[65]. Employing suitable chromophore pairs, distances on the scale of 10 to 100 Å can be accurately determined.
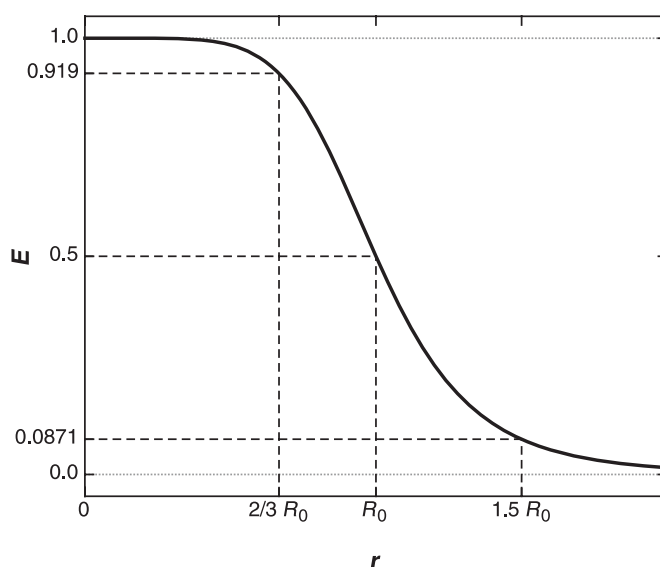
Fig. 5 – FRET efficiency *E* as a function of the distance *r* between donor and acceptor fluorophore. At the characteristic Förster distance R_0 the transfer efficiency amounts to 50 %.

The basic FRET theory has to be extended when studying unstructured systems such as unfolded polypeptides. Owing to their flexibility, the molecules in study constitute a heterogeneous ensemble with differing interchromophore distances and thus energy transfer rates. In trFRET measurements a multiexponential fluorescence decay, $F_{DA}(t)$, will be observed (eq. (37)).

$$F_{DA}(t) = \int_0^\infty p(r) \exp\left\{-\left(1+\left(\frac{R_0}{r}\right)^6\right)\frac{t}{\tau_D}\right\} dr \tag{37}$$

where $p(r)$ is the probability of a molecule having an inter-fluorophore distance r. As discussed in chapter 1.2.3, for the analysis of unstructured peptides often a skewed Gaussian probability distribution is assumed for $p(r)$ (eq. (38))[71].

$$p(r) = c\pi r^2 \exp\left(-a(r-b)^2\right) \tag{38}$$

In addition, when studying flexible molecules by FRET, one has to take into account that the interfluorophore distance governing the energy transfer rate is subject to change during the finite time needed for the measurement. Haas and coworkers[142] developed a formalism which treats conformational rearrangements en bloc as a diffusional process (eqs. (39) and (40)).

$$\frac{\partial \bar{p}(r,t)}{\partial t} = -\frac{1}{\tau_D}\left\{\left(1+\left(\frac{R_0}{r}\right)^6\right)\bar{p}(r,t)\right\} + \frac{1}{p_0(r)}\frac{\partial}{\partial r}\left(p_0(r)D\frac{\partial \bar{p}(r,t)}{\partial r}\right) \tag{39}$$

$$\bar{p}(r,t) = \frac{p(r,t)}{p_0(r)} \tag{40}$$

D signifies the diffusion constant for motions of the two fluorophores relative to each other. Even for comparatively simple probability distribution functions, $p(r)$, eq. (39) does not possess an analytical solution. However, assuming an initial probability distribution $p_0(r)$ (eq. (38)) and suitable boundary conditions (eqs. (41) and (42)), it is always possible to derive a numerical solution.

$$\left.\frac{\partial \bar{p}(r,t)}{\partial r}\right|_{r=0} = 0 \tag{41}$$

$$\left.\frac{\partial \bar{p}(r,t)}{\partial r}\right|_{r=r_{max}} = 0 \tag{42}$$

Equations (41) and (42) impose reflective boundaries at the minimum and maximum values of the interchromophore distance, 0 and r_{max}, respectively.

1.8 – Trimerization Domain of Phage T4 Fibritin (Foldon)

Another part of this thesis deals with the refolding kinetics of the foldon domain from bacteriophage T4 fibritin. Fibritin is a fibrous structural protein and forms the so-called 'whiskers' which are attached to the neck of the virus and participate in virus assembly and host recognition during the infectious cycle[168]. The 486 amino acid protein fibritin assembles into homotrimers and comprises three domains[169,170]. A large central domain (residues 47–456) forms a three-helix coiled coil and is flanked by two globular domains. Via the N-terminal anchor domain (1–46) fibritin is attached to the virus neck. Owing to its biological function of efficiently assembling and correctly aligning three fibritin monomers into a trimer, the C-terminal domain (457–486) is referred to as foldon. Its structure was elucidated by Tao *et al*.[170] who found that foldon forms a *β*-hairpin and assembles into a trimeric *β*-propeller structure (Fig. 6).

Fig. 6 – Structure of the C-terminal domain of T4 phage fibritin (foldon). (left) Topology of the intermolecular *β*-sheet structure. Dashed lines indicate hydrogen bonds. (right) Three-dimensional structure of the foldon domain. Tryptophan residues 476 are highlighted in green colour, glutamic acid 461 and arginine 471 forming an intermolecular salt bridge in yellow colour. The figure was prepared with the programs MOLMOL[171] and PovRay[172] and is based on the PDB coordinates 1RFO[173].

The trimeric structure is stabilized by intermolecular backbone hydrogen bonds, intermolecular salt bridges between glutamic acid 461 and arginine 471, and by a hydrophobic cluster formed by the tryptophan residues 476. Foldon was shown to be structured and stable as an isolated domain[174]. Its reversible unfolding reaction could be

described by a cooperative two-state equilibrium transition between unfolded monomer (U) and native trimer (N_3) according to eq. (43)[174].

$$3\,U \rightleftharpoons N_3 \tag{43}$$

Foldon has also been employed as a trimerization domain for other proteins like collagen[174,175,176,177] and an HIV envelope protein[178].

In addition to its biological relevance and its biotechnological application as a trimerization nucleus, foldon also represents a favourable model system for studying fast folding reactions of oligomeric proteins involving association steps. Previous folding studies on trimeric proteins mainly dealt with large filamentous proteins which show slow and complex folding kinetics[179]. An additional complication arises from aggregation reactions in these systems[180]. Furthermore, it is of interest that despite its small size the foldon domain displays all the stabilizing interactions also contributing to the stability of larger proteins. Since the foldon domain only comprises 27 amino acids, it can be chemically synthesized using solid-phase peptide synthesis. This facilitates targeted modification and the introduction of spectroscopic probes such as those used for triplet-triplet or fluorescence resonance energy transfer (cp. chapters 1.6 and 1.7).

2. Aims of Research

Proteins exist as an equilibrium mixture of their native and unfolded conformations. A full understanding of protein folding equilibria and kinetics requires detailed knowledge about the structure and dynamics of the participating states. For the native state of many proteins such information is provided by methods for high-resolution structure determination. In stark contrast, the unfolded state of proteins is generally less well defined. The study of the structure and dynamics of unfolded polypeptides is the main topic of this thesis.

Under physiological conditions most proteins predominantly exist in their native states. Addition of chemical denaturants is a convenient means to populate the unfolded state of proteins. Despite their widespread use, the mode of action of denaturants is still not well understood. Therefore, in this work special emphasis was put on the question of how chemical denaturants affect the structure and dynamics of unfolded polypeptide chains.

Loop closure kinetics in unstructured peptides can be studied using the triplet-triplet energy transfer (TTET) technique. Intramolecular contact formation between two sites on a polypeptide chain represents an elementary step in protein folding reactions and is thus of special interest. Using TTET, we studied chain dynamics in unstructured model peptides and addressed the following questions:

- How do denaturants affect the chain dynamics of unfolded polypeptides?
- What is the molecular origin of the denaturant effect on chain dynamics?
- How are peptide chain dynamics and their denaturant dependence influenced by length and amino acid sequence of the peptide?
- Where is the main interaction site for denaturants on the peptide chain?

To obtain information about the structure of unfolded polypeptides, we employed a second optical spectroscopic technique, namely fluorescence resonance energy transfer (FRET). An experimental strategy was developed to study the following problems:

- Which structure does the ensemble of unfolded polypeptides adopt?
- How is this structure affected by denaturant addition?
- Can we get additional information about peptide chain dynamics?
- How do the results from TTET and FRET experiments compare to each other?

We further tried to rationalize our experimental results with theoretical models of unfolded polypeptides. To this end a computer program for all-atom simulations of peptide conformations was developed and used to assess the conformational space available to unfolded polypeptide chains. Results from these simulations were correlated with the experimental observations on the dynamics and structure of unfolded polypeptides.

In a further part of this thesis we studied the folding reaction of the foldon domain of T4 phage fibritin. The trimeric foldon domain represents a suitable model system to study the folding of oligomeric proteins. In light of this, a detailed understanding of the folding mechanism of foldon seems to be desirable.

3. Summary of Published Work and Manuscripts Ready for Submission

3.1 – Effect of Denaturants on Intrachain Contact Formation in Unfolded Polypeptides

As discussed in chapter 1.4, denaturants are widely used to study protein folding and stability. The molecular mechanism by which chemical compounds exert their denaturing effect remained unclear. Krieger *et al.*[54] performed initial studies to assess the influence of chemical denaturants and other cosolvents on peptide dynamics and found linear relationships between the logarithm of the end-to-end contact formation rate constant, k_c, and the concentration of cosolvent.

In this work we examined the effect of the chemical denaturants urea and guanidinium chloride (GdmCl) on peptide dynamics in more detail. Using triplet-triplet energy transfer (TTET) from xanthonic acid to naphthalene, absolute rate constants of end-to-end contact formation were measured in a series of (glycine-serine)-repeat peptides at different concentrations of the two denaturants. As observed before[54], all peptides showed linear relationships between ln k_c and the concentrations of both denaturants. m_c-values which define the slopes of these correlations were found to be only weakly dependent on peptide length and amounted on average to (0.35 ± 0.02) and (0.52 ± 0.04) kJ·mol^{-1}·M^{-1} for urea and GdmCl, respectively.

Closer inspection revealed that these linear relationships result from a superposition of two effects. First, high concentrations of denaturants lead to an increase in solvent viscosity and secondly denaturant molecules directly interact with parts of the peptides. The effect of solvent viscosity on contact formation in the model peptides was assessed by TTET measurements in glycerol/water mixtures. The rate constant of end-to-end contact formation was found to decrease as a power function of solvent viscosity. With these data the chain dynamics in denaturant solutions could be corrected for solvent viscosity. After viscosity-correction the contact formation rate constants were still dependent on denaturant concentration. The residual effect of denaturants on chain dynamics could be described by a weak binding isotherm employing the 'site exchange' formalism developed by Schellman (cp. chapter 1.4).

Constants for denaturant binding were independent of peptide length and amounted to (0.26 ± 0.01) and (0.62 ± 0.01) M^{-1} for urea and GdmCl, indicating a stronger interaction of the latter denaturant. However, the effect of denaturant binding on peptide chain dynamics was virtually the same for the two denaturants thereby implying a common mechanism of denaturation.

Contact formation reactions followed single-exponential kinetics at all denaturant concentrations, which indicates rapid equilibration of bound denaturant molecules and thus short residence times for denaturants on the peptide. Our results indicate that m_f-values found for protein folding reactions (cp. chapter 1.4) might contain significant contributions from the denaturant dependence of peptide chain dynamics. This effect is expected to contribute most pronouncedly to m_f-values of small proteins which are on the order of 1–5 $kJ·mol^{-1}·M^{-1}$.

These results were published in[181] (see chapter 8.1, p. 74):

Möglich, A., Krieger, F., Kiefhaber, T. (2005) Molecular Basis for the Effect of Urea and Guanidinium Chloride on the Dynamics of Unfolded Polypeptide Chains. *J. Mol. Biol.* **345(1)**, 153–162.

3.2 – Effect of Denaturants on the Dynamics and Dimensions of Unfolded Polypeptide Chains

In our previous work[181] (chapter 3.1) we found that the effect of chemical denaturants on the intrachain dynamics of unstructured peptides results from a superposition of two effects. First, high denaturant concentrations lead to an increase of solvent viscosity. Secondly, denaturants bind to peptides and thereby slow down their dynamics. However, it remained unclear how exactly the binding of denaturant molecules impedes the chain dynamics of unstructured peptides.

To address this question, we studied the conformational and dynamic properties of unstructured model peptides at different guanidinium chloride (GdmCl) concentrations by fluorescence resonance energy transfer (FRET). (Glycine-serine) repeat peptides corresponding to the ones used before in triplet-triplet energy transfer (TTET) measurements (see chapter 3.1) were labelled at their termini with fluorescent dyes. Time-correlated single photon-counting was employed to measure time-resolved FRET in these peptides. As (glycine-serine) peptides are devoid of defined structure, they undergo fast conformational

rearrangements and display rapid chain dynamics on the nanosecond timescale[54,181]. These processes are expected to occur to a significant extent during the finite experimental observation time of FRET measurements, which is governed by the fluorescence lifetime of the donor chromophore. Consequently, the observable FRET signal contained contributions from intrachain diffusion. We developed a strategy to accurately resolve the influence of distance and motion in time-resolved FRET measurements. In separate FRET experiments two donor fluorophores with largely different fluorescence lifetimes were used, which entails differential weighting of diffusional events in these experiments and affords reliable data analysis. Using this approach, we determined the average dimensions of a peptide consisting of 16 repeats of the dipeptide (glycine-serine). At all GdmCl concentrations the experimental data could be sufficiently described by assuming a simple, idealized probability distribution function for the end-to-end distance of the peptide. Similar to the denaturant effect on the rate constant of loop closure (chapter 3.1), we found that the root mean square end-to-end distance increased hyperbolically with the GdmCl concentration. These results indicate that also unfolded polypeptides can undergo drastic changes in chain dimensions upon altering the solvent composition. In solutions of high denaturant concentration the average peptide dimensions were similar to predictions from polymer theory[61] and were consistent with small-angle X-ray scattering data on proteins unfolded by GdmCl[38,39]. We correlated our experimental data with results obtained from all-atom simulations of sterically allowed conformations of this peptide and found good agreement with regard to the average end-to-end distance. In these simulations only repulsive steric interactions but no attractive forces were considered. Thus, we concluded that GdmCl binding abolishes intramolecular forces between parts of the peptide and thereby causes an increase of the average molecule dimensions.

Analysis of the experimental data further revealed that the end-to-end diffusion constant varied with denaturant addition. In previous studies (cp. chapter 3.1)[145,181], we realized that peptide chain dynamics are strongly affected by solvent viscosity. Therefore, end-to-end diffusion constants were corrected for the known influence of viscosity to obtain information on the internal dynamics of the peptide. Similar to the GdmCl effect on average dimensions, also the viscosity-corrected end-to-end diffusion constants increased hyperbolically with denaturant concentration. This is consistent with the above notion that GdmCl abolishes intramolecular attractive forces: in the absence of these interactions the peptide chain might undergo faster conformational rearrangements which is reflected in an increase of the end-to-

end diffusion constant. Taken together, these results show that denaturant solutions are good solvents for polypeptides whereas water is a bad one.

Based on the results from our FRET measurements, we calculated end-to-end contact formation rate constants according to the first passage time theory by Szabo *et al.*[157] (cp. chapter 1.5). The calculated rate constants were found to be in excellent agreement with those obtained from TTET measurements. These data provide the best evidence to date that the first passage time theory is a valid description for intramolecular contact formation in long peptides.

These findings are described in detail in the following manuscript (see chapter 8.2, p. 84):

Möglich, A., Kiefhaber, T. (2005) Effect of Guanidinium Chloride on Dimensions and Dynamics of Unfolded Polypeptide Chains. *to be submitted*.

3.3 – Determination of the Main Denaturant Binding Site on Peptides

As described in chapter 3.1, part of the denaturant effect on peptide chain dynamics is due to binding of denaturant molecules to the peptide. In chapter 3.2, we used fluorescence resonance energy transfer to determine how denaturant binding affects the conformational and dynamic properties of unstructured peptides. Since all these fundamental studies were conducted on homologous (glycine-serine) repeat peptides, we could not precisely determine which parts of the peptide denaturants mainly interact with. Our previous studies[181] only provided tentative evidence that the peptide backbone might be the main binding site.

In order to address this question in more detail, we studied chain dynamics and their denaturant dependence in host-guest peptides. Different 'guest' amino acids were incorporated into common 'host' peptide contexts and end-to-end contact formation was measured by triplet-triplet energy transfer (TTET). Similar to previous studies[54], we found fastest chain dynamics for glycine-containing peptides. Replacement of glycine with hydrophilic or charged amino acids decelerated the end-to-end contact formation reaction by up to twofold. A further slowdown of chain dynamics was observed for aliphatic amino acid sidechains.

As before[181], TTET experiments were performed at different concentrations of urea and guanidinium chloride (GdmCl). For all peptides the logarithm of the rate constant of end-to-end contact formation k_c linearly decreased with denaturant concentration. The highest m_c-

values, which quantify the sensitivity of chain dynamics towards denaturant addition, were found for charged and polar amino acids. For most 'guest' amino acids the m_c-values obtained for the effect of GdmCl were about 1.5 times higher than the corresponding values found for urea. In contrast, the basic residues lysine and arginine displayed a different behaviour. In these peptides the ratio of the m_c-values for the effect of GdmCl and urea on chain dynamics only amounted to about 1.3.

To discriminate between the effects of solvent viscosity and denaturant binding on chain dynamics, we studied the viscosity dependence of end-to-end contact formation. Corresponding TTET experiments were performed in glycerol/water mixtures. For all peptides k_c decreased as a power function of the solvent viscosity. Data fitting allowed to determine the viscosity coefficients β which quantify this dependence. Similar to the denaturant effect on chain dynamics, the highest absolute β-values were observed for charged and polar amino acids. The denaturant and viscosity coefficients m_c and β are linearly correlated in the different host-guest peptides which confirms that part of the observed denaturant effect on chain dynamics is due to increases in solvent viscosity. Comparison of the β-coefficients with the m_c-values found for peptides containing the basic 'guest' amino acids clearly showed that the interaction of these residues with GdmCl is particularly weak in comparison to other amino acids.

Following our previous analysis[181], we used the β-coefficients to correct the chain dynamics measured in denaturant solutions for solvent viscosity. For all peptides the corrected values of k_c could be described by a weak binding model. Binding constants and the effect of bound denaturant molecules were found to be very similar for all peptides. In agreement with the above analysis, however, basic amino acids displayed remarkably weak interactions with GdmCl. Taken together with the above observation that the coefficients for the denaturant and viscosity effect on chain dynamics are correlated for all peptides, these data imply that denaturants interact with these peptides in a similar way. This suggests that the main denaturant binding site is the peptide backbone which was invariant in the series of host-guest peptides. As the only exception, lysine and arginine seem to differ from the other amino acids in that their interaction with GdmCl is weak. This might be due to electrostatic interactions between the like charged amino acid sidechain and the guanidinium cation.

Details of these studies can be found in the following manuscript (see chapter 8.3, p. 121):

Möglich, A., Kiefhaber, T. (2005) The Peptide Backbone is the Major Interaction Site for Urea and Guanidinium Chloride. *to be submitted*.

3.4 – Effect of Proline and Glycine Residues on Intrachain Contact Formation

Proline and glycine residues markedly differ from the other standard amino acids in their conformational properties. While glycine displays the largest flexibility of all amino acids due to its lack of a C^β-atom, the imino acid proline contains a five-membered pyrrolidine ring which is comparatively rigid. Peptide bonds preceding proline residues occur in the *cis* conformation to a significant extent. Analysis of protein structures showed that glycine and proline residues occur with high frequency within loop and turn structures[182,183]. Formation of these structure elements might represent an important step early in the protein folding reaction and thus is of special interest (cp. chapter 1.5).

In this work we used triplet-triplet energy transfer between xanthonic acid and naphthalene to study the effect of glycine and proline residues on peptide chain dynamics. Single glycine and proline residues were incorporated into oligoserine peptides of different lengths and rate constants of end-to-end contact formation, k_c, were determined. Introduction of glycine into peptides led to an up to twofold increase of k_c compared to oligoserine peptides of corresponding length. Proline residues gave rise to complex contact formation kinetics. Two exponential phases could be identified and assigned to peptide conformers with the peptide bond preceding proline in the *trans* and *cis* conformations. Contact formation in the *cis* proline isomers was always faster than in the corresponding oligoserine reference peptides, that of *trans* isomers always slower. Proline-containing peptides showed a different dependence of the intrachain contact formation rate constants on peptide length than observed before for oligo-(glycine-serine) and oligoserine peptides[54]. Both for the *cis* and *trans* isomers maximum rate constants were observed at peptide lengths of 5 and 8 amino acids, respectively. Glycine and proline residues only exert local effects on peptide dynamics; loop formation kinetics in peptides with more than four residues flanking a central glycine or proline on either side corresponded to the kinetics observed for oligoserine reference peptides of the same length.

Temperature-dependent studies revealed approximately the same Arrhenius activation barriers and preexponential factors for intrachain contact formation in oligoserine and glycine-containing peptides. The activation parameters observed for the *trans* isomers of proline-containing peptides corresponded to these values as well. In contrast, the *cis* proline conformers showed significantly higher activation barriers and preexponential factors.

The different behaviour observed for the *cis* and *trans* isomers of proline-containing peptides could be rationalized by all-atom simulations of peptide conformations employing a hard-sphere potential. As visualized in a Ramachandran plot, the conformational space amenable to an amino acid preceding a proline residue in the *cis* conformation was drastically reduced compared to the one observed for the *trans* isomer of proline. These findings could provide an explanation for the larger activation energies of contact formation in the *cis* isomers. A restricted conformational space might cause more frequent atom collisions during configurational rearrangements. Bond rotations and consequently contact formation would therefore encounter higher activation barriers.

Cis peptide bonds induce a bend in the peptide backbone, thereby bringing the peptide ends into closer proximity. Accordingly, the average end-to-end distance observed for the simulated peptides was always lower in the *cis* than in the *trans* conformers. As is evident from previous studies on the dependence of contact formation on peptide length[53,54] and from theory[156,157], larger end-to-end distances cause slower intrachain contact formation rates. This effect taken together with the restricted conformational space could account for the faster loop formation kinetics in the *cis* compared to the *trans* peptides. For longer peptides the relative difference in the end-to-end distances of the two isomers gradually diminished, which concurs well with the experimental observations. With increasing peptide length the difference in contact formation velocities of *cis* and *trans* conformers decreased. For peptides longer than six amino acids the two kinetic phases could no longer be resolved.

These results are described in detail in the following publication[184] (see chapter 8.4, p. 157):
Krieger, F., Möglich, A., Kiefhaber, T. (2005) Effect of Proline and Glycine Residues on Dynamics and Barriers of Loop Formation in Polypeptide Chains. *J. Am. Chem. Soc.* **127(10)**, 3346–3352.

3.5 – Folding Kinetics of the Trimerization Domain of Phage T4 Fibritin (Foldon)

Foldon is the C-terminal domain of the structural protein fibritin from the T4 phage. A previous study[174] showed that foldon is essential for the efficient association and correct assembly of fibritin. In this work the folding kinetics and the solution structure of the foldon domain were studied.

The three-dimensional structure of the isolated domain determined by liquid-state NMR spectroscopy closely resembled that of the foldon domain in fibritin-foldon constructs which were elucidated by X-ray crystallography[170]. As expected for an oligomeric protein, the equilibrium stability of foldon was found to be dependent on protein concentration. Denaturant-induced equilibrium transitions were used to determine the free energy of folding of (-89.2 ± 0.6) kJ·mol^{-1} corresponding to the standard state of 1 M protein concentration.

Folding kinetics of foldon were studied by rapid dilution of denaturant in a stopped-flow mixer and showed a complex pattern. Global analysis of direct refolding experiments conducted at several protein concentrations between 0.5 and 100 μM and interrupted refolding experiments allowed to determine the minimal model for the folding mechanism of foldon as shown in eq. (44).

$$
\begin{array}{ccccccccc}
81\,\% & 3\,U_t & \underset{k_{-m}}{\overset{k_m}{\rightleftharpoons}} & 3\,I_t & \underset{59\,\text{s}^{-1}}{\overset{1.9\cdot10^6\,\text{M}^{-1}\text{s}^{-1}}{\rightleftharpoons}} & D & \underset{110\,\text{s}^{-1}}{\overset{5.4\cdot10^6\,\text{M}^{-1}\text{s}^{-1}}{\rightleftharpoons}} & T & \underset{4.2\cdot10^{-4}\,\text{s}^{-1}}{\overset{210\,\text{s}^{-1}}{\rightleftharpoons}} & N \\
& k_{ct}\updownarrow k_{tc} & & 2.0\cdot10^{-2}\,\text{s}^{-1} \updownarrow 5.7\cdot10^{-3}\,\text{s}^{-1} & & +\,I_t & & & \\
19\,\% & 3\,U_c & \underset{k_{-m}}{\overset{k_m}{\rightleftharpoons}} & 3\,I_c & & & & &
\end{array}
$$

(44)
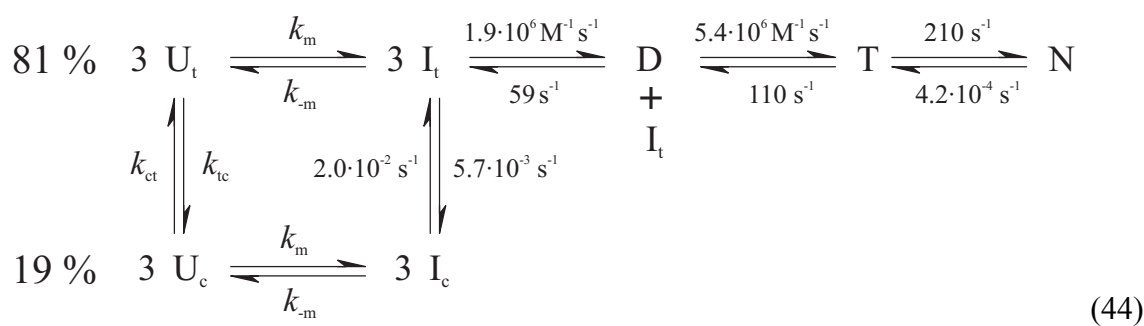
The kinetic scheme in eq. (44) defines a set of ordinary differential equations which describe the time evolution of the system. Experimental data were fitted to the numerical solution of the reaction scheme. Rate constants and relative signal amplitudes were determined as global parameters. (Details of the data analysis are given in chapter 9.1.)

Upon rapid dilution of denaturant, unfolded foldon monomers (U) undergo a fast isomerization to an intermediate state (I) occurring within the dead-time of the stopped-flow instrument (~ 2 ms). Via two fast association reactions dimeric and trimeric species (D and T) are formed. Within the trimeric state an isomerization reaction leads to the native state (N). As foldon contains two proline residues, a fraction of slow-folding molecules (I_c, ~ 19 %) was found which have at least one proline in the non-native *cis* conformation. The bimolecular rate constants of 1.9 and $5.4\cdot10^6$ M^{-1}s^{-1} are among the fastest reported so far for association reactions during protein folding.

Regarding its spectroscopic properties, the monomeric burst-phase intermediate I closely resembles a compact monomeric form A of foldon, which can be induced at pH 2.0. In an additional study[185,186,187], it could be shown that the A-state has its *β*-hairpin mostly formed. Rapid collapse and formation of the *β*-hairpin within the burst-phase intermediate I could be instrumental in the fast association reactions observed for foldon. Even at the putative *in vivo*

concentrations of foldon monomer on the order of 5 μM, folding and association of the foldon domain would be still fast and efficient. In turn, this would facilitate fast folding of the fibritin coiled-coil and ensuing fast formation of the whole virus which might be physiologically relevant, considering that the generation cycle of the T4 phage can be as short as twenty minutes.

These results are described in detail in the following publication[173] (see chapter 8.5, p. 164):
Güthe, S.*, Kapinos, L.*, Möglich, A.*, Meier, S., Grzesiek, S., Kiefhaber, T. (2004) Very Fast Folding and Association of a Trimerization Domain from Bacteriophage T4 Fibritin. *J. Mol. Biol.* **337(4)**, 905–915.

*: equally contributing

4. Summary of Unpublished Results

4.1 – Fluorescence Resonance Energy Transfer Studies

4.1.1 – Fluorescence Properties of the Dye Pairs Naphthalene→ Dansyl and Pyrene→Dansyl in Urea Solutions

The effect of the chemical denaturants urea and guanidinium chloride (GdmCl) on loop closure dynamics of unstructured peptides was studied in chapters 3.1 and 3.3. Additionally, fluorescence resonance energy transfer (FRET) was used to examine the changes in the dynamics and average dimensions of model peptides induced by GdmCl (cp. chapter 3.2). In order to also facilitate corresponding FRET studies for the denaturant urea, we determined the fluorescence properties of the dye pairs naphthalene→dansyl and pyrene→dansyl at different urea concentrations.

Fluorescence quantum yields and lifetimes of the chromophores were determined in the donor-only reference peptides, Ac-(GS)$_8$-Nal and Ac-(GS)$_8$-Pyr, displayed in Fig. 7.
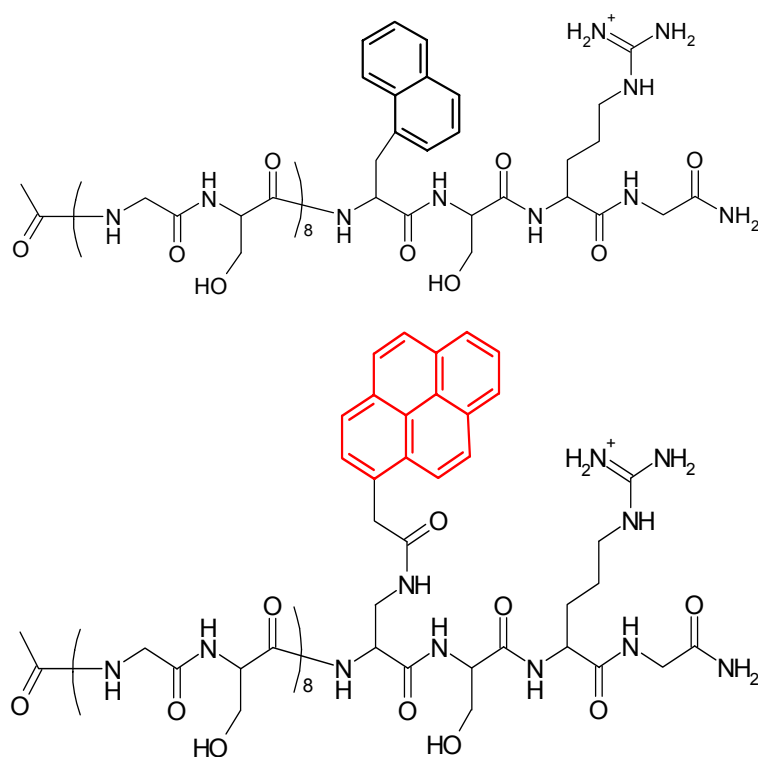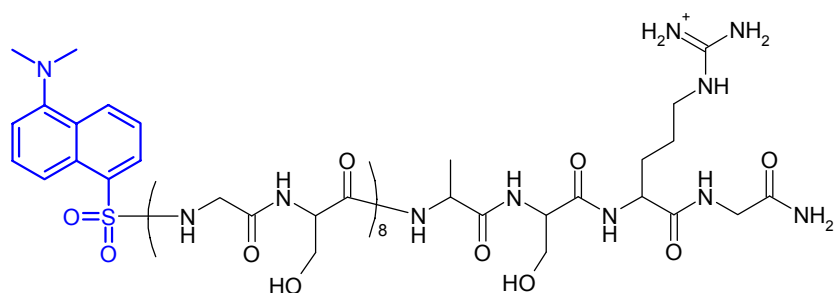
Fig. 7 – Covalent structures of reference peptides used for the determination of fluorescence properties. Donor-only peptides bearing only the FRET donor chromophores naphthalene (top) and pyrene (middle) are designated as Ac-(GS)$_8$-Nal and Ac-(GS)$_8$-Pyr peptides, respectively. The acceptor-only peptide with the dansyl chromophore (Dans-(GS)$_8$-Ala) is shown at the bottom.

Quantum yields were determined in steady-state fluorescence measurements as before (chapter 3.2), fluorescence lifetimes were measured by laser-flash photolysis. Similar to our previous findings for GdmCl (Fig. 8, left), the lifetime of pyrene monotonically decreases with increasing urea concentration (Fig. 8, right) although in a less pronounced manner. In contrast, for naphthalene the fluorescence lifetime shows an increase at low and reaches a plateau value at high urea concentrations.
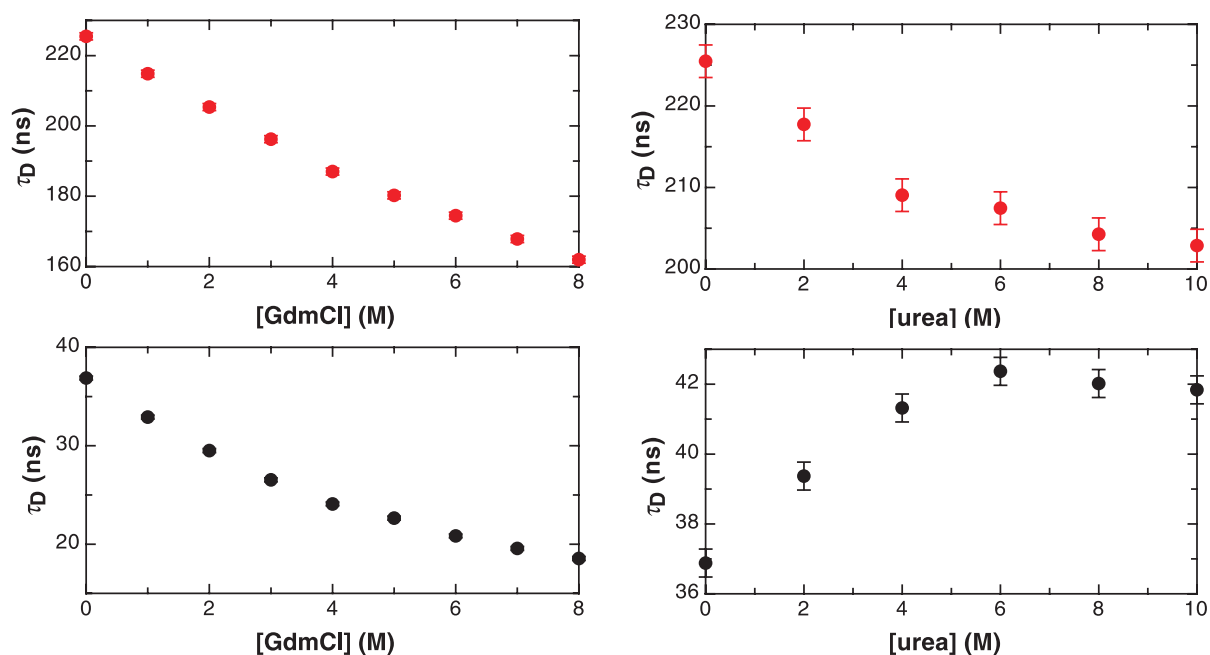
Fig. 8 – Fluorescence lifetimes τ_D of the pyrene (●, Ac-(GS)$_8$-Pyr peptide, Fig. 7) and naphthalene (●, Ac-(GS)$_8$-Nal) fluorophores at different concentrations of GdmCl (left) and urea (right).

Absorbance spectra of the FRET acceptor chromophore dansyl at different urea concentrations were measured in the acceptor-only peptide Dans-(GS)$_8$-Ala shown in Fig. 7. From these data characteristic Förster distances R_0 were calculated according to eq. (45) (cp. chapter 1.7).

$$R_0^6 = \frac{9000 \ln 10}{128\pi^5 N n^4} \kappa^2 \Phi_D \int_0^\infty F_D(\lambda) \varepsilon_A(\lambda) \lambda^4 d\lambda \qquad (45)$$

Within error the R_0 values were found to be independent of the urea concentration (Fig. 9). For the naphthalene→dansyl and pyrene→dansyl pairs average values of 23–24 Å and 20–21 Å were determined.
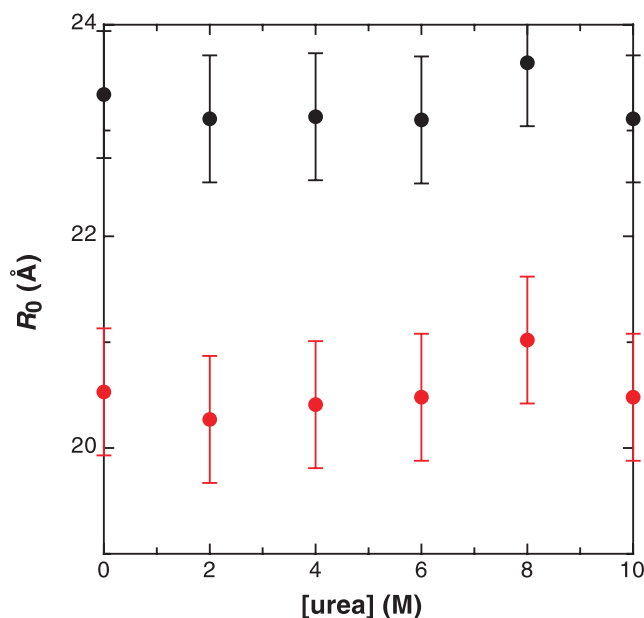

Fig. 9 – Förster distances R_0 for the naphthalene→dansyl (●) and pyrene→dansyl (●) dye pairs as a function of urea concentration.

In summary, the above data show that the two FRET dye pairs used for examining the effect of GdmCl on peptide chain dimensions and dynamics can also be employed for the study of the effect of urea. The Förster distances of the two pairs are similar at all urea concentrations rendering them sensitive for the same distance range. Further, the fluorescence lifetimes of naphthalene and pyrene differ by a factor of 4.9–6.1 depending on the urea concentration, which facilitates to resolve the influence of average distance and motion in time-resolved FRET experiments as described in chapter 3.2.

In preliminary experiments using laser-flash photolysis, time-resolved fluorescence was measured for (GS)$_{14}$ peptides, in which the donor and acceptor chromophores are separated by 14 repeats of the amino acids glycine-serine. Consistent with the findings obtained for GdmCl, an increase of the average end-to-end distance was also observed for urea. Compared to GdmCl, the effect of urea was less pronounced which agrees with the results obtained for intrachain contact formation reactions (chapter 3.1). However, the measurements were compromised by the considerable pulse width of the Nd:YAG laser employed in the flash photolysis setup. Therefore, no reliable quantitative data could be obtained. Using time-

correlated single-photon counting to measure time-resolved fluorescence as in chapter 3.2, it should be possible to acquire more accurate data.

4.1.2 – FRET Dye Pairs with Small Förster Distances

The studies of Krieger *et al.*[54,145,184] on the chain dynamics of unstructured peptides revealed two different regimes for the dependence of intrachain contact formation on peptide length (cp. chapter 1.5, Fig. 2). In short chains the contact formation rate constant, k_c, approaches a limiting value, while in long chains it decreases as a power function of the chain length n. Later studies on the temperature, viscosity and denaturant dependence of contact formation[145] confirmed these two regimes. Contact formation in short peptides seems to be determined by chain stiffness, whereas for longer peptides it appears to be limited by the low probability of forming end-to-end contact. In our previous studies (chapter 3.2) we used FRET to study the dynamics and dimensions of a peptide which had 16 repeats of the amino acids glycine-serine between the fluorophores. This peptide clearly falls in the long-chain regime of contact formation[54]. It is of great interest to also examine chain dynamics and dimensions of peptides in the short-chain limit and to compare the results to our previous findings. However, as detailed in chapter 1.7, FRET measurements are mainly sensitive to distance changes around the Förster distance R_0 which is on the order of 20–23 Å for the dye pairs used in chapter 3.2. Reliable FRET measurements for shorter peptides were precluded due to smaller end-to-end distances and fast chain dynamics. Even for medium-sized peptides with six or eight repeats of glycine-serine between the fluorophores very high FRET efficiencies and accordingly short fluorescence lifetimes were obtained. A quantitative evaluation of distances and dynamics was not possible under these conditions. To overcome these limitations, we established other FRET chromophore systems with smaller R_0 distances, which are better suited for the study of shorter peptides.

4.1.2.1 – Spectroscopic Properties

For practical reasons the same donor fluorescence dyes as used before (chapter 3.2) were also employed for the FRET systems with small Förster distances. Inspection of eq. (45) shows that for the same donor fluorophore the R_0 value can be diminished by either decreasing the donor fluorescence quantum yield or the spectral overlap integral with the acceptor

chromophore. While possible in principle, in practice the first approach is severely hampered by a concomitant decrease of the fluorescence lifetime and signal-to-noise ratio. On the other hand, changing the magnitude of the overlap integral is achieved easily and without these disadvantages by using a different acceptor chromophore.

A suitable acceptor chromophore for the donor naphthalene is 4-(dimethylamino)-phenylacetic acid (DMAP). It can be readily incorporated into peptides synthesized by F-moc solid-phase synthesis (Fig. 10).
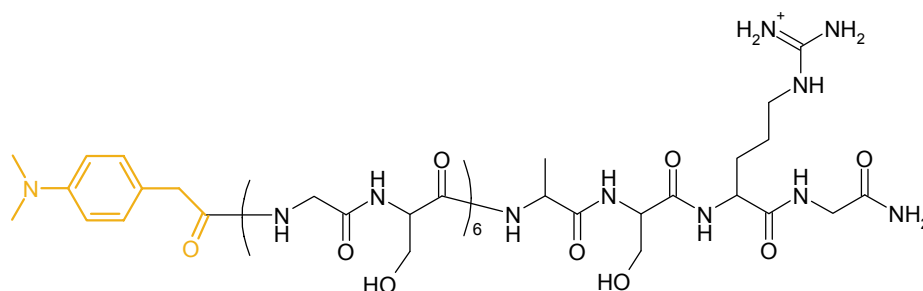
Fig. 10 – Covalent structure of the acceptor-only peptide DMAP-(GS)$_6$-Ala labelled with 4-(dimethylamino)-phenylacetic acid (DMAP).

As shown in Fig. 11, the naphthalene emission and DMAP absorption spectra overlap only to a small extent, which leads to rather ineffective fluorescence resonance energy transfer, i. e. a low R_0 value (see Fig. 14 and below).
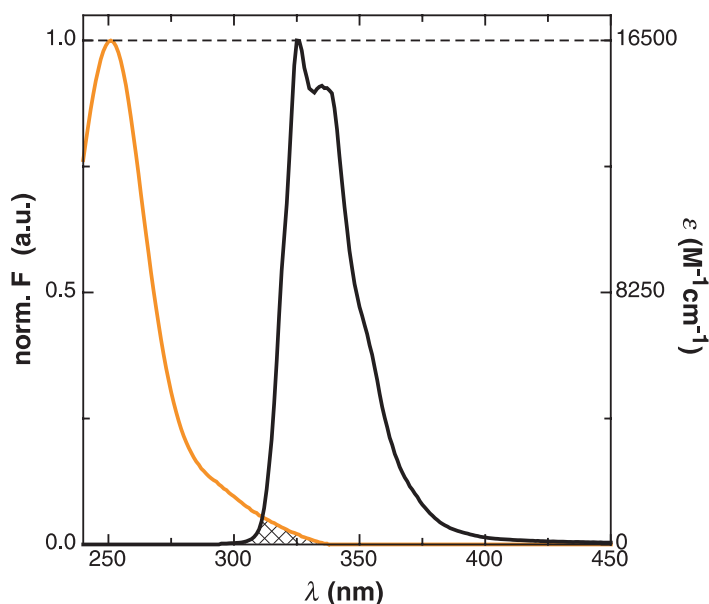
Fig. 11 – Normalized fluorescence emission spectrum of Ac-(GS)$_8$-Nal (—, Fig. 7) and absorption spectrum of DMAP-(GS)$_6$-Ala (—, Fig. 10). The FRET overlap integral shown as the cross-hatched area is rather small.

For the pyrene donor fluorophore, N,N-dimethylanthranilic acid (Anth, 2-(dimethylamino)-benzoic acid) constitutes a suitable acceptor giving rise to a small overlap integral (Fig. 12 and Fig. 13).
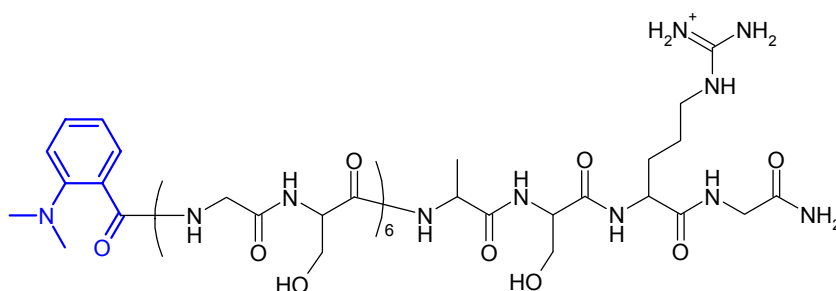

Fig. 12 – Covalent structure of the acceptor-only peptide Anth-(GS)$_6$-Ala labelled with N,N-dimethylanthranilic acid (Anth).
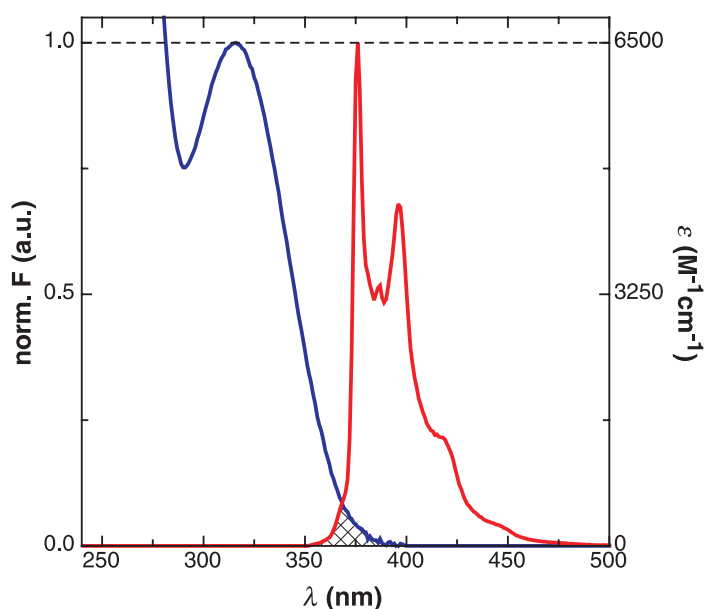

Fig. 13 – Normalized fluorescence emission spectrum of Ac-(GS)$_8$-Pyr (—, Fig. 7) and absorption spectrum of DMAP-(GS)$_6$-Ala (—, Fig. 12). The FRET overlap integral shown as the cross-hatched area is rather small.

As described before in chapter 4.1.1, Förster distances were calculated for the two dye pairs at different concentrations of the denaturants GdmCl and urea (Fig. 14).
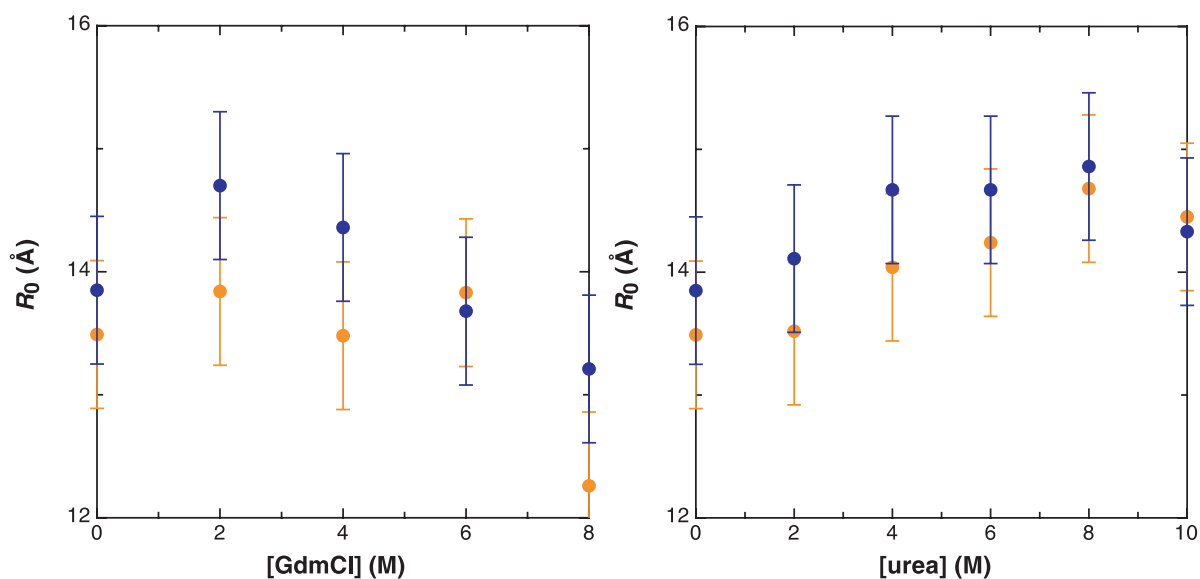
Fig. 14 – Förster distances R_0 for the naphthalene→DMAP (●) and pyrene→Anth (●) dye pairs at different denaturant concentrations.

At all denaturant concentrations the Förster distance is around 12–15 Å for both dye pairs, thus facilitating the study of systems with smaller interchromophore distances. Since the same two donor fluorophores as before were used, the same strategy to resolve average distance and motion in time-resolved FRET experiments as described in chapter 3.2 can be employed.

4.1.2.2 – Synthesis of Fluorescence-labelled Peptides

All peptides were synthesized by F-moc solid-phase peptide synthesis as described before (chapter 3.2). 4-(dimethylamino)-phenylacetic acid was purchased from Fluka (Buchs, Switzerland) and N,N-dimethylanthranilic acid was obtained from ABCR (Karlsruhe, Germany). Both compounds were coupled on resin to the N-termini of peptides using standard F-moc chemistry. Cleavage from the resin, workup and purification of the peptides were conducted as before (chapter 3.2).

4.2 – All-atom Simulations of Peptide Conformations

Following the approach employed by Pappu *et al.*[85], a simulation program was developed to assess the conformational space accessible to unstructured polypeptide chains (cp. chapter 1.2.3). Within the simulation a rather simple strategy is pursued. A large number of peptide

conformations is generated by randomly varying the torsion angles of the peptide backbone and, if applicable, those of amino acid sidechains. Then, the simulated conformers are tested for steric conflicts employing a hard-sphere potential. If any two atoms separated by at least four covalent bonds are closer in distance than their respective contact radius, the corresponding conformation is regarded as sterically not allowed and thus not accessible to the peptide chain. As described in chapters 3.2 and 3.4, results from such simulations were correlated with experimental data on peptide chain dynamics and dimensions measured by TTET and FRET.

In order to perform efficient calculations for longer peptides, the conformational restrictions occurring on the level of the dipeptide unit between neighbouring amino acids have to be known. Corresponding studies are reported in chapter 4.2.1. With these fundamental data sterically allowed conformations of oligopeptides could be evaluated (chapter 4.2.2). Finally, the spectroscopic labels used in the triplet-triplet energy transfer measurements were also incorporated into the simulation program (chapter 4.2.3). Technical details of the program and the general approach are given in chapter 4.2.4.

4.2.1 – Ramachandran Maps

To determine the range of dihedral angles accessible to a given amino acid type **Xaa**, a large number of sterically allowed conformations of tripeptides of the sequence Ser-**Xaa**-Ser was evaluated. From the observed frequency distributions of allowed peptide conformations Ramachandran maps[82,83] were generated as shown in Fig. 15 for glycine (red area).
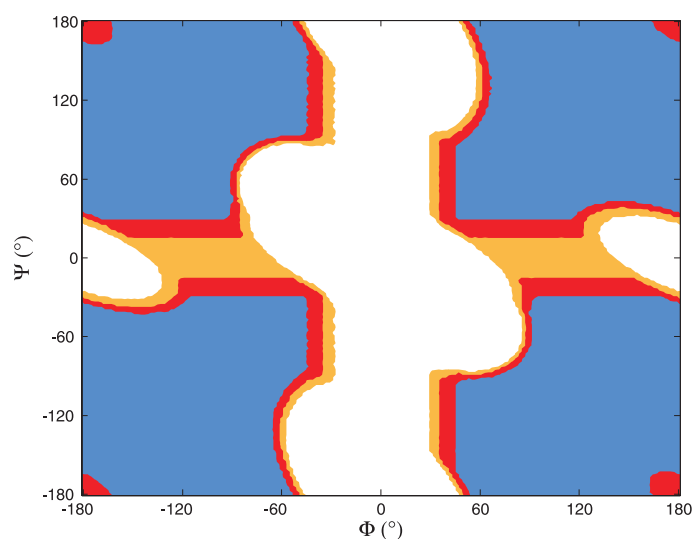

Fig. 15 – Ramachandran map for the glycine residue simulated in a Ser-Gly-Ser peptide context. The sterically allowed regions in dihedral angle space are shown as colored areas and depend on the

choice of the simulation parameters. Red color indicates results obtained for the standard set of parameters used throughout the subsequent studies. Less area is accessible when the atom contact radii are increased by 5 % (blue), more when the τ angle (between atoms N-C$^{\alpha}$-C') is allowed to slightly vary from its optimum value (orange).

The sterically allowed regions depend on the exact choice of the simulation parameters. If the atom contact radii are increased, less conformations are accessible (Fig. 15, blue area). As discussed before by Karplus[188], additional regions of the Ramachandran map become available when the bond angle τ between the atoms N, C$^{\alpha}$ and C' is allowed to slightly deviate from its optimum value of 110.5°.[72] Fig. 15 shows that conformations with a dihedral angle ψ around 0° are sterically allowed (orange) when τ is set to the value of 113.8°.[85] Richards[74] argued that within limits the exact choice of parameters for hard-sphere potentials is not too critical, since the most essential features of atom interactions are reproduced for quite different parameter sets. Accordingly, we found that the simulation results for longer peptides are affected only to a minor degree by the choice of parameters.

Apart from glycine, Ramachandran maps were also generated for proline (cp. chapter 3.4), serine and valine residues. The observed frequencies of peptide conformations within simulation runs were used to assign relative potential energies to the different conformations according to the Boltzmann equation (46).

$$N_1/N_2 = \exp\left(-\left(E_1 - E_2\right)/kT\right)$$ (46)

As shown in Fig. 16, the conformational space for serine and valine is restricted in comparison to glycine.
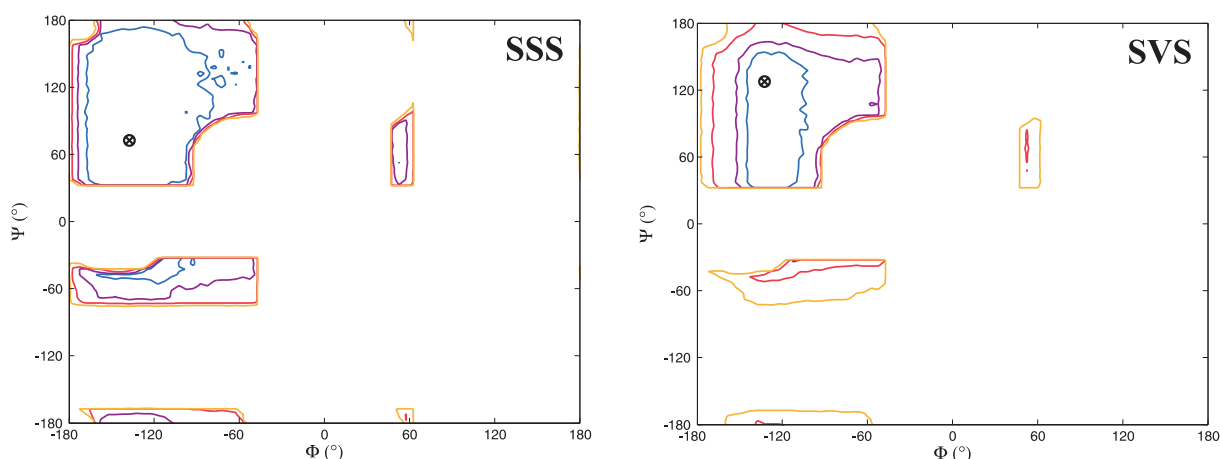

Fig. 16 – Ramachandran maps for serine (left) and valine (right) residues. Colored contour lines denote potential energies of 0.25 (–), 0.5 (–), 1 (–) and 2 kT (–) relative to the most frequent conformation which is indicated by the symbol $\otimes$.

Due to its branched sidechain, for valine even less conformations are accessible than for serine. These results largely correspond to previous observations[82,83]. However, Leach *et al.*[84] reported that the left-handed alpha helix conformation with φ angles around 60° was not amenable for valine. This discrepancy can be ascribed to the fact that in the present simulations also the sidechain dihedral angle χ_1 was varied randomly, whereas before it was fixed in a *staggered* conformation[84]. Still, in both simulations the conformational space for valine is clearly more restricted in comparison to serine. This is also underlined by the finding that during the simulation of a Ser-Val-Ser peptide steric overlap between atoms occurs twice as often as in a Ser-Ser-Ser peptide.

4.2.2 – Simulations of Oligopeptides

4.2.2.1 – Peptides Composed of Glycine and Serine Residues

Knowledge of the Ramachandran maps of the individual amino acid types allows efficient simulations of sterically allowed conformations of oligopeptides. Sterically allowed conformations were simulated for homologous series of poly-(glycine-serine) and poly-serine repeat peptides, which before were also studied experimentally using triplet-triplet energy transfer[54,145] (cp. also chapters 3.1 and 3.2). End-to-end distance probability distributions obtained from these simulations are shown in Fig. 17.
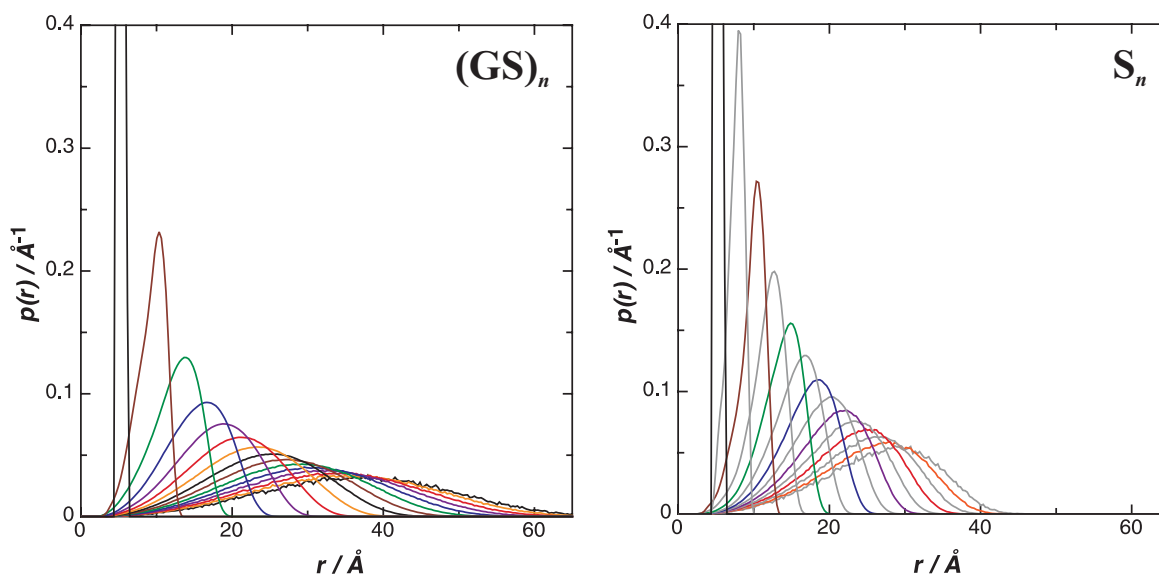

Fig. 17 – End-to-end distance probability distributions $p(r)$ for glycine-serine and serine repeat peptides obtained from all-atom simulations of sterically allowed conformations. (left) Simulation results for peptides with the canonical sequence (Gly-Ser)$_n$ (n = 1 .. 15) are shown (from left to right).

(right) Corresponding distribution functions for oligoserine peptides of the sequence $(Ser)_n$ (n = 2 .. 15). Peptides of the same length are coloured alike in the left and the right panel.

As expected from polymer theory, longer peptides show successively larger end-to-end distances and broader distribution functions. The mean square end-to-end distance $<r^2>$ of the simulated distance distributions scales as a power function of peptide length n according to eq. (47) (Fig. 18, left).

$$\left\langle r^2 \right\rangle \propto n^\nu \tag{47}$$

Oligoserine peptides show a stronger increase of the average distance with peptide length ($\nu = 1.48 \pm 0.02$) than the corresponding oligo-(glycine-serine) peptides ($\nu = 1.35 \pm 0.01$).
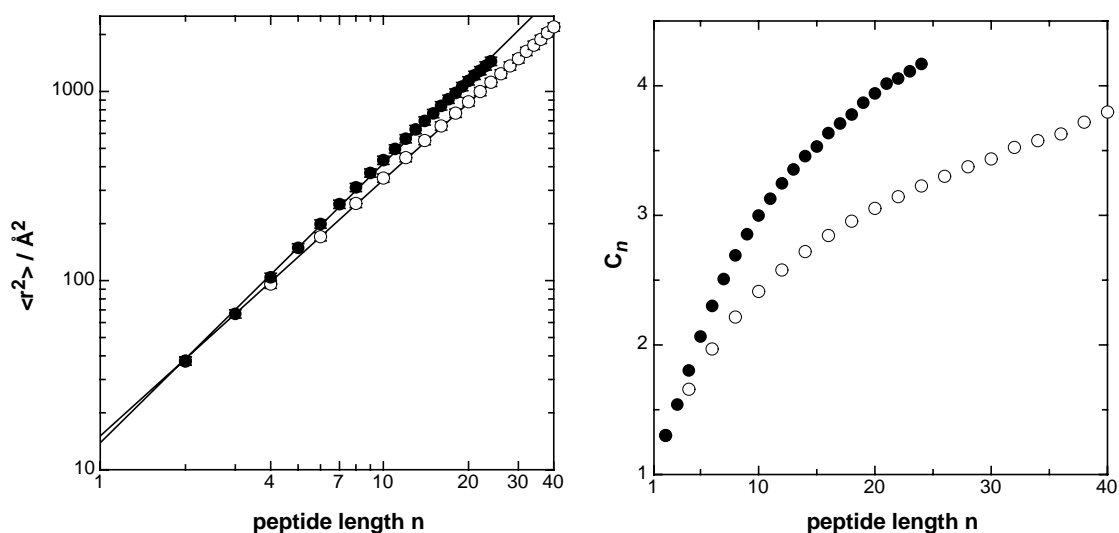

Fig. 18 – (left) Mean square end-to-end distance $<r^2>$ in simulated oligo-(glycine-serine) (○) and oligoserine (●) peptides as a function of peptide length n. The correlation plots have slopes of 1.35 ± 0.01 and 1.48 ± 0.02 for the glycine-serine and the serine peptides, respectively. (right) Corresponding characteristic ratios C_n were calculated according to eq. (48).

In order to compare the simulation results to predictions from polymer theory, characteristic ratios C_n were calculated according to eq. (48) (cp. chapter 1.2.3).

$$C_n = \left\langle r^2 \right\rangle \Big/ nl^2 \tag{48}$$

As shown in Fig. 18 (right), the characteristic ratio increases hyperbolically with peptide length n. Oligoserine peptides have higher values of C_n than oligo-(glycine-serine) peptides of corresponding length. Based on Flory's polymer theory[61], lower values for the characteristic ratio of glycine-serine peptides are expected. In the limit of very long chains a value of $C_\infty \approx 3.1$ is predicted. However, Flory's theory applies to so-called Θ-conditions, i. e. a discrete set of conditions (temperature, pressure etc.) where polymers behave like unperturbed

chains (cp. chapter 1.2.3). The deviating simulation results could be due to long-range interactions and excluded volume effects, both of which are not included in Flory's basic theory. On the other hand, for oligoserine peptides the simulated characteristic ratios are smaller than predicted from theory ($C_\infty \approx 9.4$). Flory[61] pointed out that besides steric interactions also electrostatic interactions need to be taken into account to accurately describe the conformation of non-glycine peptides. As of now only pairwise steric interactions are considered in the simulation software, but it should be straightforward to also include other kinds of interactions. For example, Pappu and Rose[88] extended their basic simulation[85] and replaced the previously used hard-sphere potential with a Lennard-Jones potential. Further, it might be worthwhile to derive the underlying Ramachandran maps from force fields used in molecular dynamics simulations[189]. Recent work by Jha *et al.*[190,191] illustrates that the conformational properties of unfolded proteins can be quite successfully modelled based on structure data found in the Brookhaven protein data base (PDB)[192]. Very recently, Blackledge and coworkers employed a closely related approach[193].

Despite the shortcomings of the simulations, we compared the results to experimental data on end-to-end contact formation in model peptides[54,145]. Jacobsen and Stockmayer[156] showed that for an ideal polymer chain the probability of forming end-to-end contact, p_c, should scale with the mean square end-to-end distance $<r^2>$ as given in eq. (49) (cp. also chapter 1.6).

$$p_c \propto \left\langle r^2 \right\rangle^{-3/2} \tag{49}$$

We therefore correlated $<r^2>^{-3/2}$ with peptide length n as shown in Fig. 19.
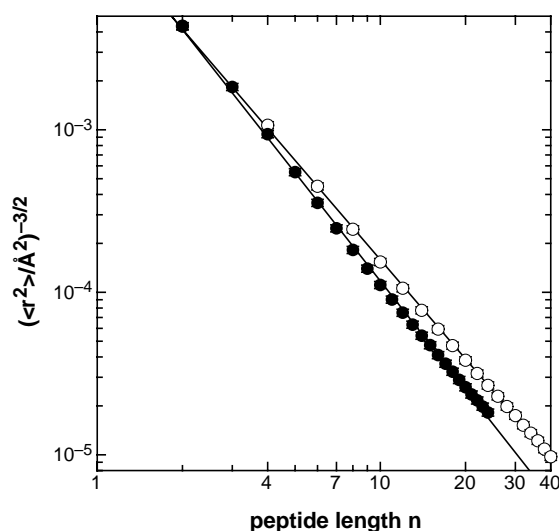

Fig. 19 – Correlation of the inverse third power of the root mean square end-to-end distance with peptide length *n*. The correlation plots have slopes of -2.0 ± 0.1 for oligo-(glycine-serine) (○) and of -2.2 ± 0.1 for oligoserine peptides (●).

Similar to the experimental data on end-to-end contact formation in long peptides[54] (cp. chapter 1.5, Fig. 2), $<r^2>^{-3/2}$ decreases as a power function of peptide length n. However, the correlation coefficients of -2.0 ± 0.1 and -2.2 ± 0.2 found for glycine-serine and serine peptides indicate a stronger length dependence for $<r^2>^{-3/2}$ than seen for the experimentally determined rate constant of end-to-end contact formation. In agreement with the experimental findings, also the simulation data predict a stronger length dependence for the oligoserine than for the oligo-(glycine-serine) peptides. For short peptides the simulation results cannot account for the experimental data on contact formation. As discussed in chapter 1.5, for short peptides the rate constant of end-to-end contact formation becomes nearly independent of peptide length. In this regime contact formation is no longer mainly limited by the probability of forming a contact but by other processes. Consistently, Krieger et al.[145] observed high activation energies and weak viscosity dependencies for contact formation in short peptides.

4.2.2.2 – Oligoproline Peptides

Proline strongly differs from the other amino acids naturally occurring in proteins regarding its structure and properties. The proline sidechain is linked to its amide nitrogen atom and thus forms a pyrrolidine ring. As a consequence, peptide bonds preceding proline residues adopt the *cis* conformation to significantly higher extents than non-prolyl bonds[194]. Proline residues are known to severely restrict the conformational space accessible to a polypeptide and to therefore increase chain stiffness[61]. Consistently, peptide motions and dynamics were also found to be strongly influenced by proline residues (cp. chapter 3.4). All-atom models of proline-containing peptides were used to rationalize the experimental data on loop closure kinetics.

Due to their relatively stiff structure, oligoproline peptides have been used as reference systems in a number of studies (for recent applications see e. g.[195,196,197]). The conformation of oligoproline peptides depends on solvent conditions and temperature[198,199]. Usually, in aqueous solution proline peptides are assumed to adopt an extended conformation corresponding to the polyproline II helix. Recently however, the validity of this assumption has been challenged by the work of Schuler et al.[200,201] Based on single-molecule and ensemble FRET measurements and simulation data, they argued that oligoproline peptides are considerably more flexible than it was previously assumed. Since the conformational space available to proline-containing peptides is mainly limited by steric interactions between

atoms[61], the application of hard-sphere simulations to these systems seems to be particularly rewarding.

Therefore, sterically allowed conformations of oligoproline peptides of varying length were simulated. In order to verify the accuracy of these simulations, in a first step we studied the conformational space of short proline peptides and compared our results to the experimental data reported by Stryer and Haugland[65]. In their study oligoproline peptides were labelled with the chromophores dansyl and naphthyl (see Fig. 20), and distances between the labels were assessed by FRET.
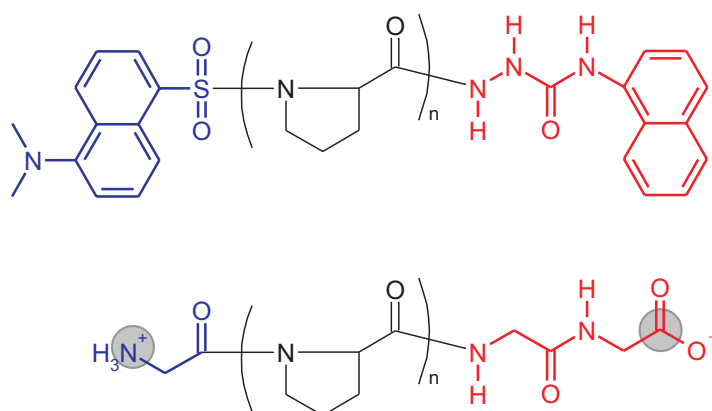

Fig. 20 – Covalent structure of the oligoproline peptides used in the study of Stryer and Haugland[65] (top) and in corresponding all-atom simulations (bottom). *n* denotes the number of proline residues, gray circles indicate atoms between which end-to-end distances were calculated.

To facilitate the all-atom simulations, in the calculations the spectroscopic labels were approximated by glycine residues as indicated in Fig. 20. End-to-end distances were calculated for the simulated peptide conformations as the distance between the N-terminal amide nitrogen and the C-terminal carbonyl carbon atom. Due to their stiffness the peptides comprising between 1 and 12 proline residues show an almost linear increase of the average end-to-end distance with peptide length (Fig. 21, left).
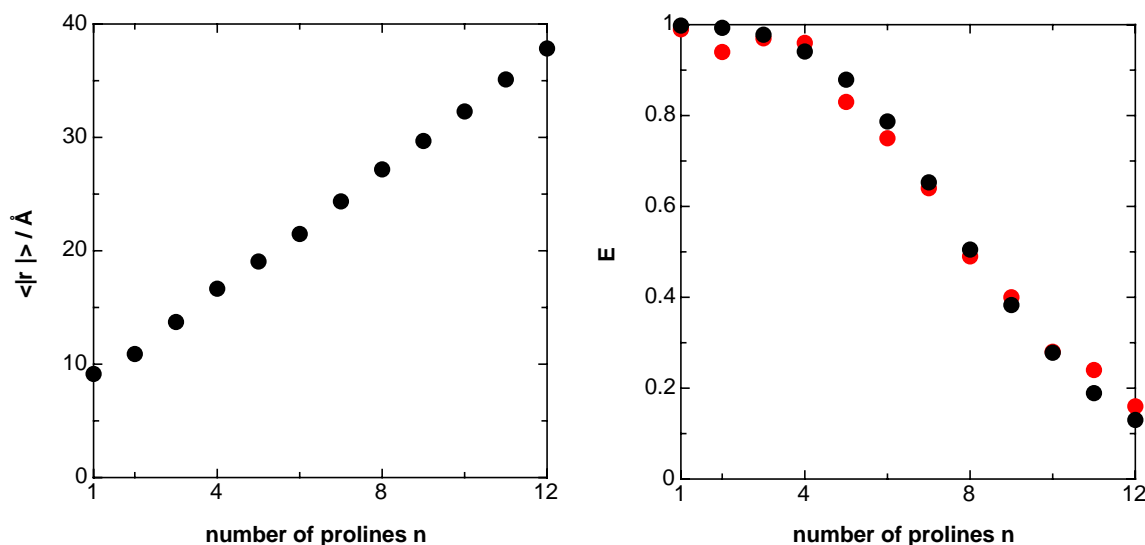
Fig. 21 – Results from all-atom simulations for the oligoproline peptides shown in Fig. 20. (left) Mean end-to-end distance <|r|> as a function of peptide length *n*. (right) Comparison of FRET efficiencies *E* based on simulation results and calculated according to eq. (50) (●) with the experimental values by Stryer and Haugland[65] (●).

Using the reported value[65] of 27.2 Å for the Förster distance R_0, efficiencies of energy transfer *E* were calculated for the simulated end-to-end distance distributions $p(r)$ according to eq. (50) (cp. chapter 1.7).

$$\langle E \rangle = \int_0^\infty p(r) \frac{R_0^6}{R_0^6 + r^6} dr \tag{50}$$

Despite the approximations of the simulation approach, the calculated transfer efficiencies *E* are found to be in good agreement with the experimental ones (Fig. 21, right). However, the simulated values seem to show a slightly steeper dependence on peptide length than the experimental data. This could be an indication that the oligoproline peptides are slightly more flexible in reality than observed in the simulations.

To compare our results to those of Schuler *et al.*[200,201], we simulated sterically allowed conformations of oligoproline peptides consisting of up to 22 proline residues. The observed end-to-end distances are narrowly distributed for all oligoproline peptides (Fig. 22, left). For the longer peptides somewhat wider distance distributions are found which can be attributed to slight bending of the peptide chain.
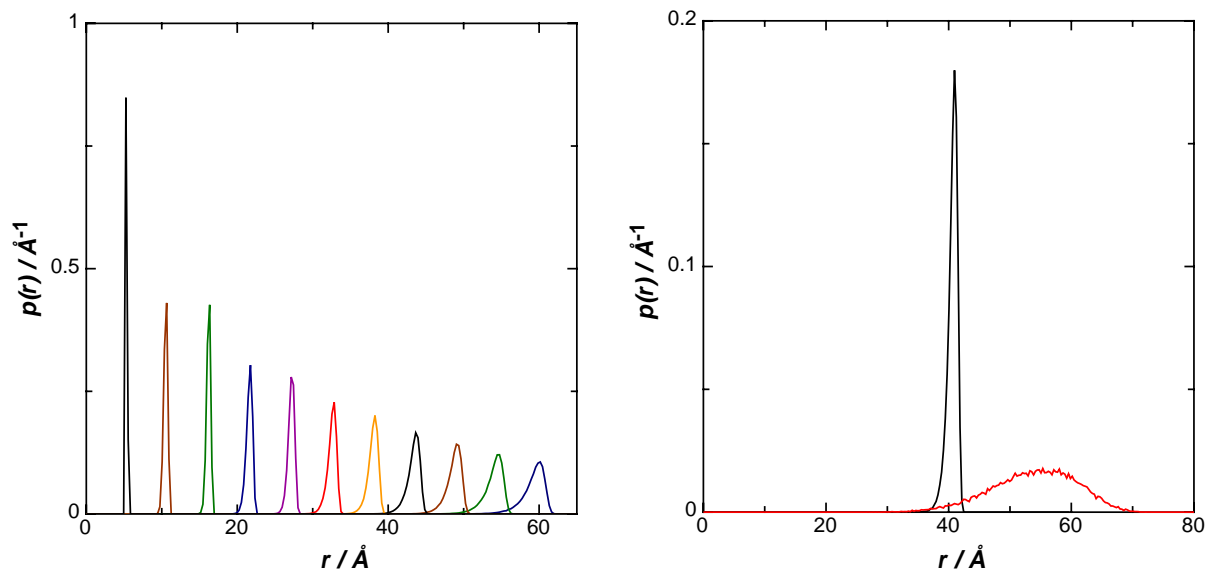
Fig. 22 – All-atom simulations of oligoproline peptides P_x. (left) End-to-end distance probability distributions for peptides with the sequence P_2, P_4, P_6, P_8, P_{10}, P_{12}, P_{14}, P_{16}, P_{20} and P_{22} (from left to right). (right) End-to-end distances found for a P_{15} peptide (–) and a $(GS)_3P_{15}(GS)_2$ peptide (–).

In contrast, Schuler et al.[201] published molecular dynamics simulation data which imply that oligoproline peptides are much more flexible. Their reported end-to-end distance probability distributions are consistently much broader than in comparable simulations performed by us. Further they are shifted towards smaller end-to-end separations. For example, a peptide of the sequence Gly-$(Pro)_{15}$-Cys was found to have end-to-end distances as low as about 23 Å. In contrast, in our simulations a $(Pro)_{15}$ peptide displays a very narrow distance distribution centered around 40 Å (Fig. 22, right).

Schuler and coworkers[201] compared their simulation results to experimental data obtained from single-molecule and ensemble FRET measurements. They found that the FRET efficiencies observed in oligoproline peptides could not be explained by assuming oligoprolines to form a rigid rod. Instead, much better agreement between experiment and theory was obtained when the simulated end-to-end distance distributions were used to calculate FRET efficiencies. However, it is interesting to note, that even when using this approach, for long proline peptides the experimentally observed transfer efficiencies are systematically higher than the predicted ones.

As discussed by Schuler et al.[201], their FRET measurements are complicated by several problems. First, the fluorescence dyes used in their study have lifetimes of only 3 ns and lower (depending on peptide context). Consequently, the orientations of donor and acceptor fluorophores are not averaged out during the experiment which is also evident from large fluorescence anisotropies. It is well known that insufficient averaging of the fluorophore

orientations severely affects FRET transfer efficiencies[202,203]. Further problems arise from the fact that the fluorophores are quite large and are attached to long linkers. Thus, it is not clear which distances the FRET measurements exactly report upon. Introduction of flexible linkers leads to a considerable broadening of the end-to-end distance distribution as illustrated in Fig. 22 (right).

In summary, it remains unclear whether the simulations reported by Schuler *et al.*[201] accurately describe the behaviour and conformation of oligoproline peptides. For the longer peptides our simulations would predict stiffer and more extended chains. At least for short peptides our simulation data are in excellent agreement with experimental results[65]. Clearly, for the longer oligoproline peptides additional explanations are required to account for the experimentally observed FRET efficiencies. As discussed above, the deviations between model and experiment could be partly due to the flexibility of oligoproline peptides. Another possible explanation could be the occurrence of *cis* peptide bonds which would introduce kinks into the peptide chain. Reimer *et al.*[194] found that the Pro-Pro amide bond in a short model peptide shows approximately 6 % *cis* conformation. While this number might be lower within the context of a polyproline II helix, for the long peptides *cis* bonds might still occur with significant frequency. Krieger and Kiefhaber (unpublished results) studied the chain dynamics of short oligoproline peptides using TTET and observed multiphasic contact formation kinetics. This indicates conformational heterogeneity in these peptides which could be caused by the occurrence of *cis* proline bonds. On the other hand, a high fraction of oligoproline peptides with amide bonds in the *cis* conformation would contradict the experimental results by Stryer and Haugland[65]. Further studies are required to resolve this problem.

4.2.3 – Simulations of Peptides with TTET Labels

In order to facilitate better comparison of the all-atom simulations with experimental data obtained by triplet-triplet energy transfer (TTET), the simulation program was extended to also include the spectroscopic probes naphthalene and xanthonic acid used in the TTET measurements. Bond lengths and angles for the triplet labels were derived from literature data on high-resolution X-ray structures[204,205] (Fig. 23).
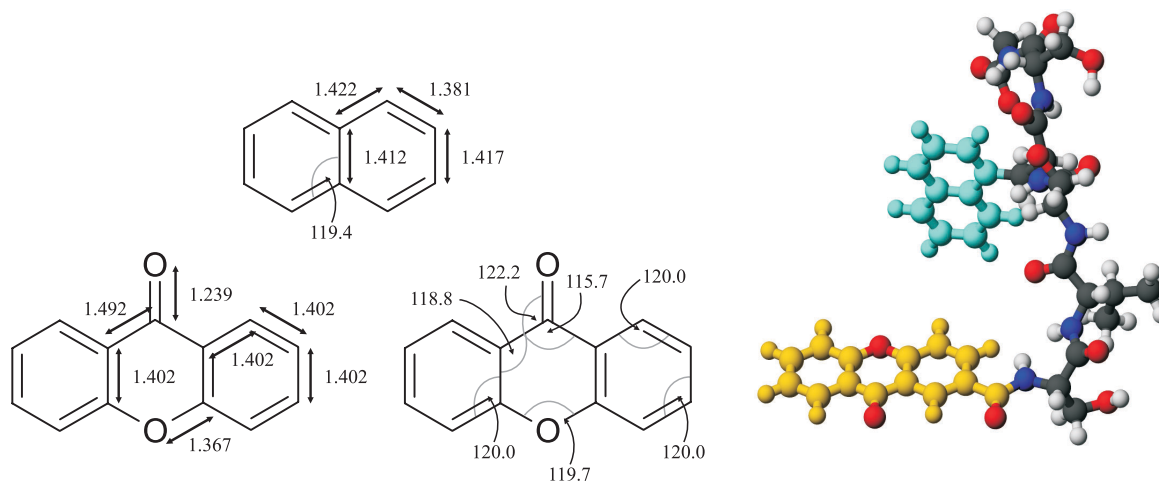
Fig. 23 – Molecular geometries of the triplet labels naphthalene (Nal, top) and xanthone (Xan, bottom) used in the all-atom simulations. Distance and angle values are given in units of Å and degree, respectively, and are derived from X-ray crystallography data[204,205]. The picture on the right-hand side shows a simulated conformation of a peptide with the sequence Xan-Ser-Val-Ser-Nal-Ser-Gly-COOH. The graphic was generated with the programs MOLMOL[171] and PovRay[172].

In chapter 3.2 we simulated sterically allowed conformations of a $(GS)_{16}$ peptide which consisted of 16 repeats of the dipeptide Gly-Ser and had the triplet labels naphthalene and xanthonic acid attached to its termini. The results were found to be consistent with distance measurements by FRET in GdmCl solutions.

Especially for short peptides it seems to be important to include the spectroscopic labels in simulations when comparing the results to corresponding experimental data. Within these peptides the labels constitute a large fraction of the whole molecule (cp. Fig. 23, right) and, accordingly, their influence on peptide conformation should not be neglected.

4.2.4 – Implementation of the Simulation Program

The basic principles of the simulation software were described before[184] (see chapter 3.4). In short, the simulation program was written in ANSI-C[206] and compiled with the Bloodshed Dev-C++[207] development environment. As in the work of Pappu *et al.*[85], atom contact radii were obtained from Hopfinger[73] and multiplied by 0.95. Bond lengths and angles for the standard amino acids were derived from high-resolution X-ray structures as reported in the compilation of Engh and Huber[72]. Geometries of the TTET labels were approximated by the reported X-ray data for naphthalene and xanthen-9-one (Landolt-Börnstein II/15[204] p. 577 and II/21[205] pp. 442–443). In the simulations torsion angles ω were fixed in the *trans* (179.5°) or *cis* (-0.5°) conformation. Dihedral angles φ, ψ, χ_1 were varied randomly except for φ of

proline residues which was fixed at -60°.[66] Pseudorandom numbers were generated using an implementation of the generalized feedback shift register algorithm[208]. Simulated peptide conformations were checked for steric overlap between atoms.

5. Summary

This thesis aimed at improving our knowledge about the dynamic and structural properties of the unfolded states of proteins. Unstructured peptides were used as model systems to probe the structure and the dynamics of such states. Special emphasis was put on the influence of chemical denaturants as they are widely used in protein folding and stability studies.

Triplet-triplet energy transfer (TTET) between xanthonic acid and naphthalene occurs in a diffusion-controlled manner and is thus well suited to measure absolute contact formation rate constants. Using TTET, intrachain loop closure dynamics in unstructured peptides were studied. We found that these dynamics are decelerated by the addition of the denaturants urea and guanidinium chloride (GdmCl). Closer inspection revealed that this behaviour has contributions from two separate effects. First, denaturants increase solvent viscosity which impedes motions of the peptide chain. Secondly, denaturants bind to peptides and thereby further slow chain dynamics. Denaturant binding was found to be rather weak and transient. Urea and GdmCl mainly differ in their binding strength but not in their effect on chain dynamics once bound. This implies similar mechanisms of denaturation for these two compounds.

In order to understand the effect of denaturants on the dynamics of unfolded polypeptide chains in more detail, we studied loop closure dynamics in a series of host-guest peptides. Different 'guest' amino acids were incorporated into common 'host' peptide contexts to assess their influence on peptide dynamics. For most peptides the effect of urea and GdmCl on intrachain contact formation was similar, indicating a common mechanism for denaturant action. These findings imply that chemical denaturants mainly interact with the polypeptide backbone. Lysine and arginine displayed significantly weaker interactions with GdmCl than all other residues, which is most likely due to charge-charge repulsion between the guanidinium cation and the basic amino acid sidechain.

To obtain information about the structure of unfolded peptides and thus deepen our understanding of their behaviour, we employed fluorescence resonance energy transfer (FRET) as a second optical spectroscopic technique. Using FRET, we studied the dynamics and dimensions of unstructured peptides at different GdmCl concentrations. The FRET measurements on these unstructured peptides were complicated by fast conformational rearrangements occurring during the experimental observation time. We developed a strategy to accurately account for these dynamics and to reliably analyze the experimental data. In

separate experiments two chromophore pairs were used, which had largely different donor fluorescence lifetimes but were sensitive to the same distance range. Data obtained on both systems were globally evaluated and allowed to assess the influence of fast dynamics on the measured FRET signal. In principle, this strategy could also be applied to the study of other flexible systems. The analysis of the FRET data revealed that the end-to-end distance of unstructured peptides increases with increasing denaturant concentration. These results show that unstructured polypeptides can undergo significant changes in their dimensions upon changes of solvent composition. GdmCl further exerted a profound effect on the internal dynamics of peptide chains. High denaturant concentrations facilitated conformational rearrangements of the peptide chain, which was only partly offset by the concomitant increase of solvent viscosity. Taken together, these data confirmed that concentrated GdmCl solutions are good solvents for polypeptides and water is a bad one. Results from FRET and TTET measurements were found to be in good agreement. The denaturant-induced changes in intrachain contact formation rates, average dimensions and internal dynamics of the peptides were clearly correlated. Comparison of the results obtained with the two spectroscopic techniques further showed that the loop closure dynamics of long peptides can be sufficiently described by a theory for polymer dynamics by Szabo *et al.*[157]

All-atom simulations of sterically allowed peptide conformations were used to rationalize experimental observations on the behaviour of unstructured peptides. Comparison of simulation results with TTET experiments on proline-containing peptides provided an explanation for the different chain dynamics observed for *cis*- and *trans*-proline peptide isomers. Further, we compared the simulations to our FRET results. The average dimensions found for simulated peptide conformations closely corresponded to the dimensions experimentally observed for peptides in concentrated GdmCl solutions. Since in the simulations no intramolecular interactions except for excluded volume effects were considered, this could mean that binding of GdmCl to peptides largely abolishes intramolecular interactions. In the absence of these interactions the polypeptide chain would be more extended and able to undergo faster conformational rearrangements.

A last part of this thesis dealt with the folding reaction of the trimeric foldon domain from T4 phage fibritin. Global evaluation of refolding experiments conducted at different protein concentrations enabled us to determine the minimal folding mechanism. Our data analysis provided an explanation for the high stability and fast assembly of the foldon trimer.

6. Acknowledgements

7. **References**

1.   Jackson, S. E. (1998). How do small single-domain proteins fold? *Fold. Des.* **3**, R81-R91.

2.   Prusiner, S. B. (1998). Prions. *Proc. Natl. Acad. Sci. USA* **95**, 13363-13383.

3.   Anson, M. L. & Mirsky, A. E. (1931). The reversibility of protein coagulation. *J. Phys. Chem.* **35**, 185-193.

4.   Mirsky, A. E. & Pauling, L. (1936). On the structure of native, denatured, and coagulated proteins. *Proc. Natl. Acad. Sci. USA* **22**, 439-447.

5.   Anfinsen, C. B., Haber, E., Sela, M. & White, F. H. (1961). The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. USA* **47**, 1309-1314.

6.   Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science* **181**, 223-230.

7.   Walter, S. & Buchner, J. (2002). Molecular chaperones - cellular machines for protein folding. *Angew. Chem. Int. Ed. Engl.* **41**, 1098-1113.

8.   Bardwell, J. C. & Collet, J.-F. (2005). The catalysis of disulfide bond formation in prokaryotes. In *Protein Folding Handbook* (Buchner, J. & Kiefhaber, T., eds.), Vol. 3, pp. 358-376. 5 vols. Wiley-VCH, Weinheim, Germany.

9.   Lang, K., Schmid, F. X. & Fischer, G. (1987). Catalysis of protein folding by prolyl isomerases. *Nature* **329**, 268-270.

10. Fischer, G. (2005). Catalysis of peptidyl-prolyl cis/trans isomerization by enzymes. In *Protein Folding Handbook* (Buchner, J. & Kiefhaber, T., eds.), Vol. 3, pp. 377-414. 5 vols. Wiley-VCH, Weinheim, Germany.

11. 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.

12. Creighton, T. E. (1993). *Proteins: structures and molecular properties*, W. H. Freeman, New York, USA.

13. Israelachvili, J. N. (1991). *Intermolecular and surface forces*, Academic Press, London, UK.

14. Kauzmann, W. (1959). Factors in the interpretation of protein denaturation. *Adv. Prot. Chem.* **14**, 1-63.

15. Baldwin, R. L. (2005). Weak Interactions in Protein Folding: Hydrophobic Free Energy, van der Waals Interactions, Peptide Hydrogen Bonds, and Peptide Solvation. In *Protein*

Folding Handbook (Buchner, J. & Kiefhaber, T., eds.), Vol. 1, pp. 127-162. 5 vols. Wiley-VCH, Weinheim, Germany.

16. Avbelj, F. (2000). Amino acid conformational preferences and solvation of polar backbone atoms in peptides and proteins. *J. Mol. Biol.* **300**, 1335-1359.

17. Avbelj, F., Luo, P. & Baldwin, R. L. (2000). Energetics of the interaction between water and the helical peptide group and its role in determining helix propensities. *Proc. Natl. Acad. Sci. USA* **97**, 10786-10791.

18. Avbelj, F. & Baldwin, R. L. (2002). Role of backbone solvation in determining thermodynamic beta propensities of the amino acids. *Proc. Natl. Acad. Sci. USA* **99**, 1309-1313.

19. Avbelj, F. & Baldwin, R. L. (2003). Role of backbone solvation and electrostatics in generating preferred peptide backbone conformations: distributions of phi. *Proc. Natl. Acad. Sci. USA* **100**, 5742-5747.

20. Kendrew, J. C., Dickerson, R. E. & Strandberg, B. E. (1960). Structure of myoglobin: a three-dimensional Fourier synthesis at 2 Å resolution. *Nature* **185**, 422-427.

21. Blake, C. C., Koenig, D. F., Mair, G. A., North, A. C., Phillips, D. C. & Sarma, V. R. (1965). Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Å resolution. *Nature* **206**, 757-763.

22. Cavanagh, J., Skelton, N. J., Palmer, A. G. & Fairbrother, W. (1996). *Protein NMR spectroscopy*, Academic Press, London, UK.

23. Zeeb, M. & Balbach, J. (2005). Kinetic Protein Folding Studies using NMR Spectroscopy. In *Protein Folding Handbook* (Buchner, J. & Kiefhaber, T., eds.), Vol. 1, pp. 536-572. 5 vols. Wiley-VCH, Weinheim, Germany.

24. Linderstrøm-Lang, K. (1955). Deuterium exchange between peptides and water. *Chem. Soc. Spec. Publ.* **2**, 1-20.

25. Wagner, G. & Wüthrich, K. (1982). Amide protein exchange and surface conformation of the basic pancreatic trypsin inhibitor in solution. Studies with two-dimensional nuclear magnetic resonance. *J. Mol. Biol.* **160**, 343-361.

26. Englander, S. W. (2000). Protein folding intermediates and pathways studied by hydrogen exchange. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 213-238.

27. Roder, H., Elöve, G. A. & Ramachandra Shasty, M. C. (2000). Early stages of protein folding. In *Mechanisms of Protein Folding* (Pain, R. H., ed.), pp. 65-104. Oxford University Press, Oxford, UK.

28. Kay, L. E. (1998). Protein dynamics from NMR. *Nature Struct. Biol.* **5 Suppl.**, 513-517.

29. Ishima, R. & Torchia, D. A. (2000). Protein dynamics from NMR. *Nature Struct. Biol.* **7**, 740-743.

30. Eisenmesser, E. Z., Bosco, D. A., Akke, M. & Kern, D. (2002). Enzyme dynamics during catalysis. *Science* **295**, 1520-1523.

31. 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**, 117-121.

32. Anderson, C. M., Stenkamp, R. E., McDonald, R. C. & Steitz, T. A. (1978). A refined model of the sugar binding site of yeast hexokinase B. *J. Mol. Biol.* **123**, 207-219.

33. Bennett, W. S., Jr. & Steitz, T. A. (1978). Glucose-induced conformational change in yeast hexokinase. *Proc. Natl. Acad. Sci. USA* **75**, 4848-4852.

34. Ansari, A., Berendzen, J., Bowne, S. F., Frauenfelder, H., Iben, I. E., Sauke, T. B., Shyamsunder, E. & Young, R. D. (1985). Protein states and proteinquakes. *Proc. Natl. Acad. Sci. USA* **82**, 5000-5004.

35. Ansari, A., Jones, C. M., Henry, E. R., Hofrichter, J. & Eaton, W. A. (1992). The role of solvent viscosity in the dynamics of protein conformational changes. *Science* **256**, 1796-1798.

36. Tanford, C. (1968). Protein denaturation. *Adv. Prot. Chem.* **23**, 121-282.

37. Tanford, C. (1970). Protein denaturation. Part C. Theoretical models for the mechanism of denaturation. *Adv. Prot. Chem.* **24**, 1-95.

38. Damaschun, G., Damaschun, H., Gast, K. & Zirwer, D. (1998). Denatured states of yeast phosphoglycerate kinase. *Biochemistry (Mosc.)* **63**, 259-275.

39. Kohn, J. E., Millett, I. S., Jacob, J., Zagrovic, B., Dillon, T. M., Cingel, N., Dothager, R. S., Seifert, S., Thiyagarajan, P., Sosnick, T. R., Hasan, M. Z., Pande, V. S., Ruczinski, I., Doniach, S. & Plaxco, K. W. (2004). Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci. USA* **101**, 12491-12496.

40. Wu, D., Chen, A. & Johnson Jr., C. S. (1995). An improved diffusion-ordered spectroscopy experiment incorporating bipolar-gradient pulses. *J. Magn. Reson. A* **115**, 260-264.

41. Tanford, C., Kawahara, K. & Lapanje, S. (1966). Proteins in 6 M Guanidine Hydrochloride. *J. Biol. Chem.* **241**, 1921-1923.

42. Neri, D., Billeter, M., Wider, G. & Wuthrich, K. (1992). NMR determination of residual structure in a urea-denatured protein, the 434-repressor. *Science* **257**, 1559-1563.

43. Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L. J., Dobson, C. M. & Schwalbe, H. (2002). Long-range interactions within a nonnative protein. *Science* **295**, 1719-1722.

44. Dyson, H. J. & Wright, P. E. (2004). Unfolded proteins and protein folding studied by NMR. *Chem. Rev.* **104**, 3607-3622.

45. Sánchez, I. E. & Kiefhaber, T. (2003). Hammond behavior versus ground state effects in protein folding: evidence for narrow free energy barriers and residual structure in unfolded states. *J. Mol. Biol.* **327**, 867-884.

46. Tiffany, M. L. & Krimm, S. (1968). New chain conformations of poly(glutamic acid) and polylysine. *Biopolymers* **6**, 1379-1382.

47. Tiffany, M. L. & Krimm, S. (1968). Circular dichroism of poly-L-proline in an unordered conformation. *Biopolymers* **6**, 1767-1770.

48. Rucker, A. L., Pager, C. T., Campbell, M. N., Qualls, J. E. & Creamer, T. P. (2003). Host-guest scale of left-handed polyproline II helix formation. *Proteins* **53**, 68-75.

49. Whittington, S. J., Chellgren, B. W., Hermann, V. M. & Creamer, T. P. (2005). Urea promotes polyproline II helix formation: implications for protein denatured states. *Biochemistry* **44**, 6269-6275.

50. Shi, Z., Olson, C. A., Rose, G. D., Baldwin, R. L. & Kallenbach, N. R. (2002). Polyproline II structure in a sequence of seven alanine residues. *Proc. Natl. Acad. Sci. USA* **99**, 9190-9195.

51. Shi, Z., Woody, R. W. & Kallenbach, N. R. (2002). Is polyproline II a major backbone conformation in unfolded proteins? *Adv. Prot. Chem.* **62**, 163-240.

52. Bieri, O. & Kiefhaber, T. (1999). Elementary steps in protein folding. *Biol. Chem.* **380**, 923-929.

53. Bieri, O., Wirz, J., Hellrung, B., Schutkowski, M., Drewello, M. & Kiefhaber, T. (1999). The speed limit for protein folding measured by triplet-triplet energy transfer. *Proc. Natl. Acad. Sci. USA* **96**, 9597-9601.

54. Krieger, F., Fierz, B., Bieri, O., Drewello, M. & Kiefhaber, T. (2003). Dynamics of unfolded polypeptide chains as model for the earliest steps in protein folding. *J. Mol. Biol.* **332**, 265-274.

55. Schwarz, G. & Seelig, J. (1968). Kinetic properties and the electric field effect of the helix-coil transition of poly(gamma-benzyl L-glutamate) determined from dielectric relaxation measurements. *Biopolymers* **6**, 1263-1277.

56. Gruenewald, B., Nicola, C. U., Lustig, A., Schwarz, G. & Klump, H. (1979). Kinetics of the helix-coil transition of a polypeptide with non-ionic side groups, derived from ultrasonic relaxation measurements. *Biophys. Chem.* **9**, 137-147.

57. Thompson, P. A., Eaton, W. A. & Hofrichter, J. (1997). Laser temperature jump study of the helix<=>coil kinetics of an alanine peptide interpreted with a 'kinetic zipper' model. *Biochemistry* **36**, 9200-9210.

58. Munoz, V., Thompson, P. A., Hofrichter, J. & Eaton, W. A. (1997). Folding dynamics and mechanism of beta-hairpin formation. *Nature* **390**, 196-199.

59. Xu, Y., Oyola, R. & Gai, F. (2003). Infrared study of the stability and folding kinetics of a 15-residue β-hairpin. *J. Am. Chem. Soc.* **125**, 15388-15394.

60. Chen, R. P., Huang, J. J., Chen, H. L., Jan, H., Velusamy, M., Lee, C. T., Fann, W., Larsen, R. W. & Chan, S. I. (2004). Measuring the refolding of β-sheets with different turn sequences on a nanosecond time scale. *Proc. Natl. Acad. Sci. USA* **101**, 7305-7310.

61.   Flory, P. J. (1988). *Statistical mechanics of chain molecules*, Hanser, Munich, Germany.

62.   Cantor, C. R. & Schimmel, P. R. (1980). *Biophysical Chemistry, Part 3: The behavior of biological macromolecules*, W. H. Freeman, New York, USA.

63.   Fierz, B. & Kiefhaber, T. (2005). Dynamics of Unfolded Polypeptide Chains. In *Protein Folding Handbook* (Buchner, J. & Kiefhaber, T., eds.), Vol. 2, pp. 809-855. 5 vols. Wiley-VCH, Weinheim, Germany.

64.   Flory, P. J. & Schimmel, P. R. (1967). Dipole moments in relation to configuration of polypeptide chains. *J. Am. Chem. Soc.* **89**, 6807-6813.

65.   Stryer, L. & Haugland, R. P. (1967). Energy transfer: a spectroscopic ruler. *Proc. Natl. Acad. Sci. USA* **58**, 719-726.

66.   Schimmel, P. R. & Flory, P. J. (1967). Conformational energy and configurational statistics of poly-L-proline. *Proc. Natl. Acad. Sci. USA* **58**, 52-59.

67.   Brant, D. A. & Flory, P. J. (1965). The Configuration of Random Polypeptide Chains. I. Experimental Results. *J. Am. Chem. Soc.* **87**, 2788-2791.

68.   Doty, P., Bradbury, J. H. & Holtzer, A. M. (1956). Polypeptides. IV. Molecular Weight, Configuration and Association of Poly-γ-benzyl-glutamate in Various Solvents. *J. Am. Chem. Soc.* **78**, 947-954.

69.   Brant, D. A. & Flory, P. J. (1965). The Configuration of Random Polypeptide Chains. II. Theory. *J. Am. Chem. Soc.* **87**, 2791-2800.

70.   Edwards, S. F. (1965). The statistical mechanics of polymers with excluded volume. *Proc. Phys. Soc. London* **85**, 613-624.

71.   Haas, E., Wilchek, M., Katchalski-Katzir, E. & Steinberg, I. Z. (1975). Distribution of end-to-end distances of oligopeptides in solution as estimated by energy transfer. *Proc. Natl. Acad. Sci. USA* **72**, 1807-1811.

72.   Engh, R. A. & Huber, R. (1991). Accurate Bond and Angle Parameters for X-ray Protein Structure Refinement. *Acta Cryst.* **A47**, 392-400.

73.   Hopfinger, A. J. (1973). *Conformational Properties of Macromolecules*, Academic Press, New York, USA.

74.   Richards, F. M. (1977). Areas, Volumes, Packing, and Protein Structure. *Ann. Rev. Biophys. Bioeng.* **6**, 151-176.

75.   Pauling, L. & Corey, R. B. (1950). Two Hydrogen-Bonded Spiral Configuration of the Polypeptide Chain. *J. Am. Chem. Soc.* **72**, 5349.

76.   Pauling, L. & Corey, R. B. (1951). Configuration of polypeptide chains. *Nature* **168**, 550-551.

77.   Pauling, L., Corey, R. B. & Branson, H. R. (1951). The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. USA* **37**, 205-211.

78.   Pauling, L. & Corey, R. B. (1951). Atomic coordinates and structure factors for two helical configurations of polypeptide chains. *Proc. Natl. Acad. Sci. USA* **37**, 235-240.

79.   Pauling, L. & Corey, R. B. (1951). The structure of synthetic polypeptides. *Proc. Natl. Acad. Sci. USA* **37**, 241-250.

80.   Pauling, L. & Corey, R. B. (1951). The pleated sheet, a new layer configuration of polypeptide chains. *Proc. Natl. Acad. Sci. USA* **37**, 251-256.

81.   Pauling, L. & Corey, R. B. (1951). The structure of fibrous proteins of the collagen-gelatin group. *Proc. Natl. Acad. Sci. USA* **37**, 272-281.

82.   Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). Stereochemistry of Polypeptide Chain Configurations. *J. Mol. Biol.* **7**, 95-99.

83.   Ramachandran, G. N. & Sasisekharan, V. (1968). Conformation of Polypeptides and Proteins. *Adv. Prot. Chem.* **23**, 283-438.

84.   Leach, S. J., Némethy, G. & Scheraga, H. A. (1966). Computation of the Sterically Allowed Conformations of Peptides. *Biopolymers* **4**, 369-407.

85.   Pappu, R. V., Srinivasan, R. & Rose, G. D. (2000). The Flory isolated-pair hypothesis is not valid for polypeptide chains: implications for protein folding. *Proc. Natl. Acad. Sci. USA* **97**, 12565-12570.

86.   Fleming, P. J. & Rose, G. D. (2005). Conformational Properties of Unfolded Proteins. In *Protein Folding Handbook* (Buchner, J. & Kiefhaber, T., eds.), Vol. 2, pp. 710-736. 5 vols. Wiley-VCH, Weinheim, Germany.

87.   Baldwin, R. L. & Zimm, B. H. (2000). Are denatured proteins ever random coils? *Proc. Natl. Acad. Sci. USA* **97**, 12391-12392.

88.   Pappu, R. V. & Rose, G. D. (2002). A simple model for polyproline II structure in unfolded states of alanine-based peptides. *Protein Sci.* **11**, 2437-2455.

89.   Fitzkee, N. C. & Rose, G. D. (2004). Reassessing random-coil statistics in unfolded proteins. *Proc. Natl. Acad. Sci. USA* **101**, 12497-12502.

90.   Fitzkee, N. C. & Rose, G. D. (2005). Sterics and Solvation Winnow Accessible Conformational Space for Unfolded Proteins. *J. Mol. Biol.* **353**, 873-887.

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

92.   Levinthal, C. (1969). How to fold graciously. In *Mössbauer spectroscopy in biological systems* (DeBrunner, J. T. P. & Munck, E., eds.), pp. 22-24. University of Illinois Press, Urbana, IL, USA.

93.   Zwanzig, R., Szabo, A. & Bagchi, B. (1992). Levinthal's paradox. *Proc. Natl. Acad. Sci. USA* **89**, 20-22.

94.   Daggett, V. & Fersht, A. R. (2003). Is there a unifying mechanism for protein folding? *Trends Biochem. Sci.* **28**, 18-25.

95. Dolgikh, D. A., Gilmanshin, R. I., Brazhnikov, E. V., Bychkova, V. E., Semisotnov, G. V., Venyaminov, S. & Ptitsyn, O. B. (1981). Alpha-Lactalbumin: compact state with fluctuating tertiary structure? *FEBS Lett.* **136**, 311-315.

96. Ptitsyn, O. (1995). Molten globule and protein folding. *Adv. Prot. Chem.* **47**, 83-229.

97. Karplus, M. & Weaver, D. L. (1976). Protein-folding dynamics. *Nature* **260**, 404-406.

98. Karplus, M. & Weaver, D. L. (1994). Protein folding dynamics: the diffusion-collision model and experimental data. *Prot. Sci.* **3**, 650-658.

99. Bachmann, A. & Kiefhaber, T. (2001). Apparent two-state tendamistat folding is a sequential process along a defined route. *J. Mol. Biol.* **306**, 375-386.

100. Sánchez, I. E. & Kiefhaber, T. (2003). Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J. Mol. Biol.* **325**, 367-376.

101. Wagner, C. & Kiefhaber, T. (1999). Intermediates can accelerate protein folding. *Proc. Natl. Acad. Sci. USA* **96**, 6716-6721.

102. Wetlaufer, D. B. (1973). Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc. Natl. Acad. Sci. USA* **70**, 697-701.

103. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. (1995). Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **21**, 167-195.

104. Eyring, H. (1935). The activated complex in chemical reactions. *J. Chem. Phys.* **3**, 107-115.

105. Kramers, H. A. (1940). Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **4**, 284-304.

106. Hänggi, P., Talkner, P. & Borkovec, M. (1990). Reaction-rate theory: fifty years after Kramers. *Rev. Mod. Phys.* **62**, 251-341.

107. Leffler, J. E. (1953). Parameters for the description of transition states. *Science* **117**, 340-341.

108. Sánchez, I. E. & Kiefhaber, T. (2003). Non-linear rate-equilibrium free energy relationships and Hammond behavior in protein folding. *Biophys. Chem.* **100**, 397-407.

109. Kiefhaber, T., Sánchez, I. E. & Bachmann, A. (2005). Characterization of Protein Folding Barriers with Rate Equilibrium Free Energy Relationships. In *Protein Folding Handbook* (Buchner, J. & Kiefhaber, T., eds.), Vol. 1, pp. 411-444. 5 vols. Wiley-VCH, Weinheim, Germany.

110. Pappenberger, G., Saudan, C., Becker, M., Merbach, A. E. & Kiefhaber, T. (2000). Denaturant-induced movement of the transition state of protein folding revealed by high-pressure stopped-flow measurements. *Proc. Natl. Acad. Sci. USA* **97**, 17-22.

111. Pohl, F. M. (1976). Temperature-dependence of the kinetics of folding of chymotrypsinogen A. *FEBS Lett.* **65**, 293-296.

112. Schätzle, M. (2005). Thesis. Properties of the Free Energy Barriers for Folding of the α-Amylase Inhibitor Tendamistat, Biozentrum, University of Basel, Switzerland.

113. Schindler, T. & Schmid, F. X. (1996). Thermodynamic properties of an extremely rapid protein folding reaction. *Biochemistry* **35**, 16833-16842.

114. Fersht, A. R., Matouschek, A. & Serrano, L. (1992). The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* **224**, 771-782.

115. Fersht, A. R. (1999). *Structure and mechanism in protein science.*, W. H. Freeman, New York.

116. Spiro, K. (1900). Über die Beeinflußung der Eiweißcoagulation durch stickstoffhaltige Substanzen. *Z. Physiol. Chem.* **30**, 182-199.

117. Ramsden, W. (1902). Some new properties of urea. *J. Physiol.* **28**, 23-27.

118. Greenstein, J. P. (1938). Sulfhydryl groups in proteins. *J. Biol. Chem.* **125**, 501-513.

119. Frensdorff, K. H., Watson, M. T. & Kauzmann, W. (1953). The kinetics of protein denaturation. V. The viscosity of urea solutions of serum albumin. *J. Am. Chem. Soc.* **75**, 5167-5172.

120. Tanford, C. (1964). Isothermal unfolding of globular proteins in aqueous urea solutions. *J. Am. Chem. Soc.* **86**, 2050-2059.

121. Gibson, Q. H. & Milnes, L. (1964). Apparatus for Rapid and Sensitive Spectrophotometry. *Biochem. J.* **91**, 161-171.

122. Regenfuss, P., Clegg, R. M., Fulwyler, M. J., Barrantes, F. J. & Jovin, T. M. (1985). Mixing liquids in microseconds. *Rev. Sci. Instrum.* **56**, 283-290.

123. Shastry, M. C. R., Luck, S. D. & Roder, H. (1998). A continuous-flow capillary mixer to monitor reactions on the microsecond time scale. *Biophys. J.* **74**, 2714-2721.

124. Pace, C. N., Grimsley, G. R. & Scholtz, J. M. (2005). Denaturation of Proteins by Urea and Guanidine Hydrochloride. In *Protein Folding Handbook* (Buchner, J. & Kiefhaber, T., eds.), Vol. 1, pp. 45-69. 5 vols. Wiley-VCH, Weinheim, Germany.

125. Pace, C. N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Meth. Enzymol.* **131**, 266-280.

126. Nozaki, Y. & Tanford, C. (1963). The Solubility of Amino Acids and Related Compounds in Aqueous Urea Solutions. *J. Biol. Chem.* **238**, 4074-4081.

127. Robinson, D. R. & Jencks, W. P. (1965). The effect of compounds of the urea-guanidinium class on the activity coefficient of acetyltetraglycine ethyl ester and related compounds. *J. Am. Chem. Soc.* **87**, 2462-2470.

128. Wetlaufer, D. B., Malik, S. K., Stoller, L. & Coffin, R. L. (1964). Non-polar group participation in the denaturation of proteins by urea and guanidinium salts. Model compound studies. *J. Am. Chem. Soc.* **86**, 508-514.

129. Nozaki, Y. & Tanford, C. (1970). The solubility of amino acids, diglycine, and triglycine in aqueous guanidine hydrochloride solutions. *J. Biol. Chem.* **245**, 1648-1652.

130. Nozaki, Y. & Tanford, C. (1971). The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. Establishment of a hydrophobicity scale. *J. Biol. Chem.* **246**, 2211-2217.

131. Aune, K. C. & Tanford, C. (1969). Thermodynamics of the denaturation of lysozyme by guanidine hydrochloride. II. Dependence on denaturant concentration at 25 degrees. *Biochemistry* **11**, 4586-4590.

132. Schellman, J. A. (2002). Fifty years of solvent denaturation. *Biophys. Chem.* **96**, 91-101.

133. Timasheff, S. N. (2002). Thermodynamic binding and site occupancy in the light of the Schellman exchange concept. *Biophys. Chem.* **101-102**, 99-111.

134. Schellman, J. A. (1975). Macromolecular binding. *Biopolymers* **14**, 999-1018.

135. Schellman, J. A. (1978). Solvent denaturation. *Biopolymers* **17**, 1305-1322.

136. Schellman, J. A. (1987). Selective binding and solvent denaturation. *Biopolymers* **26**, 549-559.

137. Greene, R. F. J. & Pace, C. N. (1974). Urea and guanidine-hydrochloride denaturation of ribonuclease, lysozyme, alpha-chymotrypsin and beta-lactoglobulin. *J. Biol. Chem.* **249**, 5388-5393.

138. Pace, C. N. & Tanford, C. (1968). Thermodynamics of the unfolding of β-lactoglobulin A in aqueous urea solutions between 5 and 55°. *Biochemistry* **7**, 198-208.

139. Alexander, S. S. & Pace, C. N. (1971). A comparison of the denaturantion of bovine-lactoglobulins A and B and goat-lactoglobulin. *Biochemistry* **10**, 2738-2743.

140. 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**, 8063-8068.

141. 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.

142. Haas, E., Katchalski-Katzir, E. & Steinberg, I. Z. (1978). Brownian Motion of the Ends of Oligopeptide Chains in Solution as Estimated by Energy Transfer Between the Chain Ends. *Biopolymers* **17**, 11-31.

143. Hagen, S. J., Hofrichter, J., Szabo, A. & Eaton, W. A. (1996). Diffusion-limited contact formation in unfolded cytochrome c: estimating the maximum rate of protein folding. *Proc. Natl. Acad. Sci. USA* **93**, 11615-11617.

144. Chang, I. J., Lee, J. C., Winkler, J. R. & Gray, H. B. (2003). The protein-folding speed limit: intrachain diffusion times set by electron-transfer rates in denatured Ru(NH3)5(His-33)-Zn-cytochrome c. *Proc. Natl. Acad. Sci. USA* **100**, 3838-3840.

145. Krieger, F. (2004). Thesis. Dynamics in Unfolded Polypeptide Chains as Model for Elementary Steps in Protein Folding, Biozentrum, University of Basel, Switzerland.

146. Krieger, F., Fierz, B., Axthelm, F., Joder, K., Meyer, D. & Kiefhaber, T. (2004). Intrachain diffusion in a protein loop fragment from carp parvalbumin. *Chem. Phys.* **307**, 209-215.

147. Lapidus, L. J., Eaton, W. A. & Hofrichter, J. (2000). Measuring the rate of intramolecular contact formation in polypeptides. *Proc. Natl. Acad. Sci. USA* **97**, 7220-7225.

148. Lapidus, L. J., Eaton, W. A. & Hofrichter, J. (2001). Dynamics of intramolecular contact formation in polypeptides: distance dependence of quenching rates in a room-temperature glass. *Phys. Rev. Lett.* **87**, 258101.

149. Lapidus, L. J., Eaton, W. A. & Hofrichter, J. (2002). Measuring dynamic flexibility of the coil state of a helix-forming peptide. *J. Mol. Biol.* **319**, 19-25.

150. Buscaglia, M., Schuler, B., Lapidus, L. J., Eaton, W. A. & Hofrichter, J. (2003). Kinetics of intramolecular contact formation in a denatured protein. *J. Mol. Biol.* **332**, 9-12.

151. Buscaglia, M., Kubelka, J., Eaton, W. A. & Hofrichter, J. (2005). Determination of ultrafast protein folding rates from loop formation dynamics. *J. Mol. Biol.* **347**, 657-664.

152. Hudgins, R. R., Huang, F., Gramlich, G. & Nau, W. M. (2002). A fluorescence-based method for direct measurement of submicrosecond intramolecular contact formation in biopolymers: an exploratory study with polypeptides. *J. Am. Chem. Soc.* **124**, 556-564.

153. Huang, F. & Nau, W. M. (2003). A conformational flexibility scale for amino acids in peptides. *Angew. Chem. Int. Ed. Engl.* **42**, 2269-2272.

154. Huang, F., Hudgins, R. R. & Nau, W. M. (2004). Primary and secondary structure dependence of peptide flexibility assessed by fluorescence-based measurement of end-to-end collision rates. *J. Am. Chem. Soc.* **126**, 16665-16675.

155. Neuweiler, H., Schulz, A., Bohmer, M., Enderlein, J. & Sauer, M. (2003). Measurement of submicrosecond intramolecular contact formation in peptides at the single-molecule level. *J. Am. Chem. Soc.* **125**, 5324-5330.

156. Jacobsen, H. & Stockmayer, W. H. (1950). Intramolecular reaction in polycondensations. I. The theory of linear systems. *J. Phys. Chem.* **18**, 1600-1606.

157. Szabo, A., Schulten, K. & Schulten, Z. (1980). First passage time approach to diffusion controlled reactions. *J. Chem. Phys.* **72**, 4350-4357.

158. Turro, N. J. (1991). *Modern Molecular Photochemistry*, University Science Books, Sausalito, CA, USA.

159. Dexter, D. L. (1953). A Theory of Sensitized Luminescence in Solids. *J. Chem. Phys.* **21**, 836-850.

160. Murov, S. L., Carmichael, I. & Hug, G. L. (1993). *Handbook of Photochemistry*, Marcel Dekker, New York, USA.

161. Pownall, H. J. & Huber, J. R. (1971). Absorption and Emission Spectra of Aromatic Ketones and Their Medium Dependence. Excited States of Xanthone. *J. Am. Chem. Soc.* **93**, 6429-6436.

162. Lewis, G. N. & Kasha, M. (1944). Phosphorescence and the Triplet State. *J. Am. Chem. Soc.* **66**, 2100-2116.

163. Einstein, A. (1906). Eine neue Bestimmung der Moleküldimensionen. *Ann. Phys.* **19**, 289-306.

164. von Smoluchowski, M. (1906). Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen. *Ann. Phys.* **21**, 756-780.

165. von Smoluchowski, M. (1916). Drei Vorträge über Diffusion, Brownsche Molekularbewegung und Koagulation von Kolloidteilchen. *Phys. Z.* **17**, 557-571.

166. Satzger, H., Schmidt, B., Root, C., Zinth, W., Fierz, B., Krieger, F., Kiefhaber, T. & Gilch, P. (2004). Ultrafast Quenching of the Xanthone Triplet by Energy Transfer: New Insight into the Intersystem Crossing Kinetics. *J. Phys. Chem. A* **108**, 10072-10079.

167. Förster, T. (1948). Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann. Phys.* **2**, 55-75.

168. Efimov, V. P., Nepluev, I. V., Sobolev, B. N., Zurabishvili, T. G., Schulthess, T., Lustig, A., Engel, J., Haener, M., Aebi, U., Venyaminov, S. Y., Potekhin, S. A. & Mesyanzhinov, V. V. (1994). Fibritin encoded by bacteriophage T4 gene wac has a parallel triple-stranded alpha-helical coiled-coil structure. *J. Mol. Biol.* **242**, 470-486.

169. Letarov, A. V., Londer, Y. Y., Boudko, S. P. & Mesyanzhinov, V. V. (1999). The carboxy-terminal domain initiates trimerization of bacteriophage T4 fibritin. *Biochemistry (Mosc.)* **64**, 817-823.

170. Tao, Y., Strelkov, S. V., Mesyanzhinov, V. V. & Rossmann, M. G. (1997). Structure of bacteriophage T4 fibritin: a segmented coiled coil and the role of the C-terminal domain. *Structure* **5**, 789-798.

171. Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51-55, 29-32.

172. Povray. (2004). Persistence of Vision Raytracer. Persistence of Vision Pty. Ltd., Williamstown, Victoria, Australia.

173. Güthe, S., Kapinos, L., Möglich, A., Meier, S., Grzesiek, S. & Kiefhaber, T. (2004). Very Fast Folding and Association of a Trimerization Domain from Bacteriophage T4 Fibritin. *J. Mol. Biol.* **337**, 905-915.

174. Frank, S., Kammerer, R. A., Mechling, D., Schulthess, T., Landwehr, R., Bann, J., Guo, Y., Lustig, A., Bächinger, H. P. & Engel, J. (2001). Stabilization of short collagen-like triple helices by protein engineering. *J. Mol. Biol.* **308**, 1081-1089.

175. Boudko, S., Frank, S., Kammerer, R. A., Stetefeld, J., Schulthess, T., Landwehr, R., Lustig, A., Bächinger, H. P. & Engel, J. (2002). Nucleation and propagation of the collagen triple helix in single-chain and trimerized peptides: transition from third to first order kinetics. *J. Mol. Biol.* **317**, 459-470.

176. Frank, S., Boudko, S., Mizuno, K., Schulthess, T., Engel, J. & Bächinger, H. P. (2003). Collagen triple helix formation can be nucleated at either end. *J. Biol. Chem.* **278**, 7747-7750.

177. Stetefeld, J., Frank, S., Jenny, M., Schulthess, T., Kammerer, R. A., Boudko, S., Landwehr, R., Okuyama, K. & Engel, J. (2003). Collagen stabilization at atomic level: crystal structure of designed (GlyProPro)$_{10}$foldon. *Structure (Camb.)* **11**, 339-346.

178. Yang, X., Lee, J., Mahony, E. M., Kwong, P. D., Wyatt, R. & Sodroski, J. (2002). Highly stable trimers formed by human immunodeficiency virus type 1 envelope glycoproteins fused with the trimeric motif of T4 bacteriophage fibritin. *J. Virol.* **76**, 4634-4642.

179. Bächinger, H. P. & Engel, J. (2005). The Thermodynamics and Kinetics of Collagen Folding. In *Protein Folding Handbook* (Buchner, J. & Kiefhaber, T., eds.), Vol. 2, pp. 1059-1110. 5 vols. Wiley-VCH, Weinheim, Germany.

180. Lilie, H. & Seckler, R. (2005). Folding and Association of Multi-domain and Oligomeric Proteins. In *Protein Folding Handbook* (Buchner, J. & Kiefhaber, T., eds.), Vol. 3, pp. 32-72. 5 vols. Wiley-VCH, Weinheim, Germany.

181. Möglich, A., Krieger, F. & Kiefhaber, T. (2005). Molecular Basis for the Effect of Urea and Guanidinium Chloride on the Dynamics of Unfolded Polypeptide Chains. *J. Mol. Biol.* **345**, 153-162.

182. Wilmot, C. M. & Thornton, J. M. (1988). Analysis and prediction of the different types of beta-turn in proteins. *J. Mol. Biol.* **203**, 221-232.

183. Richardson, J. S. (1981). The anatomy and taxonomy of protein structure. *Adv. Prot. Chem.* **34**, 167-339.

184. Krieger, F., Möglich, A. & Kiefhaber, T. (2005). Effect of proline and glycine residues on dynamics and barriers of loop formation in polypeptide chains. *J. Am. Chem. Soc.* **127**, 3346-3352.

185. Güthe, S. (2005). Thesis. Mechanisms of Evolutionary Optimized Complex Folding Reactions, Biozentrum, University of Basel, Switzerland.

186. Meier, S. (2004). Thesis. Novel weak alignment techniques for nuclear magnetic resonance spectroscopy and applications to biomolecular structure determination, Biozentrum, University of Basel, Switzerland.

187. Meier, S., Güthe, S., Kiefhaber, T. & Grzesiek, S. (2004). Foldon, the natural trimerization domain of T4 fibritin, dissociates into a monomeric A-state form containing a stable beta-hairpin: atomic details of trimer dissociation and local beta-hairpin stability from residual dipolar couplings. *J. Mol. Biol.* **344**, 1051-1069.

188. Karplus, P. A. (1996). Experimentally observed conformation-dependent geometry and hidden strain in proteins. *Protein. Sci.* **5**, 1406-1420.

189. Zaman, M. H., Shen, M. Y., Berry, R. S., Freed, K. F. & Sosnick, T. R. (2003). Investigations into sequence and conformational dependence of backbone entropy, inter-basin dynamics and the Flory isolated-pair hypothesis for peptides. *J. Mol. Biol.* **331**, 693-711.

190. Jha, A. K., Colubri, A., Zaman, M. H., Koide, S., Sosnick, T. R. & Freed, K. F. (2005). Helix, sheet, and polyproline II frequencies and strong nearest neighbor effects in a restricted coil library. *Biochemistry* **44**, 9691-9702.

191. Jha, A. K., Colubri, A., Freed, K. F. & Sosnick, T. R. (2005). Statistical coil model of the unfolded state: resolving the reconciliation problem. *Proc. Natl. Acad. Sci. USA* **102**, 13099-13104.

192. Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535-542.

193. Bernado, P., Blanchard, L., Timmins, P., Marion, D., Ruigrok, R. W. & Blackledge, M. (2005). A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc. Natl. Acad. Sci. USA* **102**, 17002-17007.

194. 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.

195. Serron, S. A., Aldridge Iii, W. S., Fleming, C. N., Danell, R. M., Baik, M. H., Sykora, M., Dattelbaum, D. M. & Meyer, T. J. (2004). Evidence for through-space electron transfer in the distance dependence of normal and inverted electron transfer in oligoproline arrays. *J. Am. Chem. Soc.* **126**, 14506-14514.

196. Malak, R. A., Gao, Z., Wishart, J. F. & Isied, S. S. (2004). Long-range electron transfer across peptide bridges: the transition from electron superexchange to hopping. *J. Am. Chem. Soc.* **126**, 13888-13889.

197. Giese, B., Napp, M., Jacques, O., Boudebous, H., Taylor, A. M. & Wirz, J. (2005). Multistep electron transfer in oligopeptides: direct observation of radical cation intermediates. *Angew. Chem. Int. Ed. Engl.* **44**, 4073-4075.

198. Cantor, C. R. & Schimmel, P. R. (1980). *Biophysical Chemistry, Part 1: The conformation of biological macromolecules*, W.H. Freeman, New York, USA.

199. Lin, L. N. & Brandts, J. F. (1980). Kinetic Mechanism for Conformational Transitions between Poly-L-prolines I and II: A Study Utilizing the Cis-Trans Specificity of a Proline-Specific Protease. *Biochemistry* **19**, 3055-3059.

200. Schuler, B., Lipman, E. A. & Eaton, W. A. (2002). Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* **419**, 743-747.

201. Schuler, B., Lipman, E. A., Steinbach, P. J., Kumke, M. & Eaton, W. A. (2005). Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence. *Proc. Natl. Acad. Sci. USA* **102**, 2754-2759.

202. van der Meer, B. W., Coker, G. & Chen, S. S. Y. (1994). *Resonance Energy Transfer. Theory and Data*, Wiley-VCH, Weinheim, Germany.

203. Dale, R. E., Eisinger, J. & E., B. W. (1979). The orientational freedom of molecular probes. The orientation factor in intramolecular energy transfer. *Biophys. J.* **26**, 161-193.

204. Landolt-Börnstein. (1987). *Structure Data of Free Polyatomic Molecules*, II/15, Springer, Heidelberg, Germany.

205. Landolt-Börnstein. (1992). *Structure Data of Free Polyatomic Molecules*, II/21, Springer, Heidelberg, Germany.

206. Kernighan, B. W. & Ritchie, D. (1988). *C Programming Language*, Prentice Hall, Upper Saddle River, NJ, USA.

207. Bloodshed. (2000). Bloodshed Dev-C++ 4.0. Bloodshed Software, http://www.bloodshed.net.

208. Lewis, T. G. & Payne, W. H. (1973). Generalized Feedback Shift Register Pseudorandom Number Algorithm. *J. ACM* **20**, 456-468.

209. MATLAB. (2004). MATLAB 7.0 (R14). The Mathworks, Natick, MA, USA.

8. Published Work and Manuscripts Ready for Submission

JMB

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

ELSEVIER

Molecular Basis for the Effect of Urea and Guanidinium Chloride on the Dynamics of Unfolded Polypeptide Chains

Andreas Möglich, Florian Krieger and Thomas Kiefhaber*

Biozentrum der Universität Basel, Division of Biophysical Chemistry, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

Chemical denaturants are frequently used to unfold proteins and to characterize mechanisms and transition states of protein folding reactions. The molecular basis of the effect of urea and guanidinium chloride (GdmCl) on polypeptide chains is still not well understood. Models for denaturant–protein interaction include both direct binding and indirect changes in solvent properties. Here we report studies on the effect of urea and GdmCl on the rate constants (k_c) of end-to-end diffusion in unstructured poly(glycine-serine) chains of different length. Urea and GdmCl both lead to a linear decrease of $\ln k_c$ with denaturant concentration, as observed for the rate constants for protein folding. This suggests that the effect of denaturants on chain dynamics significantly contributes to the denaturant-dependence of folding rate constants for small proteins. We show that this linear dependency is the result of two additive non-linear effects, namely increased solvent viscosity and denaturant binding. The contribution from denaturant binding can be quantitatively described by Schellman's weak binding model with binding constants (K) of $0.62(\pm 0.01)\,\mathrm{M}^{-1}$ for GdmCl and $0.26(\pm 0.01)\,\mathrm{M}^{-1}$ for urea. In our model peptides the number of binding sites and the effect of a bound denaturant molecule on chain dynamics is identical for urea and GdmCl. The results further identify the polypeptide backbone as the major denaturant binding site and give an upper limit of a few nanoseconds for residence times of denaturant molecules on the polypeptide chain.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: protein folding; polypeptide chain dynamics; solvent denaturation; denaturant binding; Schellman model

*Corresponding author

Introduction

Chemical denaturants like guanidinium chloride (GdmCl) and urea have long been known to unfold proteins[1,2] by stabilizing the unfolded state compared to the native state.[3–8] Unfolding by GdmCl and urea has since then become the standard method to determine protein stability and to investigate the kinetics and mechanisms of protein folding reactions. Empirically, a linear correlation between the free energy for unfolding (ΔG^0) and denaturant concentration ([D]) was found in the region of the unfolding transition.[9,10] The

correlation coefficient of this linear free energy relationship is commonly termed m_{eq}-value ($m_{eq} = \partial \Delta G^0/\partial[D]$). It reflects the denaturing strength of the denaturant. Linear relationships were also found between denaturant concentration and the logarithm of the rate constants for protein folding (k_f) and unfolding (k_u) indicating that the activation free energies for folding ($\Delta G_f^{0\ddagger}$) and unfolding ($\Delta G_u^{0\ddagger}$) are also linearly dependent on denaturant concentration.[11] Accordingly, kinetic m-values were defined as:

$$m_f = \frac{\partial \Delta G_f^{0\ddagger}}{\partial[D]} \text{ and } m_u = \frac{\partial \Delta G_u^{0\ddagger}}{\partial[D]}$$

Despite the wealth of experimental data on the effect of chemical denaturants on protein folding and stability, the physical basis of denaturant action and the origin of the linear free energy relationships are still under discussion. Myers *et al.*[12] found a

correlation between the denaturant m_{eq}-values and the change in solvent accessible surface area (ΔASA) upon unfolding. Mainly two models have been proposed to account for the destabilizing effect of denaturants. The transfer model correlates the effect of denaturants on protein stability to free energies of transfer of the polypeptide chain from water to denaturant solutions.[5,13–17] These data showed that GdmCl and urea increase the solubility of most parts of a protein compared to water explaining the stabilization of the more solvent exposed unfolded state relative to the native state. An alternative model ascribes chemical denaturation to weak binding of denaturant molecules to the polypeptide chain[8,9,18–23] and explains denaturation by the larger number of binding sites in the unfolded state compared to the native state. Both models are able to explain chemical denaturation curves of proteins in the region of the unfolding transition. Santoro & Bolen[24] compared the linear extrapolation method with the binding model and the transfer model by employing different denaturants and extrapolating the stabilities measured in the transition region to zero denaturant. Only the linear extrapolation method yielded identical ΔG^0-values for all applied denaturants. This suggests the applicability of the empirical linear extrapolation method although to date no physical explanation has been given to explain the linearity over the complete range of denaturant concentrations.

The observed effects of denaturants on protein stability and folding are the result of differences between effects on the native state and on the unfolded state for m_{eq}-values and between the transition state and the respective ground states for m_f and m_u. This causes a major difficulty in the interpretation of the experimental data, since the effect of denaturants on a single state cannot be obtained directly. We recently reported that GdmCl and urea affect the rate constants for end-to-end diffusion (k_c) in unstructured polypeptide chains in the same way as they influence the microscopic rate constants for folding and unfolding, i.e. a linear correlation between $\ln k_c$ and $[D]$ was observed.[25] Here we report studies on the effect of urea and GdmCl on intra-chain diffusion in poly(glycine-serine) chains of different length, which serve as models for unstructured polypeptide chains. This allows us to specifically characterize the effect of denaturants on the ensemble of unfolded polypeptide chains and to asses the contributions of chain dynamics to m_f-values in protein folding.

Results

Effect of denaturants on end-to-end diffusion in unstructured polypeptide chains

We have previously used triplet-triplet energy transfer between xanthone (Xan) and naphthylalanine (NAla) to determine absolute rate constants for intra-chain diffusion (k_c) and the effect of

denaturants on chain dynamics in various unstructured polypeptide chains. To understand the molecular basis of denaturant action we investigated the effect of urea and GdmCl on unstructured polypeptide chains of varying length in more detail. Figure 1(A) shows representative kinetic traces for triplet-triplet energy transfer in a Xan-(Gly-Ser)$_8$-NAla-Ser-Arg-Gly peptide at various GdmCl concentrations. Single exponential kinetics for contact formation were observed in all measurements. The presence of GdmCl leads to a significant decrease in k_c from $4.0(\pm 0.1) \times 10^7$ s^{-1} in water to $1.7(\pm 0.1) \times 10^7$ s^{-1} in 8 M GdmCl. A similar effect was observed for urea (data not shown). A plot of $\ln k_c$
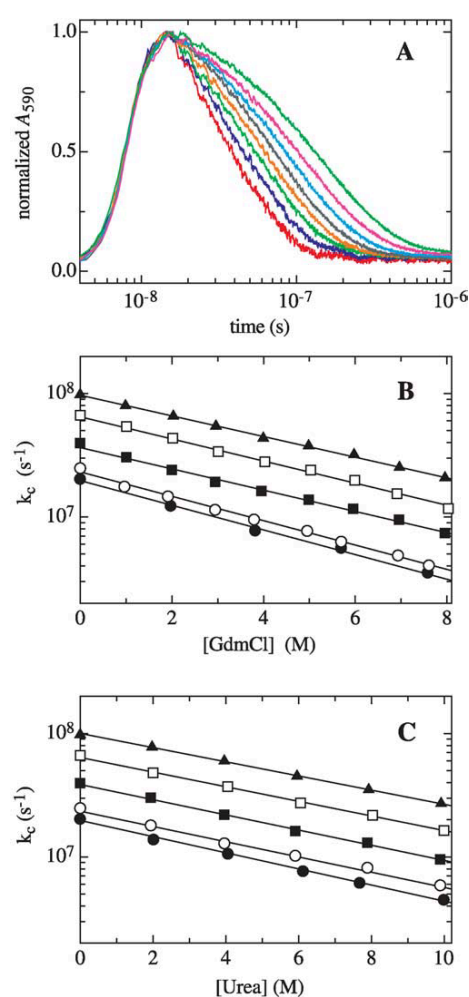
Figure 1. (A) Kinetics of end-to-end diffusion measured by triplet-triplet energy transfer in a Xan-(Gly-Ser)$_8$-NAla-Ser-Arg-Gly peptide as a function of GdmCl concentration (from left to right 1, 2, 3, 4, 5, 6, 7, and 8 M GdmCl). The reaction is followed by the decrease in xanthone triplet absorbance at 590 nm. (B) GdmCl dependence and (C) urea dependence of the rate constants for end-to-end contact formation (k_c) for various (Gly-Ser)$_n$ peptides with $n = 3$ (▲), 5 (□), 8 (■), 12 (○) and 14 (●) measured by triplet-triplet energy transfer as shown in (A). Solid lines represent fits of the data to equation (13).

versus [D] reveals approximately linear relationships both for GdmCl and urea (Figure 1(B) and (C)). In analogy to protein folding, we determined m_c-values ($m_c = -RT \partial \ln k_c/\partial[D]$) for the effect of urea and GdmCl on chain dynamics (Figure 1(B) and (C)). For the (Gly-Ser)$_8$ peptide a fit to equation (13) gave m_c values of 0.49 (kJ/mol)/M for GdmCl and 0.35 (kJ/mol)/M for urea (Table 1). To investigate the effect of chain length on the m_c-values we measured the urea and GdmCl dependence of end-to-end diffusion in various Xan-(Gly-Ser)$_n$-NAla-Ser-Arg-Gly peptides with $n = 3, 5, 8, 12, 14$. For all peptides an approximately linear relationship between $\ln k_c$ and [D] is observed (Figure 1). Both for urea and for GdmCl the m_c-values increase slightly with peptide length (Table 1).

Viscosity-corrected chain dynamics

Addition of denaturants significantly increases solvent viscosity,[26] which is known to affect end-to-end diffusion in unstructured peptides.[27] To determine the contributions from increased solvent viscosity to the m_c-values we measured chain dynamics in various glycerol/water mixtures (Figure 2). All peptides show a linear correlation between $\ln k_c$ and $\ln \eta$. The effect of η on k_c was analyzed using:

$$k_c = k_c^0 \left(\frac{\eta}{\eta_0} \right)^{\beta} \tag{1}$$

where η_0 is the reference solvent viscosity at zero denaturant concentration, k_c^0 is the rate constant of end-to-end diffusion at η_0, k_c is the rate constant at a given viscosity η and the exponent β reflects the sensitivity of the reaction to solvent viscosity. A β-value of -1 indicates that k_c is inversely proportional to solvent viscosity and a β-value of 0 indicates that k_c is independent of solvent viscosity. Fitting the data to equation (1) gives β-values ranging from -0.90 for the shortest peptide to -0.99 for the longest peptide (Table 1). To test for contributions from specific interactions of the co-solutes with the peptides we also used ethylene glycol, glucose and sucrose as co-solutes to increase solvent viscosity. All co-solutes gave the same β-values within error, indicating that the
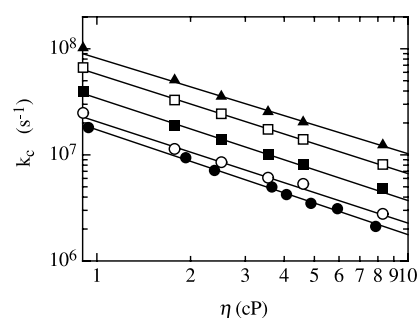

Figure 2. Effect of solvent viscosity (η) on k_c for the series of (Gly-Ser)$_n$ peptides. Peptides and symbols are the same as shown in Figure 1(B) and (C). Measurements were performed in glycerol/water mixtures. Continuous lines show fits of the experimental data to equation (2) with the slopes (β) given in Table 1.

observed effect is due to an increase in solvent viscosity and not caused by specific co-solute-peptide interactions (data not shown and F.K. & T.K., unpublished results). To correct for the contributions of changes in η we used the β-values in combination with the viscosities of the respective denaturant solutions to calculate viscosity-corrected rate constants for end-to-end diffusion (k_c') according to equation (2):

$$k_c' = k_c \left(\frac{\eta}{\eta_0} \right)^{-\beta} \tag{2}$$

This correction reduces the denaturant dependence of chain dynamics. However, the corrected k_c-values still depend on denaturant concentration (Figure 3(A) and (B)) but asymptotically approach a limiting value at high denaturant concentrations. This indicates that the effect of denaturants on end-to-end diffusion is partly due to a non-specific increase in solvent viscosity and partly due to a denaturant-specific effect, which becomes saturated at high denaturant concentrations. Correction for viscosity effects completely eliminates the denaturant-dependence of bimolecular triplet-triplet energy transfer (TTET) from xanthone to naphthylalanine,

Table 1. Denaturant dependence of end-to-end diffusion in Xan-(Gly-Ser)$_n$-Ser-Arg-Gly peptides

Peptide	β[a]	Urea			GdmCl		
		m_c[b] ((J/mol)/M)	K_{ex}[c]	γ[d]	m_c ((J/mol)/M)	K_{ex}	γ
(Gly-Ser)$_3$	-0.90 ± 0.01	325.2 ± 2.5	13.1 ± 0.8	0.70 ± 0.01	474.1 ± 1.3	34.9 ± 1.1	0.65 ± 0.01
(Gly-Ser)$_5$	-0.94 ± 0.01	335.3 ± 1.6	13.6 ± 0.6	0.68 ± 0.01	507.2 ± 1.9	23.8 ± 0.6	0.72 ± 0.01
(Gly-Ser)$_8$	-0.97 ± 0.01	348.3 ± 2.1	13.3 ± 0.6	0.69 ± 0.01	488.1 ± 1.8	33.8 ± 0.9	0.66 ± 0.01
(Gly-Ser)$_{12}$	-0.96 ± 0.01	349.8 ± 3.3	15.6 ± 1.0	0.68 ± 0.01	559.0 ± 3.0	26.0 ± 0.9	0.77 ± 0.01
(Gly-Ser)$_{14}$	-0.99 ± 0.04	370.2 ± 3.4	14.5 ± 1.2	0.71 ± 0.02	565.7 ± 4.9	27.4 ± 1.5	0.73 ± 0.01
Global	–	–	13.8 ± 0.3	0.69 ± 0.01	–	32.6 ± 0.4	0.68 ± 0.01

[a] Viscosity dependence of k_c; see equation (3).
[b] $m_c = RT$ d $\ln K_c$/d[D]; see equation (13).
[c] Equilibrium constant for exchange of a water molecule with a denaturant molecule; see equation (3).
[d] Effect of a bound denaturant molecule on chain dynamics; see equation (4).
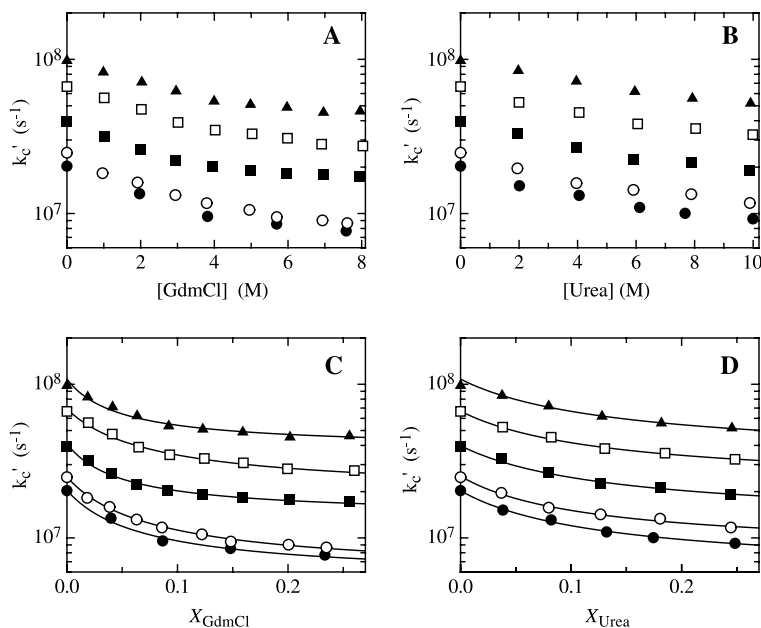
Figure 3. Effect of GdmCl (A) and urea (B) on the viscosity-corrected rate constants of end-to-end contact formation (k'_c) displayed on a molar scale. (C) and (D) The same data, respectively, on a mole fraction scale. In addition, the results of global fits of the data to the Schellman "site exchange" model are shown in (C) and (D). The results of the global fits and of all individual fits are given in Table 1. The (Gly-Ser)$_n$ peptides and symbols are the same as for Figure 1.

indicating that denaturants do not influence the electron transfer reaction.

Schellman's weak binding model *versus* transfer model

We analyzed the denaturant-specific effect of urea and GdmCl on chain dynamics with the "site exchange" formalism, which was developed by Schellman to describe weak interactions of denaturants and other osmolytes with proteins.[8,22] Assuming independent binding sites on a polypeptide chain, the average fractional binding site occupancy, $\bar{\nu}$, can be expressed as a function of the mole fraction of denaturant, X_D:

$$\bar{\nu} = \frac{(K_{ex} - 1)X_D}{(K_{ex} - 1)X_D + 1} \quad (3)$$

K_{ex} is the equilibrium constant for the exchange of a water molecule with a denaturant molecule at a single binding site. Further details and the derivation of equation (3) are discussed by Schellman.[8,22] To analyze the effect of denaturants on k'_c in terms of the weak binding model we converted the denaturant concentrations to the mole fraction scale (Figure 3(C) and (D)). We made the assumption that the effect of denaturant binding on chain dynamics is proportional to the fractional binding site occupancy ($\bar{\nu}$), which means that all denaturant binding sites on the polypeptide chains are identical and independent of each other. These assumptions lead to equation (4):

$$k'_c = k^0_c - \gamma k^0_c \bar{\nu} \quad (4)$$

where k^0_c is the rate constant for intra-chain diffusion in the absence of denaturant. The proportionality constant, γ, reflects the effect exerted by bound denaturant molecules on chain dynamics.

Fitting the data for urea and GdmCl to equations (3) and (4) shows that the denaturant dependence of k'_c is well-described by Schellman's weak binding model for all peptides (Figure 3(C) and (D); Table 1). Interestingly, both K_{ex} and γ do not vary systematically with peptide length (Table 1), which prompted us to fit the data for all peptides globally for each denaturant. This gave K_{ex}-values of 32.6 ± 0.4 and 13.8 ± 0.3 for binding of GdmCl and urea, respectively (Table 1), which corresponds to respective binding constants (K) of $0.62(\pm 0.01)$ M^{-1} and $0.26(\pm 0.01)$ M^{-1} on the molar scale. Remarkably, binding of GdmCl and urea slows down contact formation by the same amount as indicated by the virtually identical γ-values of 0.68 ± 0.01 for GdmCl and 0.69 ± 0.01 for urea (Table 1).

To also test whether the transfer model is able to describe the effect of denaturants on k'_c we compared the data to predictions from ΔG^0_{tr}-values for different amino acids reported by Nozaki & Tanford.[13,16,17] The correlation between $\ln k'_c$ and ΔG^0_{tr} shows significant deviations from linearity especially for GdmCl (Figure 4). In addition, the slope of the correlation changes significantly with peptide length, both for urea and for GdmCl. This is not compatible with a general effect of transfer free energies on the chain properties. Similar results were obtained when only the ΔG^0_{tr}-values for the polypeptide backbone were considered.

Discussion

Direct and indirect effects of denaturants on polypeptide dynamics

Our results show that the effect of denaturants on intra-chain diffusion in unstructured polypeptides
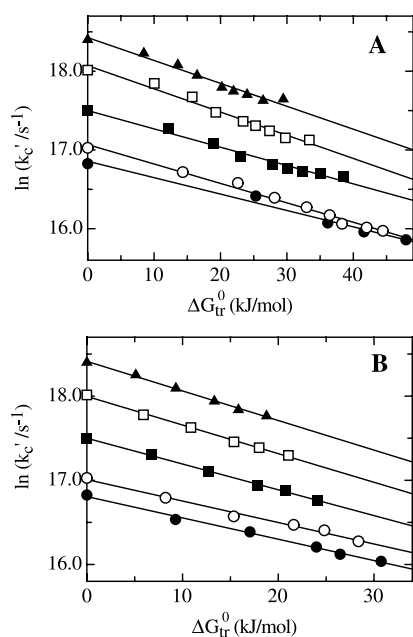
Figure 4. Correlation between k'_c and ΔG^0_{tr} for GdmCl (A) and urea (B) calculated from the ΔG^0_{tr}-values given by Nozaki & Tanford (see Materials and Methods). The (Gly-Ser)$_n$ peptides and symbols are the same as for Figure 1. The linear fits of the data give slopes ($-d \ln k'_c/d \ln K_{tr} = d \ln k'_c/d \ln \Delta G^0_{tr}/RT$) of $-7.2(\pm 0.5)\times 10^{-2}$, $-7.2(\pm 0.4)\times 10^{-2}$, $-5.7(\pm 0.3)\times 10^{-2}$, $-6.1(\pm 0.2)\times 10^{-2}$ and $-5.1(\pm 0.4)\times 10^{-2}$ for $n = 3,5,8,12$ and 14, respectively, for GdmCl and of $-8.6(\pm 0.3)\times 10^{-2}$, $-8.3(\pm 0.2)\times 10^{-2}$, $-7.5(\pm 0.2)\times 10^{-2}$, $-6.3(\pm 0.3)\times 10^{-2}$, $-6.2(\pm 0.2)\times 10^{-2}$, respectively, for urea.

can be dissected into two components. Denaturants indirectly slow down chain dynamics by increasing solvent viscosity,[27] and denaturants directly interact with polypeptide chains. The individual contributions from both components have non-linear effects on the denaturant dependence of $\ln k_c$. k_c is proportional to η^β (see equation (1)) with β-values between -0.90 and -0.99 for the different peptides (Figure 2). Solvent viscosity increases approximately hyperbolically with denaturant concentration,[26] which leads to a corresponding decrease in $\ln k_c$ with denaturant concentration. The effect of denaturant binding, in contrast, strongly changes $\ln k_c$ at low denaturant concentrations and approaches a limiting value at high concentrations (Figure 3). The sum of both effects results in an approximately linear denaturant dependence of $\ln k_c$, as seen in Figure 1. This model predicts that the linearity will break down at very high denaturant concentrations, which are, however, experimentally not accessible. A closer inspection of the additive effects of viscosity and binding reveals small deviations from linearity in the experimentally accessible concentration range. These small deviations are indeed seen both in the curve

calculated from the fitted parameters and in the experimental data (Figure 5) supporting the model that solvent viscosity and weak denaturant binding are the molecular origin of the observed effect of denaturants on end-to-end diffusion.

The effect of the direct denaturant–polypeptide interaction on chain dynamics can be quantitatively described by Schellman's "site exchange" model (Figure 3). The site exchange model gives K_{ex}-values of 32.6 ± 0.4 and 13.8 ± 0.3 for binding of GdmCl and urea, respectively (Table 1), which corresponds to respective binding constants (K) of $0.62(\pm 0.01)$ M^{-1} and $0.26(\pm 0.01)$ M^{-1} on the molar scale. Urea and GdmCl have virtually identical γ-values (0.69 ± 0.01 *versus* 0.68 ± 0.01; Table 1), which reflect the effect of bound denaturant molecules on chain dynamics (see equation (4)). This indicates that the polypeptide chains have the same number of binding sites for urea and GdmCl and that bound urea and GdmCl molecules decrease the rate constant for intra-chain contact formation by the same amount. The different efficiency of GdmCl and urea in slowing down chain dynamics is thus solely based on their different binding affinities to polypeptide chains. Full binding site occupancy decreases the rate of contact formation by about 3.3-fold for both denaturants. The binding constants and γ-values for urea and GdmCl are independent of the chain length, indicating that the slightly higher m_c-values for longer chains are due to their higher sensitivity to changes in solvent viscosity (Figure 2 and Table 1). The relative hydrophobicity should be significantly higher for the shorter peptides compared to the longer peptides, due to the contributions of the aromatic TTET labels. The insensitivity of the binding constant for chain length indicates only little contribution from side-chains and argues for the polypeptide backbone as the major denaturant binding site. This is consistent with the observation that the m_{eq}-values for protein unfolding are
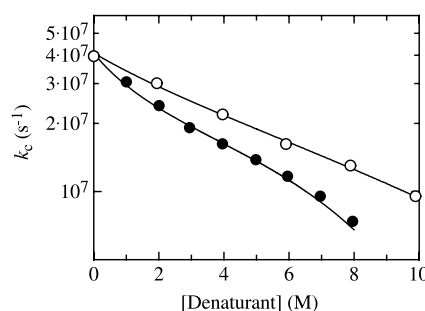
Figure 5. Comparison of the experimental data for the GdmCl ($\bullet$) and urea ($\circ$) dependence of $\ln k_c$ in the (Gly-Ser)$_8$ peptide with the simulated curve assuming additive effects of solvent viscosity and denaturant binding on chain dynamics. The curve was simulated using equations (3) and (4) with the experimentally determined values for the viscosity dependence ($\beta = -0.97$) and denaturant-binding (K_{ex}(urea) $= 13.3$; K_{ex}(GdmCl) $= 33.8$).

approximately independent of the amino acid sequence[12] and with recent results from molecular dynamics simulations, which suggested the polypeptide backbone as the major site for interaction of denaturants with unfolded proteins.[28,29]

Comparison to binding constants measured in other systems

The binding constants of $0.26(\pm 0.01)\,\mathrm{M}^{-1}$ for urea and $0.62(\pm 0.01)\,\mathrm{M}^{-1}$ for GdmCl measured for their effect on end-to-end diffusion represent specific binding constants of the denaturants to the unfolded polypeptide chain. These values are in good agreement with values determined for denaturant binding to model compounds. Robinson & Jencks[15] reported solubility measurements and transfer studies of acetyl tetraglycyl ester from water to urea and GdmCl solutions and obtained binding constants of $0.29\,\mathrm{M}^{-1}$ for urea and $0.90\,\mathrm{M}^{-1}$ for GdmCl at 25 °C by using a classical binding model. Tanford & Nozaki determined transfer free energies of various amino acids and peptide bond mimics from water to solutions of various denaturant concentrations.[13,16,17] From these data Pace[6] calculated binding constants in the range of $0.08\,\mathrm{M}^{-1}$ to $0.29\,\mathrm{M}^{-1}$ for urea and $0.51\,\mathrm{M}^{-1}$ to $1.03\,\mathrm{M}^{-1}$ for GdmCl. The values found for binding to peptide groups are $0.09\,\mathrm{M}^{-1}$ for urea and $0.60\,\mathrm{M}^{-1}$ for GdmCl. The binding constants determined from the effect of GdmCl on chain dynamics also agree well with calorimetrically determined values for GdmCl binding to unfolded proteins, which were $0.60\,\mathrm{M}^{-1}$ at 25 °C for RNase A and lysozyme [30] and $1.16\,\mathrm{M}^{-1}$ for apocytochrome *c*.[31] However, calorimetrically determined binding constants for interactions of urea with unfolded proteins are significantly lower ($K = 0.06\,\mathrm{M}^{-1}$ at 25 °C) than those derived from the effect of urea on chain dynamics.[30]

Comparison of our data to binding constants derived from unfolding transitions in proteins is not straightforward. Analysis of unfolding transitions usually assumes identical binding constants for denaturants on the unfolded and native state but different number of denaturant binding sites as the origin of the transitions. Still, the binding constants determined for the effect of urea and GdmCl on the helix-coil transition in short model peptides agree well with the values derived from chain dynamics. Applying the weak binding model to analyze the effect of urea on the helix-coil transition in short peptides gave a binding constant of $0.14\,\mathrm{M}^{-1}$.[32] Using the weak binding model to analyze the data reported by Scholtz and co-workers for the effect of GdmCl in the same peptides[33] we obtained values of $0.67\,\mathrm{M}^{-1}$ and $0.58\,\mathrm{M}^{-1}$ in the presence 10 mM and 100 mM NaCl, respectively.

A major question in the analysis of the effect of denaturants on proteins is whether there is direct evidence for binding or whether high denaturant concentrations non-specifically lead to the presence of denaturants near the polypeptide chain. Recent magnetization transfer NMR studies on native and unfolded proteins have shown direct evidence for specific protein-denaturant interactions. For urea-binding to native BPTI a binding constant in the range of $0.1\,\mathrm{M}^{-1}$ to $0.36\,\mathrm{M}^{-1}$ was determined at 36 °C.[34] Experiments on urea binding to unfolded intestinal fatty acid-binding protein revealed both strong and weak binding sites.[35] The experimental data for the weak binding sites are in agreement with binding constants between $0.05\,\mathrm{M}^{-1}$ and $0.2\,\mathrm{M}^{-1}$, which are similar to the value of $0.26\,\mathrm{M}^{-1}$ found in our experiments. However, in the analysis of the magnetization transfer data the binding constant was fixed to the indicated values and not included in the data fitting. For GdmCl binding to unfolded and reduced α-lactalbumin an average binding constant of $0.16(\pm 0.07)\,\mathrm{M}^{-1}$ was reported,[36] which is significantly lower than the value of $0.62\,\mathrm{M}^{-1}$ found in our studies.

Residence time of denaturant molecules on the polypeptide chain

The observation of single exponential kinetics for contact formation in all peptides and at all denaturant concentrations suggests residence times of a few nanoseconds or less for denaturants bound to the unfolded chain. Exchange of denaturant molecules slower than 5 ns would lead to heterogeneous populations of unfolded molecules and result in complex kinetics. Fast exchange of bound denaturants is in agreement with results from NMR studies on native and unfolded proteins. Binding of urea to native BPTI and Pec-60 showed average residence times of a few nanoseconds at 4 °C.[34] For urea and GdmCl-binding to weak binding sites in unfolded fatty acid binding protein average residence times of around 1 ns or shorter were reported.[35] Short residence times of urea on unfolded CI2 were also observed in molecular dynamics simulations.[29]

Implications for protein folding *m*-values

Our results demonstrate that the rate constants for intra-chain diffusion in unfolded polypeptide chains significantly depend on denaturant concentration and exhibit the same linear relationship between $\ln k_c$ and $[D]$ as rate constants for protein folding reactions. This will contribute to the denaturant dependence of the rate constants for protein folding (m_f-values). In general, the rate constant (k) of a reaction contains contributions from the prefactor (k_0) and the activation free energy ($\Delta G^{0\ddagger}$) according to:

$$k = k_0 \mathrm{e}^{-\Delta G^{0\ddagger}/RT} \tag{5}$$

It is very difficult to estimate absolute values for pre-exponential factors for the folding k_0^f and unfolding reaction k_0^u. They are most likely different for each protein, since the pre-exponential factors depend on the shape of the potential and on the

dynamics in the individual kinetic states (unfolded protein, native protein and folding intermediates). According to Kramers theory for barrier crossing reactions in the high friction limit, which should correspond to most reactions in solution and also to protein folding reactions, the rate of a reaction depends on the friction (γ) encountered by motions on a free energy landscape, by the shape of the barriers and by the frequency of motion (ω) on the free energy landscape according to equation (6):[37,38]

$$k = \frac{\omega_0 \omega_B}{2\pi\gamma} e^{(-E_b/kT)} \qquad (6)$$

ω_0 and ω_B denote the frequencies of motion in the starting well and on top of the barrier, respectively. This shows that the pre-exponential factor for protein folding depends on the shape of the potential and on the dynamics in the individual wells (states). ω_0 will be different for the refolding reaction starting from unfolded protein (ω_0^U) and for the unfolding reaction starting from native protein (ω_0^N) and hence the pre-factors will also be different for the forward and backward reaction (Figure 6). Eaton an co-workers measured an effective viscosity of 4.1 cP for internal motions in native myoglobin, in agreement with increased friction and/or decreased frequency of chain motions in native proteins compared to unfolded proteins.[39]

In the light of these considerations our results suggest that the denaturant dependence of chain dynamics has major effects on the pre-factors for protein folding reactions. For a two-state folding reaction:

$$U \underset{k_u}{\overset{k_f}{\rightleftharpoons}} N \qquad (7)$$

equation (8) is valid under all conditions:[40]

$$K_{eq} = \frac{k_f}{k_u} \qquad (8)$$

Thus, the equilibrium constant for folding (K_{eq})

contains contributions from the pre-exponential factors for folding (k_0^f) and unfolding (k_0^u) and from the free energies of activation, $\Delta G_f^{0\ddagger}$ and $\Delta G_u^{0\ddagger}$, according to:[41]

$$K_{eq} = \frac{k_0^f e^{-\Delta G_f^{0\ddagger}/RT}}{k_0^u e^{-\Delta G_u^{0\ddagger}/RT}} \qquad (9)$$

The contribution of differences in the pre-factors for folding and unfolding to ΔG^0 ($= -RT \ln K_{eq}$) are consequently:

$$\Delta G^0 = \Delta G_f^{0\ddagger} - \Delta G_u^{0\ddagger} + \Delta G_{pref}^0$$
$$\Delta G_{pref}^0 = RT \ln \frac{k_0^u}{k_0^f} \qquad (10)$$

Differences in k_0 for folding and unfolding can thus significantly contribute to apparent protein stabilities. A fivefold difference in k_0 between unfolded and native state, e.g. caused by differences in ω_0 (see Figure 6) would contribute 4 kJ/mol to the experimentally determined ΔG^0-values at 25 °C.[41] Consequently, changes in k_0^f caused by solvent-induced changes in the dynamics of the unfolded state, will influence both k_f and K_{eq}, even if the height of the barrier is not changed. Experimentally it can thus not be distinguished, whether a solvent-induced change in a rate constant arises from changes in the free energy barriers or from changes in the pre-factors, since equation (8) always holds. The effect of denaturants on the pre-factor for folding (m_f^0) will contribute to the observed kinetic and equilibrium m-values, m_f(app) and m_{eq}(app), according to equation (11):
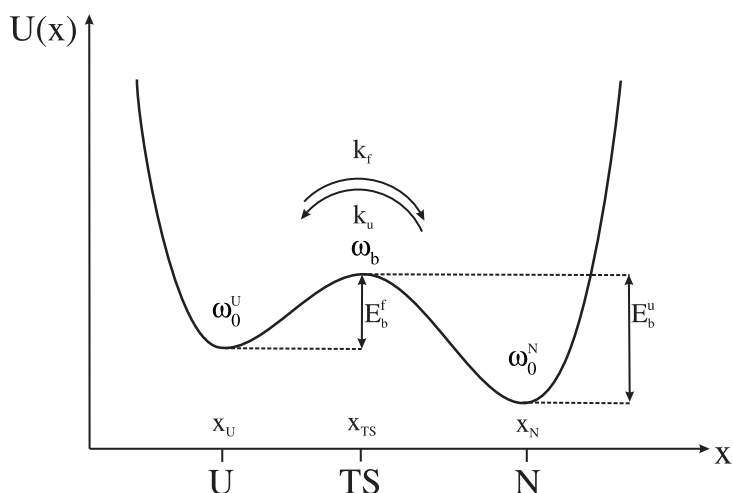

Figure 6. Schematic representation of a barriers crossing reaction on a double-well potential surface with two states N and U separated by an energy barriers. Escape from state U to N occurs *via* the rates k_f and k_u. E_b^f and E_b^u are the activation energies for the forward and back reaction. ω_0^U, ω_0^N and ω_B denote the frequencies of motion in the unfolded state, the native state and at the top of the barrier, respectively. The drawing was adopted from Hänggi *et al.*[38] and adapted for a schematic protein folding reaction.

$$\frac{\partial \Delta G_f^{0\ddagger}(\mathrm{app})}{\partial [D]} = \frac{\partial \Delta G_f^{0\ddagger}}{\partial [D]} - RT \frac{\partial \ln(k_0^f)}{\partial [D]} = m_f - m_f^0 = m_f(\mathrm{app})$$

$$\frac{\partial \Delta G_u^{0\ddagger}(\mathrm{app})}{\partial [D]} = \frac{\partial \Delta G_f^{0\ddagger}}{\partial [D]} - RT \frac{\partial \ln(k_0^u)}{\partial [D]} = m_u - m_u^0 = m_u(\mathrm{app})$$

$$\frac{\partial \Delta G^{0}(\mathrm{app})}{\partial [D]} = \frac{\partial \Delta G_u^{0\ddagger}(\mathrm{app})}{\partial [D]} - \frac{\partial \Delta G_f^{0\ddagger}(\mathrm{app})}{\partial [D]} = m_u - m_f + (m_f^0 - m_u^0) = m_{eq}(\mathrm{app}) \qquad (11)$$

This allows to estimate the contributions of the denaturant dependence of intra-chain diffusion (m_c) to experimentally determined m-values for small proteins. The m_c-values for peptide dynamics vary only little with chain length and show values around 0.35 (kJ/mol)/M for urea and 0.50 (kJ/mol)/M for GdmCl. Typical m_f-values for folding of the smallest fast folding proteins with chain length between 40 and 50 amino acid residues are around 1.0 (kJ/mol)/M for urea indicating that up to 30% of the measured m_f-values may arise from contributions of chain dynamics. Larger two-state folders consisting of 70–100 amino acid residues typically show m_f-values around 3 (kJ/mol)/M for urea and 5 (kJ/mol)/M for GdmCl,[42,43] indicating that the denaturant dependence of chain dynamics constitutes up to 10% of the experimental m_f-values. It is difficult to judge the effect of denaturants on the dynamics of the native state, which will contribute to the experimental m_u-values. However, we have observed that the effect of denaturants on chain dynamics is largely based on increased end-to-end distances (A.M. & T.K., unpublished results). This would argue for only little effects of denaturant binding on the internal dynamics of the native state and the transition state structures, which were shown to be structurally well-defined and robust against changes in denaturant concentration.[43–46] In addition, native proteins have less denaturant binding sites and the dynamics of native proteins, which contribute to k_0 for the unfolding reaction (see equation (6) and Figure 6), are strongly influenced by intra-molecular interactions. This reduces the frequency of chain motions as observed for myoglobin.[39] These motions should thus be less sensitive to the presence of denaturants. Consequently, the m_u-values are likely to have only little contributions from the effects of denaturants on chain dynamics ($m_u^0 \approx 0$) and m_{eq}-values should contain similar contributions from m_f^0 as the m_f-values. (see equation (11)).

Comparison of kinetic and equilibrium m-values is frequently used to characterize protein folding transition states according to the rate-equilibrium free energy relationship:[11,44,45]

$$\alpha_D = \frac{m_f}{m_{eq}} \qquad (12)$$

α_D is interpreted as a measure for the solvent accessibility of the transition state. According to our results, the contributions from chain dynamics to the m_f-values of small proteins lead to apparently higher α_D-values and thus to apparently less solvent exposed transition states. This might explain the commonly observed higher α_D-values compared to α_C-values ($\alpha_C = \Delta C_P^{0\ddagger}/\Delta C_P^0$), which are also believed to monitor the solvent accessibility of transition states.[44,45] Temperature dependence of intra-chain diffusion in unstructured peptides showed that this process is not associated with a measurable change in heat capacity (F.K. & T.K., unpublished results). This suggests that α_C-values give a more reliable picture of the solvent accessibility of the transition state than α_D-values.

The ratio of binding constants determined for the binding of GdmCl and urea to unfolded chains is 2.4 (Table 1), which is virtually identical to the average ratio of 2.3 found for the m_{eq}-values of the two denaturants,[12] whereas the ratio of m_c-values for GdmCl and urea is significantly lower (1.5). This indicates that denaturant binding is the major source for the effect of denaturants on protein stability. We cannot compare these values to m_f-values for urea and GdmCl in protein folding reactions due to the small number of proteins that were characterized both with urea and GdmCl. However, the contribution of chain dynamics on m_f-values should depend on the location of the transition state along the reaction coordinate so a larger scattering in the ratio of urea and GdmCl m_f-value is expected.

Materials and Methods

Xanthonic acid and peptides were synthesized and purified as described.[25] The peptides are based on repetitive units of glycine-serine and have the canonical sequence Xan-(Gly-Ser)$_n$-NAla-Ser-Arg-Gly:
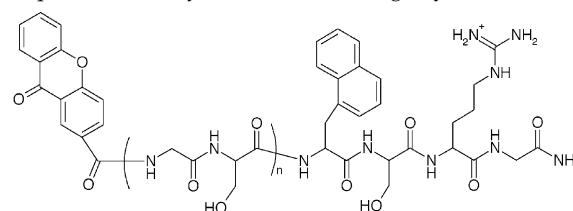
A peptide with n repeats of (Gly-Ser) between the two chromophores xanthone (Xan) and naphthylalanine (NAla) is referred to as (Gly-Ser)$_n$ peptide. All measurements were performed in degassed solutions containing 10 mM potassium phosphate (pH 7) at 22.5 °C. Peptide concentrations were between 30 μM and 100 μM. Denaturant concentrations were determined by measuring the refractive index according to Pace.[6]

Triplet-triplet energy transfer (TTET) measurements

were performed as described by Krieger *et al.*[25] Xanthone triplet states were generated by a 4 ns laser pulse (Nd:YAG, 355 nm) and transient absorbance of the xanthone triplet state was monitored at 590 nm. The decay of the xanthone triplet absorbance is accompanied by an increase in the naphthyl triplet absorbance band at 420 nm indicating exchange of two electrons between xanthone and naphthalene upon contact formation between the chromophores.[25] Generally, four to eight kinetic traces at 590 nm were averaged and analyzed by non-linear least-squares fitting of the experimental data to exponential functions. Single exponential kinetics were observed for contact formation in all peptides. Errors were less than 5%. Data analysis was carried out using the computer programs ProFit (Quantum Soft, Zürich, Switzerland) and Matlab (The MathWorks, Natick, MA, USA).

The m_c-values for end-to-end diffusion were determined by fitting the data to equation (13):

$$\ln k_c = \ln k_c^0 - \frac{m_c[D]}{RT} \tag{13}$$

where k_c^0 is the rate constant for end-to-end diffusion at zero denaturant concentration and k_c is the respective value at the denaturant concentration $[D]$. Viscosities of denaturant solutions were calculated according to Perl *et al.*[26] These values agree well with viscosities measured in a falling ball viscosimeter (Haake, Germany). Viscosities of water/glycerol solutions were determined with a falling ball viscometer. Transfer free energies (ΔG_{tr}^0) from water to solutions containing various concentrations of urea or GdmCl were calculated from the values given by Nozaki & Tanford.[13,16,17] The values for Trp were used to calculate ΔG_{tr}^0 for naphthylalanine and 1.5 times the value for Trp was used to calculate the ΔG_{tr}^0 value for xanthone. The values for Ser were approximated by the values of Thr minus the difference in ΔG_{tr}^0 between Leu and Val.

Acknowledgements

References

1. Spiro, K. (1900). Über die Beeinflussung der Eiweiss-coagulation durch stickstoffhaltige Substanzen. *Z. Physiol. Chem.* **30**, 182–199.
2. Ramsden, W. (1902). Some new properties of urea. *J. Physiol.* **28**, 23–27.
3. Simpson, R. B. & Kauzman, W. (1953). The kinetics of protein denaturation. I. The behavior of the optical rotation of ovalbumin in urea solutions. *J. Am. Chem. Soc.* **75**, 5139–5152.
4. Frensdorff, K. H., Watson, M. T. & Kauzman, W. (1953). The kinetics of protein denaturation. V. The viscosity of urea solutions of serum albumin. *J. Am. Chem. Soc.* **75**, 5167–5172.
5. Tanford, C. (1964). Isothermal unfolding of globular proteins in aqueous urea solutions. *J. Am. Chem. Soc.* **86**, 2050–2059.
6. Pace, C. N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* **131**, 266–280.
7. Makhatadze, G. I. (1999). Thermodynamics of protein interactions with urea and guanidinium hydrochloride. *J. Phys. Chem.* **103**, 4781–4785.
8. Schellman, J. A. (2002). Fifty years of solvent denaturation. *Biophys. Chem.* **96**, 91–101.
9. Aune, K. C. & Tanford, C. (1969). Thermodynamics of the denaturation of lysozyme by guanidine hydrochloride. II Dependence on denaturant concentration at 25 degrees. *Biochemistry,* **11**, 4586–4590.
10. Greene, R. F. J. & Pace, C. N. (1974). Urea and guanidine-hydrochloride denaturation of ribonuclease, lysozyme, alpha-chyomtrypsinn and beta-lactoglobulin. *J. Biol. Chem.* **249**, 5388–5393.
11. Tanford, C. (1970). Protein denaturation. Part C. Theoretical models for the mechanism of denaturation. *Adv. Protein Chem.* **24**, 1–95.
12. 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.
13. Nozaki, Y. & Tanford, C. (1963). The solubility of amino acids and related compounds in aqueous urea solutions. *J. Biol. Chem.* **238**, 4074–4081.
14. Wetlaufer, D. B., Malik, S. K., Stoller, L. & Coffin, R. L. (1964). Non-polar group participation in the denaturation of proteins by urea and guanidinium salts. Model compund studies. *J. Am. Chem. Soc.* **86**, 508–514.
15. Robinson, D. R. & Jencks, W. P. (1965). The effect of compounds of the urea-guanidinium class on the activity coefficient of acetyltetraglycine ethyl ester and related compounds. *J. Am. Chem. Soc.* **87**, 2462–2470.
16. Nozaki, Y. & Tanford, C. (1970). The solubility of amino acids, diglycine, and triglycine in aqueous guanidine hydrochloride solutions. *J. Biol. Chem.* **245**, 1648–1652.
17. Nozaki, Y. & Tanford, C. (1971). The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. Establishment of a hydrophobicity scale. *J. Biol. Chem.* **246**, 2211–2217.
18. Schellman, J. A. (1955). The stability of hydrogen-bonded peptide structures in aqueous solutions. *Compt. rend. Carlsberg Lab. Ser. Chim.* **29**, 230–259.
19. Schellman, J. A. (1958). The factors affecting the stability of hydrogen-bonded polypeptide structures in solution. *J. Phys. Chem.* **62**, 1485–1494.
20. Schellman, J. A. (1975). Macromolecular binding. *Biopolymers,* **14**, 999–1018.
21. Schellman, J. A. (1978). Solvent denaturation. *Biopolymers,* **17**, 1305–1322.
22. Schellman, J. A. (1987). Selective binding and solvent denaturation. *Biopolymers,* **26**, 549–559.
23. Timasheff, S. N. (2002). Thermodynamic binding and site occupancy in the light of the Schellman exchange concept. *Biophys. Chem.* **101–102**, 99–111.
24. 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**, 8063–8068.
25. Krieger, F., Fierz, B., Bieri, O., Drewello, M. &

Kiefhaber, T. (2003). Dynamics of unfolded polypeptide chains as model for the earliest steps in protein folding. *J. Mol. Biol.* **332**, 265–274.

26. Perl, D., Jacob, M., Bánó, M., Stupâk, M., Antalík, M. & Schmid, F. X. (2002). Thermodynamics of a diffusional protein folding reaction. *Biophys. Chem.* **2–3**, 173–190.

27. Bieri, O., Wirz, J., Hellrung, B., Schutkowski, M., Drewello, M. & Kiefhaber, T. (1999). The speed limit for protein folding measured by triplet-triplet energy transfer. *Proc. Natl Acad. Sci. USA*, **96**, 9597–9601.

28. Tobi, D., Elber, R. & Thirumalai, D. (2003). The dominant interaction between peptide and urea is electrostatic in nature: a moelcular dynamics simulation study. *Biopolymers*, **68**, 359–369.

29. Bennion, B. J. & Daggett, V. (2003). The molecular basis for the chemical denaturation of proteins by urea. *Proc. Natl Acad. Sci. USA*, **100**, 5142–5147.

30. Makhatadze, G. I. & Privalov, P. L. (1992). Protein interactions with urea and guanidinium chloride. A calorimetric study. *J. Mol. Biol.* **226**, 491–505.

31. Pfeil, W., Welfle, K. & Bychkova, V. E. (1991). Guanidine hydrochloride titration of the unfolded apocytochrome c studied by calorimetry. *Studia Biophysica*, **140**, 5–12.

32. Scholtz, J. M., Barrick, D., York, E. J., Stewart, J. M. & Baldwin, R. L. (1995). Urea unfolding of peptide helices as a model for interpreting protein unfolding. *Proc. Natl Acad. Sci. USA*, **92**, 185–189.

33. Smith, J. S. & Scholtz, J. M. (1996). Guanidine hydrochloride unfolding of peptide helices: separation of denaturant and salt effects. *Biochemistry*, **35**, 7292–7297.

34. Liepinsh, E. & Otting, G. (1994). Specificity of urea binding to proteins. *J. Am. Chem. Soc.* **116**, 9670–9674.

35. Modig, K., Kurian, E., Prendergast, F. G. & Halle, B. (2003). Water and urea interactions with the native and unfolded forms of a beta-barrel protein. *Protein Sci.* **12**, 2768–2781.

36. Denisov, V. P., Jonsson, B. H. & Halle, B. (1999). Hydration of denatured and molten globule proteins. *Nature Struct. Biol.* **6**, 253–260.

37. Kramers, H. A. (1940). Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, **4**, 284–304.

38. Hänggi, P., Talkner, P. & Borkovec, M. (1990). Reaction-rate theory: fifty years after Kramers. *Rev. Mod. Phys.* **62**, 251–341.

39. Ansari, A., Jones, C. M., Henry, E., Hofrichter, J. & Eaton, W. A. (1992). The role of solvent viscosity in the dynamics of protein conformational changes. *Science*, **256**, 1796–1798.

40. van't Hoff, J. H. (1884). *Etudes de dynamique*. Muller, Amsterdam.

41. Bieri, O. & Kiefhaber, T. (2000). Kinetic models in protein folding. In *Protein Folding: Frontiers in Molecular Biology* (Pain, R., ed.) 2nd edit., pp. 34–64, Oxford University Press, Oxford.

42. Jackson, S. E. (1998). How do small single-domain proteins fold? *Fold. Des.* **3**, R81–R91.

43. Sánchez, I. E. & Kiefhaber, T. (2003). Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J. Mol. Biol.* **325**, 367–376.

44. Sánchez, I. E. & Kiefhaber, T. (2003). Non-linear rate-equilibrium free energy relationships and Hammond behavior in protein folding. *Biophys. Chem.* **100**, 397–407.

45. Sánchez, I. E. & Kiefhaber, T. (2003). Hammond behavior versus ground state effects in protein folding: evidence for narrow free energy barriers and residual structure in unfolded states. *J. Mol. Biol.* **327**, 867–884.

46. Sánchez, I. E. & Kiefhaber, T. (2003). Origin of unusual phi-values in protein folding: evidence against specific nucleation sites. *J. Mol. Biol.* **334**, 1077–1085.

Effect of Guanidinium Chloride on Dimensions and Dynamics

of Unfolded Polypeptide Chains

Andreas Möglich & Thomas Kiefhaber[*]

Biozentrum der Universität Basel

Division of Biophysical Chemistry

Klingelbergstrasse 70

CH-4056 Basel

Switzerland

[*]to whom correspondence should be addressed: Thomas Kiefhaber, Tel. ++41 61 267 2194,

Fax ++41 61 267 2189, e-mail: t.kiefhaber@unibas.ch

Abstract

Denaturants like urea and guanidinium chloride (GdmCl) are frequently used to unfold proteins and to probe mechanisms and transition states of protein folding reactions. Despite their wide use in biophysical and biochemical studies, the molecular basis of the effects that denaturants exert on polypeptide chains is still not well understood. Here, we use time-resolved FRET measurements to determine the effect of GdmCl on dynamics and dimensions of an unfolded model polypeptide chain consisting of 16 glycine-serine repeats. We use a novel approach that is based on global analysis of data obtained from different donor-acceptor pairs. This allows us to measure the end-to-end distance distribution and the intrachain diffusion coefficient as a function of denaturant concentration. The results show that the average end-to-end distance increases from (18.7 ± 1.0) Å to (39.2 ± 1.0) Å between 0 M and 8 M GdmCl. At the same time the chain becomes more flexible as indicated by an increase in diffusion constant from $(3.6 \pm 1.0) \cdot 10^{-7}$ cm^2 s^{-1} at 0 M GdmCl to $(14.8 \pm 2.5) \cdot 10^{-7}$ cm^2 s^{-1} at 8 M GdmCl. These results suggest that intrachain backbone interactions exist even in very flexible unfolded chains, which leads to major chain compaction upon changing the solvent from high denaturant concentrations to water. These results imply that rapid collapse of polypeptide chains during refolding of denaturant-unfolded proteins can at least in part be ascribed to non-specific intrachain backbone interactions and does not require the presence of hydrophobic side chains. Comparison of our FRET results to end-to-end contact formation rate constants determined by triplet-triplet energy transfer revealed that the denaturant effects on the average dimensions, the internal dynamics and loop closure kinetics of unstructured peptides are correlated. Further, we were able to show that end-to-end contact formation in long peptides can be sufficiently described by the first passage time theory.

Introduction

Despite the widespread use of chemical denaturants in protein folding and stability studies the molecular mechanism by which they unfold proteins is still largely unknown. It has long been known that chemical denaturants, such as guanidinium chloride[1] (GdmCl) or urea[2,3], denature proteins by selectively stabilizing the unfolded state relative to the native state[4,5]. Empirically, linear relations between the Gibbs free energy of unfolding (ΔG^0) and the concentration of chemical denaturant ([D]) were observed with the so-called $m_{eq.}$-value defining the slope of this correlation ($m_{eq.} = \partial \Delta G^0 / \partial [D]$)[6,7,8]. These linear free energy relationships are now routinely employed to determine protein stabilities[9]. Similar relations were also reported for the free activation energies ($\Delta G^{0\ddagger}$) of protein folding and unfolding rate constants[5,6], allowing the determination of corresponding kinetic m-values ($m_f = \partial \Delta G_f^{0\ddagger} / \partial [D]$ and $m_u = \partial \Delta G_u^{0\ddagger} / \partial [D]$). Myers *et al.*[10] showed that the magnitude of the m-values correlates with the amount of solvent-accessible surface area exposed upon unfolding and thus with protein size. Consequently, the effect of denaturants on protein stability and folding kinetics has been used to characterize protein folding transition states in terms of their solvent accessibility. A detailed knowledge of the mechanism of the denaturant effect on proteins would provide further insight into the molecular basis of protein stability and would allow a more detailed characterization of transition state properties.

Several models for the mechanism of denaturant action were proposed and found to adequately describe the experimental data[11,12,13,14,15]. Recently, studies on the dynamics of intramolecular loop formation in peptides have shed new light on the mode of action of chemical denaturants[16]. Using the triplet-triplet energy transfer (TTET) technique[17,18,19], it was observed that the rate constant for intrachain loop closure, k_c, is decreased in the presence of denaturants. In analogy to protein folding reactions linear relationships between the

logarithm of k_c and denaturant concentration were found. The effect of guanidinium chloride and urea on the rate constant of end-to-end contact formation could be dissected into two components. Part of the observed slowdown of contact formation in the presence of denaturants is due to increased solvent viscosity. The residual effect could be described by weak binding of denaturant molecules to the peptide. Here, we report time-resolved fluorescence resonance energy transfer (FRET) studies on unstructured model peptide chains to investigate the effect of GdmCl on the chain dimensions and on the diffusion constant between the FRET labels. Comparison to the previously measured end-to-end contact formation rate constants in the same peptide chains allowed us to assess the effect of GdmCl binding to the peptide on a molecular level.
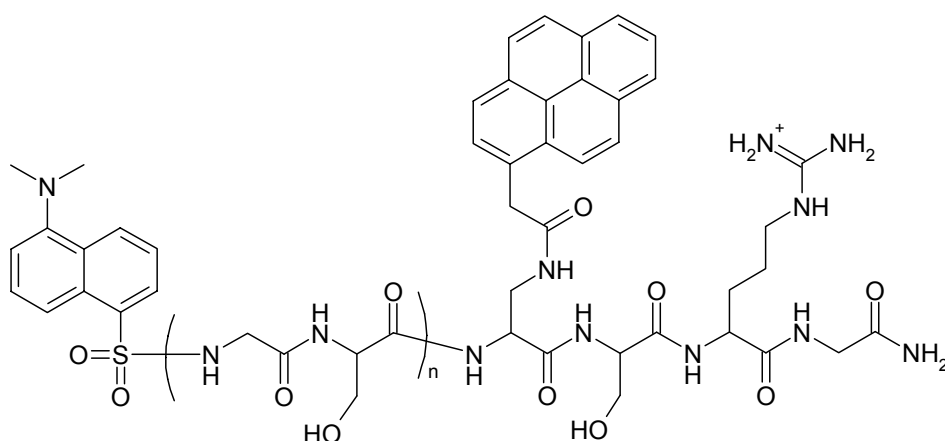
Materials and Methods

Peptide synthesis and purification

All peptides were synthesized using standard fluorenylmethoxycarbonyl (F-moc) chemistry on an Applied Biosystems 433A peptide synthesizer as described[18] and have the following structures.

(A)
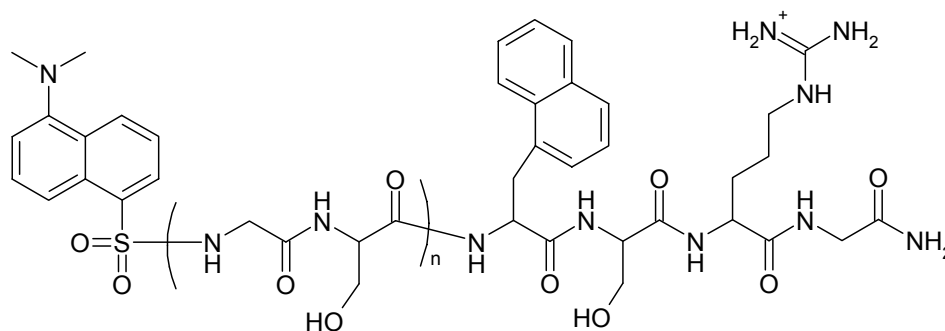
(B)

n (n = 16) denotes the number of repeats of the glycine-serine dipeptide. The peptides were synthesized with an amidated C-terminus using TentaGel S RAM resin (Rapp Polymere, Tübingen, Germany). Resin-bound peptides were N-terminally labelled with dansyl chloride (Fluka, Buchs, Switzerland). The resin was incubated for 30 minutes in a solution of 10 % (v/v) diisopropylethylamine in dimethylformamide containing a tenfold excess of dansyl chloride. The naphthyl moiety in peptide (A) was introduced via the nonnatural amino acid F-moc-1-naphthylalanine (NAla, Bachem, Bubendorf, Switzerland). The pyrene label (peptide (B)) was attached to the sidechain of the nonnatural amino acid diaminopropionic acid (Dpr, NovaBiochem, Laufelingen, Switzerland). During synthesis and dansyl labelling the amine side chain of Dpr was protected by a methyltrityl group which is selectively cleaved under mild acidic conditions (2 % (v/v) trifluoroacetic acid (TFA) in dichloromethane). 1-pyreneacetic acid (Aldrich, Buchs, Switzerland) was coupled to Dpr using standard F-moc chemistry.

To determine the properties of the fluorescence labels, control peptides consisting of eight glycine-serine repeats (n = 8) were synthesized. These substrates bear either only the acceptor or only the donor chromophore, referred to as donor-only and acceptor-only peptides,

respectively. In the case of the donor-only peptides the N-terminal dansyl group was replaced by an acetyl moiety. The acceptor-only peptide was obtained by replacing NAla in peptide (A) with alanine. All peptides were purified by reversed-phase HPLC on a RP-8-HPLC column (Merck, Darmstadt, Germany) using acetonitrile/0.1 % TFA and water/0.1 % TFA as eluents. In the case of the FRET peptides labelled with two fluorophores a second purification on the RP-8-HPLC column was carried out employing a mixture of acetonitrile and 50 mM ammonium acetate buffer pH 6.0. Higher than 95 % purity of the peptides was confirmed by analytical HPLC and mass spectrometry. All chemicals were of peptide synthesis or higher grade and were purchased from Fluka (Buchs, Switzerland) and Acros Organics (Basel, Switzerland).

Fluorescence measurements

Steady-state fluorescence measurements were carried out on an Aminco Bowman series 2 fluorimeter (SLM Aminco, USA). Characteristic FRET R_0 distances were calculated according to eq. (2). Fluorescence quantum yields and spectra were measured for the above donor-only peptides. N-acetyl-tryptophanamide was used as a standard for determining the quantum yield ($\Phi_F = 0.144$)[20]. Absorbance spectra of the dansyl chromophore in acceptor-only peptides were acquired on an Agilent 8453 UV-visible spectrophotometer (Agilent, Basel, Switzerland). The orientational term κ^2 in eq. (2) was taken to equal 2/3.

Time-resolved fluorescence measurements were conducted with a time-correlated single-photon counting fluorimeter (FLS900, Edinburgh Instruments, Edinburgh, UK). Donor fluorophores were excited by irradiation with a H_2-flashlamp at 284 nm with 1.5 ns pulse width operating at a frequency of 40 kHz. Fluorescence decay curves were monitored at the emission maxima of 325 and 377 nm, respectively, for the naphthalene and pyrene chromophores. Measurements were carried out on solutions containing 8 μM peptide, 10 mM

potassium phosphate buffer pH 7.0, 0.5 % DMSO and varying concentrations of guanidinium

chloride. Concentrations of the denaturant were checked by refractive index measurements[11].

Samples were degassed prior to measurement and equilibrated at 22.5 ˚C.

Control measurements

Fluorescence measurements were further conducted on a mixture of 4 µM acceptor-only and

4 µM donor-only peptide. No energy transfer was detected under these conditions, indicating

that the energy transfer observed in the doubly-labelled peptides is intramolecular. The donor-

only peptides showed low steady-state fluorescence anisotropies of 0.005 ± 0.006 and

0.001 ± 0.014 in the case of naphthalene and pyrene, respectively. From these values

rotational correlation times smaller than 0.5 ns can be estimated for the donor fluorophores

which is much faster than the average fluorescence decay observed in the time-resolved FRET

measurements. Therefore, the use of a value of 2/3 for the orientational term κ^2 in eq. (2)

appears justified.

Data analysis

Data were analyzed with the programs Matlab (The MathWorks, Natick, NA, USA) and

ProFit (QuantumSoft, Zürich, Switzerland). Fluorescence decay curves obtained with the

chromophore pairs naphthyl/dansyl and pyrene/dansyl were fitted globally to the numerical

solution of equations (6-8). The parameters a and b, which determine the end-to-end

probability distribution, and the end-to-end diffusion constant D were calculated as global

parameters.

Monte-Carlo simulation of peptide conformations

Random conformations for a $(GS)_{16}$ peptide without steric overlap were generated using an all-atom hard-sphere model as described[21]. The spectroscopic labels used in our earlier TTET experiments[16] were attached to the peptide termini. Xanthonic acid (Xan) was linked to the N-terminus via an amide bond. The nonnatural amino acid naphthylalanine (NAla) was appended to the C-terminus. Bond lengths and angles for Xan and NAla were derived from the corresponding crystal structures[22,23]. End-to-end distances were calculated as the vector connecting the middle of the xanthone aromatic ring system with the middle of the naphthalene aromatic ring system. In total, 62000 valid conformations without steric clashes were simulated in this way.

Results

Using FRET to measure molecular distances

Fluorescence resonance energy transfer (FRET) designates the radiationless transfer of excited-state energy between two fluorophores commonly termed donor and acceptor. Förster developed the theoretical background and the formalism describing this process[24]. Energy transfer occurs through space and is mediated by the resonant oscillation of transition dipoles in the two chromophores. As the resonance condition has to be met, FRET only occurs between pairs of chromophores with spectral overlap between the donor emission (fluorescence) and the acceptor absorbance spectra. The rate of energy transfer, k_T, depends on the properties of the chromophores and on the distance r separating them according to eq. (1).

$$k_T = \frac{1}{\tau_D}\left(\frac{R_0}{r}\right)^6 \tag{1}$$

Here, τ_D denotes the intrinsic fluorescence lifetime of the donor fluorophore and R_0 the characteristic Förster distance which is a function of the properties of the given pair of fluorophores (eq. (2)).

$$R_0^6 = \frac{9000 \ln 10\, \kappa^2 \Phi_D}{128\pi^5 N n^4} \int_0^\infty F_D(\lambda)\varepsilon_A(\lambda)\lambda^4 d\lambda \qquad (2)$$

In eq. (2) N is Avogadro's number, n the refractive index of the medium, κ^2 an orientational term assumed to equal 2/3 for rapidly reorienting fluorophores, Φ_D the fluorescence quantum yield of the donor, $F_D(\lambda)$ the normalized emission of the donor and $\varepsilon_A(\lambda)$ the extinction coefficient of the acceptor at wavelength λ. The dependence of the energy transfer rate constant on the inverse sixth power of the interchromophore separation can be utilized to accurately determine distances on a molecular scale. To this end usually the efficiency of energy transfer, E, is calculated by comparing the fluorescence properties of the donor chromophore in the absence and in the presence of the acceptor (eq. (3)).

$$E = \frac{F_D - F_{DA}}{F_D} = \frac{\tau_D - \tau_{DA}}{\tau_D} = \frac{k_T}{1/\tau_D + k_T} \qquad (3)$$

FRET efficiencies can be determined by steady-state fluorescence measurements where the fluorescence intensities in the absence (F_D) and the presence (F_{DA}) of the acceptor chromophore are measured. Typically, a more accurate determination of E can be achieved by comparing the corresponding fluorescence lifetimes, τ_D and τ_{DA}, obtained from time-resolved FRET (trFRET) measurements. Combining eqs. (1-3) shows that transfer efficiencies can be directly correlated with the distance separating the two fluorophores according to eq. (4).

$$E = \frac{R_0^6}{R_0^6 + r^6} \tag{4}$$

At the characteristic distance R_0 the transfer efficiency amounts to 50 %. The applicability of the above theory has first been shown experimentally by Stryer and Haugland[25]. Employing suitable chromophore pairs, distances on the scale of 10 to 100 Å can be accurately determined.

FRET measurements on unfolded polypeptide chains

The basic FRET theory has to be extended when studying highly dynamic systems which undergo conformational rearrangements on the timescale of the donor lifetime, such as unstructured polypeptide chains. Due to their flexibility, unfolded polypeptide chains represent heterogeneous ensembles of conformations with differing interchromophore distances and thus with differing energy transfer rate constants. In trFRET measurements this leads to a multiexponential fluorescence decay, $F_{DA}(t)$, (eq. (5)).

$$F_{DA}(t) = \int_0^\infty p(r) \exp\left\{-\left(1 + \left(\frac{R_0}{r}\right)^6\right)\frac{t}{\tau_D}\right\} dr \tag{5}$$

where $p(r)$ is the probability of a molecule having an interfluorophore distance r. For analysis of experimental data often a three-dimensional Gaussian probability distribution is assumed for $p(r)$ (eq. (6))[26,27].

$$p(r) = c\pi r^2 \exp\left(-a(r-b)^2\right) \tag{6}$$

The Gaussian distribution is offset from zero by b. The parameter a determines the width of the distribution and $c = f(a,b)$ is a normalization constant.

In addition, when studying flexible molecules by FRET one has to take into account that the interfluorophore distance governing the energy transfer rate is subject to change during the finite time needed for the measurement. Haas and coworkers[27] developed a formalism which treats conformational rearrangements as a diffusional process (eqs. (7) and (8)).

$$\frac{\partial \bar{p}(r,t)}{\partial t} = -\frac{1}{\tau_D}\left\{\left(1+\left(\frac{R_0}{r}\right)^6\right)\bar{p}(r,t)\right\} + \frac{1}{p_0(r)}\frac{\partial}{\partial r}\left(p_0(r)D\frac{\partial \bar{p}(r,t)}{\partial r}\right) \qquad (7)$$

$$\bar{p}(r,t) = \frac{p(r,t)}{p_0(r)} \qquad (8)$$

D signifies the diffusion constant for motions of the two fluorophores relative to each other. Even for simple probability distribution functions, $p(r)$, eq. (7) does not have an analytical solution.

Spectroscopic properties of the fluorophore systems

In the present work we performed FRET experiments on glycine-serine repeat peptides previously studied by TTET measurements . The dansyl group was attached to the N terminus as an acceptor chromophore. In our studies we used two different donor peptides in which either naphthalene (peptide (A)) or pyrene (peptide (B)) were introduced near the C terminus as fluorescence donor groups. Donor and acceptor labels were separated by 16 repeats of the Gly-Ser dipeptide ($(GS)_{16}$ peptide). The spectroscopic properties of the individual chromophores and of the FRET pairs they form were determined in control peptides bearing only one fluorophore (referred to as donor-only resp. acceptor-only peptides). The results obtained at different concentrations of GdmCl are summarized in Table 1. An increase in the denaturant concentration leads to a decrease in both the donor fluorescence lifetime and in the Förster distance. For the naphthalene/dansyl pair in the absence of denaturant a donor fluorescence lifetime of (36.9 ± 0.2) ns and an R_0 distance of (23.3 ± 0.4) Å were measured.

At a concentration of 8 M GdmCl these values decrease to (18.6 ± 0.2) ns and (20.2 ± 0.4) Å, respectively (Table 1). In the case of the pyrene/dansyl system donor lifetimes of (225.5 ± 1.0) and (162.0 ± 1.0) ns, and R_0 distances of (20.5 ± 0.4) and (20.2 ± 0.4) Å were measured in the absence and in the presence of 8 M GdmCl, respectively. At all denaturant concentrations both FRET pairs show similar characteristic distances but differ considerably in their donor fluorescence lifetimes by a factor of 6.1–8.7 depending on the GdmCl concentration (Table 1).

Time-resolved FRET measurements

Time-resolved fluorescence was monitored by time-correlated single-photon counting after excitation with a H_2-flashlamp (284 nm). Representative fluorescence data monitored at 325 nm for the $(GS)_{16}$ peptide labelled with the dansyl and naphthyl moieties (Dan-$(GS)_{16}$-NAla peptide) in the absence of GdmCl are shown in Fig. 1A (blue trace). In comparison to the corresponding donor-only peptide (black trace) the fluorescence decays faster which is indicative of FRET. Additionally, fluorescence originating from the dansyl group could be observed at 550 nm (data not shown). In our previous studies on the dynamics of unstructured peptides we observed end-to-end contact formation rate constants on the nanoseconds timescale. Thus, a significant contribution of end-to-end diffusion occurring on the timescale of the present FRET experiments, which is governed by the donor fluorescence lifetime, is expected. To account for this effect, we evaluated our data according to the applicable model described above (eqs. (6-8)). However, analysis of a single fluorescence decay curve did not allow us to reliably determine both the end-to-end distance probability distribution and the diffusion constant, since these quantities are highly correlated in the data fitting procedure. Haas and coworkers[27] suggested to globally evaluate donor and acceptor fluorescence traces to resolve this problem. Yet, this procedure did not turn out to be successful in the present case.

To resolve the contributions of average end-to-end distance and rapid chain dynamics to the observed fluorescence lifetimes, we performed FRET experiments on peptides with pyrene as a donor group. Using dansyl as the acceptor fluorophore, pyrene displays similar R_0 values but considerably longer fluorescence lifetimes compared to naphthalene. Fig. 1B shows FRET data acquired for the $(GS)_{16}$ peptide labelled with pyrene and dansyl (Dan-$(GS)_{16}$-Pyr peptide). Time-resolved fluorescence was monitored at 377 nm. For comparison the fluorescence decay in the corresponding donor-only peptide is shown as well. Upon introduction of the acceptor group dansyl, the fluorescence lifetime of pyrene is strongly decreased. Compared to the naphthalene-labelled peptide this decrease of fluorescence lifetime is relatively more pronounced (Fig. 1A). In order to quantitate this difference between the two fluorophores, we approximated the time-resolved FRET data, $F_{DA}(t)$, by single exponential functions according to eq. (9)[28].

$$\tau_{app.} \approx \frac{1}{F_{DA}(0)} \cdot \int_0^\infty F_{DA}(t)dt \qquad (9)$$

With the apparent time constant, $\tau_{app.}$, apparent values for the transfer efficiency and the end-to-end distance were calculated according to eqs. (3) and (4) (Table 2). At all denaturant concentrations the apparent FRET efficiencies with pyrene as a donor fluorophore are considerably higher than in the case of naphthalene. The apparent interchromophore distances increase with denaturant concentration but are under all conditions 5–7 Å larger for the peptide with naphthalene as the fluorescence donor. These results clearly show a significant influence of diffusional events on the observable FRET efficiency.

Global analysis of the two fluorophore systems

We exploited the difference in donor lifetimes and the resulting difference in the contributions of diffusional processes to the fluorescence decay curves observed for the two fluorophore systems to determine both the end-to-end distance probability distribution and the corresponding diffusion constant in a global fitting procedure. Time-resolved fluorescence data acquired separately for the two chromophore systems were evaluated globally where common values for the distance distribution and the end-to-end diffusion constant in the two peptides were assumed. This assumption seems justified taking into account the small size of the chromophores relative to the rest of the peptides. Global analysis of the time-resolved fluorescence curves allowed to resolve the parameter correlation between mean distance and diffusion constant. A representative example of a global fit is shown in Fig. 1 (orange lines). Results from the global evaluation of FRET data at different concentrations of GdmCl are summarized in Figs. 2 and 3 and in Table 3.

On average, the peptide shows increasingly larger end-to-end distances with increasing GdmCl concentration (Fig. 2). The effect is strongest at low denaturant concentrations and seems to level off at higher concentrations. To quantify this behaviour, the root mean-square end-to-end distances of the distribution functions were calculated (Fig. 3A and Table 3). In the absence of denaturant a root mean square end-to-end distance $\sqrt{\langle r_0^2 \rangle}$ of (18.7 ± 1.0) Å was obtained. With increasing GdmCl concentration this value increases in a hyperbolic fashion and reaches a value of (39.2 ± 1.0) Å at 8 M GdmCl. Similar to our previous analysis of the denaturant effect on contact formation rate constants[16], the dependency of $\sqrt{\langle r_0^2 \rangle}$ on denaturant concentration can be quantitatively described by a binding isotherm employing the Schellman model[15,29,30,31] for weak interactions (eq. (10)).

$$\sqrt{\left\langle r_0^2 \right\rangle} = \sqrt{\left\langle r_0^2 \right\rangle^0} \cdot \left(1 + \gamma \frac{(K_{ex}-1)X_D}{(K_{ex}-1)X_D + 1} \right) \tag{10}$$

Here, $\sqrt{\left\langle r_0^2 \right\rangle^0}$ denotes the root mean square end-to-end distance in the absence of denaturant,

K_{ex} a binding constant for GdmCl on the mole fraction scale (X_D), and γ is a factor which

quantifies the sensitivity of $\sqrt{\left\langle r_0^2 \right\rangle}$ towards denaturant binding. Fitting the experimentally

determined root mean square end-to-end distances to eq. (10) yielded parameter values of

$\sqrt{\left\langle r_0^2 \right\rangle^0} = (18.9 \pm 0.9)$ Å, $K_{ex} = (12.0 \pm 2.7)$ and $\gamma = (1.46 \pm 0.13)$.

A different behaviour is found for the end-to-end diffusion constants. A value of

$(3.6 \pm 1.0) \cdot 10^{-7}$ cm^2 s^{-1} is obtained at 0 M GdmCl and maximum diffusion constants of around

$8.4 \cdot 10^{-7}$ cm^2 s^{-1} are reached at denaturant concentrations around 4 M ($X_D = 0.12$). At higher

GdmCl concentrations a slight decrease in D is observed leading to a value of $(6.0 \pm 1.0) \cdot 10^{-7}$

cm^2 s^{-1} at 8 M GdmCl (Fig. 3B and Table 3). In our previous studies we observed that the rate

constant for end-to-end loop formation in a (GS)$_{16}$ peptide is inversely proportional to

macroscopic solvent viscosity η.[16] Accordingly, we corrected the end-to-end diffusion

constants for solvent viscosity (eq. (11)), and found that the corrected values increase

hyperbolically with denaturant concentration. This dependency can be described by a weak

binding model according to eq. (12).

$$D' = \left(\frac{\eta}{\eta_0} \right) \cdot D \tag{11}$$

$$D' = D'^0 \cdot \left(1 + \gamma \frac{(K_{ex}-1)X_D}{(K_{ex}-1)X_D + 1} \right) \tag{12}$$

Data fitting allowed us to determine parameter values of $D^{\prime 0} = (3.7 \pm 0.9)\cdot 10^{-7}$ $\mathrm{cm}^2\,\mathrm{s}^{-1}$, $K_{ex} = (5.9 \pm 4.0)$ and $\gamma = (5.6 \pm 2.3)$.

Discussion

Effect of guanidinium chloride binding on chain dimensions

Our results show that the average dimensions of the studied model peptide increase with denaturant concentration in a hyperbolic manner. At 8 M GdmCl the root mean square end-to-end distance is more than twice than the value in the absence of denaturant. The dependence of the dimensions of the peptide chain on the denaturant concentration can be described by a Schellman weak binding model indicating that binding of GdmCl to the peptide is the origin of its extension (Fig. 3A). We compared the average dimensions determined for the $(GS)_{16}$ peptide at varying concentrations of GdmCl to the corresponding data on proteins unfolded by GdmCl. Damaschun *et al.*[32] measured and compiled small-angle X-ray scattering (SAXS) data on 12 proteins with no or reduced disulfide bonds unfolded in GdmCl and found that the observed radii of gyration, R_g, are a power function of the number of residues N (eq. (13)).

$$R_g = R_{g0} \cdot N^{\nu} \tag{13}$$

Damaschun and coworkers[32] reported parameter values of $R_{g0} = (4.4 \pm 0.3)$ Å and $\nu = 0.5 \pm 0.02$. In a recent study Kohn *et al.*[33,34] evaluated SAXS data from 28 proteins unfolded by GdmCl and determined parameter values of $R_{g0} = (2.08 \pm 0.19)$ Å and $\nu = 0.598 \pm 0.028$. Assuming a Gaussian chain[35,36], the radii of gyration can be converted to root mean square end-to-end distances according to eq. (14).

$$R_g^2 = 1/6 \cdot \langle r_0^2 \rangle \tag{14}$$

For a hypothetical GdmCl-unfolded protein of 33 residues, corresponding to our model peptides, $\langle r_0^2 \rangle$-values of 61.9 Å and 41.2 Å are calculated, using the different models. These values are in good agreement with our results, which gave $\langle r_0^2 \rangle$-value of 39.2 Å at 8 M GdmCl. The slightly higher values observed in proteins can be ascribed to the high content of glycine residues in our model peptide which leads to shorter end-to-end distances[35].

Further, we compared our experimental data to predictions from polymer theory. Flory and coworkers[35] developed a formalism to calculate the average dimensions of real polymers. Although the validity of the underlying "isolated-pair hypothesis" has been recently challenged[37], Flory's theory still provides a useful estimate. We calculated a root mean square end-to-end distance of 37.6 Å with the conformational energies of serine residues approximated by those reported for alanine[35]. However, this estimate should only apply to Θ-conditions where a real polymer chain behaves like an ideal chain[35]. Under these conditions repulsive steric interactions (excluded volume) are exactly compensated by intramolecular attractive forces within the polymer. To assess the influence of the repulsive steric interactions we conducted all-atom Monte-Carlo simulations of sterically allowed peptide conformations as described before[21]. As these simulations employ a simple hard-sphere potential, they only consider steric, i.e. repulsive interactions but no attractive forces. In a $(GS)_{16}$ peptide with the labels used for TTET experiments attached at its termini (see Materials and Methods) we found a root mean square distance between the spectroscopic labels of (43.9 ± 1.0) Å. This value is in close agreement with the limiting value of (46.4 ± 1.5) Å of the binding isotherm for the effect of GdmCl on the average peptide dimensions (Fig. 3A, Table 3). Taken together these results show that water is a bad solvent for peptide chains. In contrast, aqueous solutions of guanidinium chloride constitute a good

solvent for unstructured peptides. Binding of GdmCl to the peptide abolishes intramolecular attractive forces and thus causes an extension of the peptide chain.

Effect of GdmCl on the internal flexibility of unfolded polypeptide chains

Our findings indicate that the viscosity-corrected end-to-end diffusion constants D' increase with denaturant concentration and approach a limiting value at high denaturant concentrations. The observed dependency can be described by a Schellman weak binding model albeit with a lower binding constant than determined for the denaturant effect on the average dimensions of the peptide. This discrepancy in the binding constants may be explained by a different dependence of the root mean square end-to-end distance and diffusion on the degree of denaturant binding. The viscosity-corrected end-to-end diffusion constants should specifically report on internal motions and on energy barriers between different conformations. Thus, denaturant binding seems to lead to a substantial increase in internal mobility of the peptide which is only partly offset by the increase in macroscopic solvent viscosity induced by the addition of high amounts of denaturants. This result is consistent with our above observation that GdmCl binding weakens intramolecular attractive forces in peptides. In the absence of such interactions faster configurational rearrangements become possible. These findings also concur well with previous TTET experiments which imply increased chain mobility in the presence of GdmCl[18,38].

Correlation of the results from FRET experiments with dynamics of intrachain loop formation

Our previous studies on the dynamics of intrachain loop formation in poly-(glycine-serine) peptides showed that the effect of denaturants on the rate constant of end-to-end contact formation, k_c, can be divided into two components according to eq. (15). In part, the slower dynamics are due to an increase in solvent viscosity. The remaining effect could be described by binding of denaturant molecules to the polypeptide chain[16].

$$k_c = k_c^0 \cdot (\eta/\eta_0)^\beta \cdot \left(1 - \gamma \frac{(K_{ex}-1)X_D}{1+(K_{ex}-1)X_D}\right) \tag{15}$$

Here, k_c^0 denotes the rate constant of end-to-end contact formation in the absence of

denaturants at the reference solvent viscosity η_0. The coefficients β and γ reflect the

sensitivity of contact formation towards viscosity and denaturant binding, respectively. The

constant K_{ex} for GdmCl binding to the peptide was determined to be 32.6 ± 0.4 on the mole

fraction scale. β is a function of peptide length and is essentially equal -1.0 for a $(GS)_{16}$

peptide. γ was found to be 0.68 ± 0.01 independent of peptide length. To eliminate the

contributions from solvent viscosity, we correlated the viscosity-corrected rate constants, k_c',

observed for a $(GS)_{16}$ peptide at different concentrations of GdmCl with the average peptide

dimensions measured by FRET under the same experimental conditions. Fig. 4A shows that

the two quantities which were determined in independent experiments are clearly correlated.

Fitting these data to eq. (17) yielded a slope for the correlation of $\rho = (-0.63 \pm 0.04)$.

$$\ln\left(k_c'/k_c'^0\right) = \rho \cdot \ln\left(\langle r_0^2 \rangle \big/ \langle r_0^2 \rangle^0\right) \tag{17}$$

It is interesting to note that k_c' is correlated to the viscosity-corrected diffusion constants D' in

the same manner (Fig. 4B). The corresponding correlation has a slope of $\rho = (-0.62 \pm 0.08)$.

These data imply that the average dimensions and the diffusion constants of unstructured

peptides are affected by denaturant binding in the same way. In contrast, the end-to-end

contact formation rate constants depend on denaturant binding in a different way.

Accordingly, different denaturant binding constants were obtained for the different quantities.

We compared the correlation between end-to-end distance and the dynamics of loop formation (ρ) to predictions from theory. Szabo, Schulten and Schulten[28] developed a theory for intramolecular end-to-end contact formation in polymers (SSS-theory). For polymers in which only a small fraction of all molecules makes contact simultaneously and in which the conformations not forming contact equilibrate fast the theory predicts a single relaxation time constant, τ, for contact formation (eq. (18)).

$$\Sigma(t) \approx \Sigma(t)_{\text{approx.}} = \exp(-t/\tau) \tag{18}$$

In agreement with the theory, we observed single exponential loop closure kinetics in TTET experiments.[16,17,18] The theory of Szabo *et al.*[28] further correlates this single observable relaxation time constant for contact formation with the underlying end-to-end distance probability distribution, $p(x)$, of the polymer (eq. (19)).

$$\tau = \frac{1}{D} \int\limits_{r_b}^{\infty} \frac{1}{p(x)} \left(\int\limits_{x}^{\infty} p(y)dy \right)^2 dx \left/ \int\limits_{r_b}^{\infty} p(x)dx \right. \tag{19}$$

The reactive boundary r_b designates the end-to-end distance at and below which a reaction occurs, i.e. contact is formed. For end-to-end probability distributions of the kind used presently (eq. (6)) and a given fixed diffusion constant the coefficient ρ in eq. (17) should assume a value of approximately -1.5. The considerably higher value of -0.63 obtained in our FRET experiments is clearly due to the effect of denaturants on the diffusion constant for chain motions (Figs. 3B and 4B).

We employed the values for the chain diffusion constants and probability distributions determined in this work to calculate rate constants for intramolecular end-to-end contact

formation according to the SSS-theory. The only unknown quantity in eq. (19) is the reactive boundary r_b. We found best agreement between calculated rate constants and experimentally observed ones, when a value of 4.2 Å was used as the reactive boundary which seems plausible considering the dimensions of the spectroscopic labels used in the TTET experiments. This value agrees nicely with calculations based on TTET from xanthone to naphthalene which has a bimolecular quenching rate constant of $k_{bi} \approx 4 \cdot 10^9$ M^{-1}s^{-1}.[18] Comparing this value with diffusion constants for naphthalene and xanthone of $9.0 \cdot 10^{-6}$ cm^{2}s^{-1} and $5.5 \cdot 10^{-6}$ cm^{2}s^{-1}, respectively, yields a reactive boundary of 3.6 Å based on the theory developed by von Smoluchowski[39,40]. Fig. 5 shows that the rate constants calculated with a reactive boundary of 4.2 Å are in very good agreement with the measured values. A linear correlation yields a slope of 1.01 ± 0.05 and an intercept of essentially zero [$(-7.2 \cdot 10^4 \pm 2.4 \cdot 10^5)$ s^{-1}]. The linear correlation coefficient is 0.994. The fact that two independent sets of experiments employing two different spectroscopic techniques yield very similar results indicates that the SSS-theory is able to sufficiently describe dynamics of end-to-end loop formation in long polypeptide chains.

Implications for FRET studies of flexible systems

Our results show that it is crucial to take into account dynamical processes occurring on the timescale of the donor fluorescence lifetime when studying FRET in unstructured and thus flexible systems. Both close interchromophore distances and fast diffusion of the two labels relative to each other lead to high FRET efficiencies. This holds true for steady-state and for time-resolved FRET measurements. When the dynamic effect is not properly taken into account, the distances derived from transfer efficiencies systematically underestimate the real donor-acceptor distances. This effect is expected to occur most pronouncedly for donor fluorophores with long fluorescence lifetimes.

Especial care should be taken when external variables are changed which might affect the interchromophore diffusion constant. Although for most complex experimental systems it is not precisely clear how these diffusion constants depend on external variables a significant contribution of macroscopic solvent viscosity is expected. Denaturants lead to an increase in solvent viscosity and can further affect the internal dynamics of biomolecules beyond this effect as shown in this work. Solvent viscosity is also a function of temperature. At ambient temperatures the temperature-dependence of water viscosity alone shows an apparent activation energy of about 17.7 kJ mol^{-1}.[21,41] Thus, results from temperature-dependent FRET experiments, including temperature jump studies, should be corrected for this.

Muñoz and coworkers[42] reported that the FRET efficiency in the fluorescent-labelled acid-denatured form of the protein BBL increases from about 0.21 at 280 K to 0.45 at 360 K and interpreted these findings as a decrease of the root mean square interchromophore distance from about 32 Å to 20 Å over this temperature range. However, the viscosity of pure water decreases by a factor of 4.3 from 1.43 cP to 0.33 cP over this temperature interval. It is not clear how this translates into changes of the interchromophore diffusion constant and how this affects the reported distances.

In summary, the method we present here shows that the use of two donor/acceptor pairs with different donor lifetimes can be utilized to resolve the influence of distance and diffusion in FRET measurements. In steady-state experiments the use of two donor/acceptor pairs could serve at least as a qualitative indicator for diffusional influence on the FRET efficiency. When FRET efficiencies obtained with two suitably chosen chromophore pairs coincide, diffusional processes should have no significant influence on the measurements. Conceptually similar methods suitable to resolve average interchromophore distance and diffusion in FRET measurements on flexible systems were proposed before by Haas and coworkers[27] and

Lakowicz's group. Fluorescence data were measured at different solvent viscosities[43] or in the presence of different amounts of fluorescence quenchers[44] and were globally analyzed. Lakowicz et al.[45] further suggested the parallel use of fluorescent donor-acceptor and donor-quencher pairs as spectroscopic labels. Finally, in steady-state experiments Wiczk et al.[46] also employed different FRET chromophore pairs to vary the Förster distance R_0.[47]

Implications for protein folding

Our previous findings on end-to-end contact formation dynamics in unstructured peptides showed that the denaturant dependence of protein folding rates contains significant contributions from the effect on intrachain diffusion in the unfolded state. The results from FRET measurements reported in the present work provide insight into the effect of denaturants on chain dynamics at the molecular level. The results have shown that the effect of denaturants on peptide chain dynamics comprises several components. The increased macroscopic solvent viscosity in denaturant solutions impedes chain motions. In addition to the viscosity-effect, denaturants bind to unfolded peptide chains. This causes an extension of average chain dimensions and an increase in the internal mobility of unfolded peptide chains. The substantial increase in end-to-end distance from 18.7 Å in water to 39.2 Å in 8 M GdmCl is surprising, considering that the model peptides contain nearly exclusively glycine and serine residues. This should result in very flexible chains which do not exhibit strong side chain interactions. Our results therefore imply that flexible unfolded polypeptide chains in water form significant amount of intramolecular interactions between groups of the polypeptide backbone. These interactions are weakened upon addition of GdmCl in accordance with a change from a bad to a good solvent. Since this effect is observed for the flexible glycine-serine based model peptides, it is most likely a general effect of denaturants on the dimensions of a polypeptide chain caused by interactions of denaturants with the polypeptide backbone. This strong and non sequence-specific reduction of the chain

dimensions in water compared to denaturant solutions has to be taken into account when FRET measurements on unfolded states of proteins are interpreted in terms of the presence of folding intermediates[48].

Acknowledgements

Tables

Table 1 – Properties of the fluorophore systems at different denaturant concentrations.

[GdmCl] (M)	naphthalene → dansyl		pyrene → dansyl	
	τ_D (10^{-9} s) [a]	R_0 (Å) [b]	τ_D (10^{-9} s) [a]	R_0 (Å) [b]
0	36.9 ± 0.2	23.3 ± 0.4	225.5 ± 1.0	20.5 ± 0.4
1	32.9 ± 0.2	23.0 ± 0.4	214.9 ± 1.0	19.9 ± 0.4
2	29.5 ± 0.2	22.4 ± 0.4	205.4 ± 1.0	19.8 ± 0.4
3	26.5 ± 0.2	22.0 ± 0.4	196.2 ± 1.0	19.8 ± 0.4
4	24.1 ± 0.2	21.4 ± 0.4	187.0 ± 1.0	19.8 ± 0.4
5	22.7 ± 0.2	21.1 ± 0.4	180.3 ± 1.0	19.9 ± 0.4
6	20.8 ± 0.2	20.8 ± 0.4	174.5 ± 1.0	20.0 ± 0.4
7	19.6 ± 0.2	20.6 ± 0.4	167.8 ± 1.0	20.1 ± 0.4
8	18.6 ± 0.2	20.2 ± 0.4	162.0 ± 1.0	20.2 ± 0.4

[a] Determined in fluorescence lifetime measurements for the corresponding donor-only peptide.

[b] Calculated according to eq. (2) from steady-state fluorescence properties of the fluorophore system.

Table 2 – Apparent fluorescence lifetimes and transfer efficiencies in the FRET peptides at different denaturant concentrations.

[GdmCl]	naphthalene → dansyl			pyrene → dansyl		
(M)	$\tau_{app.}$ (10^{-9} s) [a]	$E_{app.}$ [b]	$<r>_{app.}$ (Å) [b]	$\tau_{app.}$ (10^{-9} s) [a]	$E_{app.}$ [b]	$<r>_{app.}$ (Å) [b]
0	13.8	0.62	21.4	35.3	0.84	15.5
1	14.6	0.56	22.1	40.7	0.81	15.7
2	15.7	0.47	22.9	50.6	0.75	16.4
3	15.9	0.40	23.5	55.4	0.72	16.9
4	16.3	0.32	24.1	60.6	0.68	17.5
5	15.9	0.30	24.3	68.4	0.62	18.4
6	15.7	0.25	25.1	71.2	0.59	18.8
7	15.4	0.21	25.6	77.7	0.54	19.6
8	15.0	0.19	25.7	82.0	0.49	20.3

[a] Time-resolved FRET traces were approximated by single exponential decay curves according to eq. (9).

[b] Calculated according to eqs. (3) and (4).

Table 3 – Average molecular dimensions, diffusion coefficients and viscosity-corrected contact formation rates in a $(GS)_{16}$ peptide determined at different denaturant concentrations.

[GdmCl] (M)	$<r_0^2>^{1/2}$ (Å) [a]	D (10^{-7} cm^2 s^{-1}) [a]	k_c' (s^{-1}) [b]
0	18.7 ± 1.0	3.6 ± 1.0	$1.56\cdot10^7$
1	24.0 ± 1.0	5.4 ± 1.0	$1.16\cdot10^7$
2	27.6 ± 1.0	6.8 ± 1.0	$9.64\cdot10^6$
3	29.8 ± 1.0	6.7 ± 1.0	$8.48\cdot10^6$
4	33.1 ± 1.0	8.2 ± 1.0	$7.70\cdot10^6$
5	34.4 ± 1.0	8.4 ± 1.0	$7.16\cdot10^6$
6	36.4 ± 1.0	7.5 ± 1.0	$6.74\cdot10^6$
7	38.2 ± 1.0	7.5 ± 1.0	$6.42\cdot10^6$
8	39.2 ± 1.0	6.0 ± 1.0	$6.15\cdot10^6$

[a] Determined by globally fitting time-resolved FRET data from $(GS)_{16}$ peptides with the two fluorophore systems.

[b] Calculated according to eqs. (15) and (16) based on the parameter values determined by Krieger *et al.*[18] and Möglich *et al.*[16]

References

1.   Greenstein, J. P. (1938). Sulfhydryl groups in proteins. *J. Biol. Chem.* **125**, 501-513.

2.   Spiro, K. (1900). Über die Beeinflußung der Eiweißcoagulation durch stickstoffhaltige Substanzen. *Z. Physiol. Chem.* **30**, 182-199.

3.   Ramsden, W. (1902). Some new properties of urea. *J. Physiol.* **28**, 23-27.

4.   Tanford, C. (1968). Protein denaturation. *Adv. Prot. Chem.* **23**, 121-282.

5.   Tanford, C. (1970). Protein denaturation. Part C. Theoretical models for the mechanism of denaturation. *Adv. Prot. Chem.* **24**, 1-95.

6.   Pace, C. N. & Tanford, C. (1968). Thermodynamics of the unfolding of β-lactoglobulin A in aqueous urea solutions between 5 and 55°. *Biochemistry* **7**, 198-208.

7.   Alexander, S. S. & Pace, C. N. (1971). A comparison of the denaturantion of bovine-lactoglobulins A and B and goat-lactoglobulin. *Biochemistry* **10**, 2738-2743.

8.   Greene, R. F. J. & Pace, C. N. (1974). Urea and guanidine-hydrochloride denaturation of ribonuclease, lysozyme, alpha-chymotrypsin and beta-lactoglobulin. *J. Biol. Chem.* **249**, 5388-5393.

9.   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**, 8063-8068.

10.   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.

11.   Pace, C. N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Meth. Enzymol.* **131**, 266-280.

12.   Pace, C. N., Grimsley, G. R. & Scholtz, J. M. (2005). Denaturation of Proteins by Urea and Guanidine Hydrochloride. In *Protein Folding Handbook* (Buchner, J. & Kiefhaber, T., eds.), Vol. 1, pp. 45-69. 5 vols. Wiley-VCH, Weinheim, Germany.

13.   Tanford, C. (1964). Isothermal unfolding of globular proteins in aqueous urea solutions. *J. Am. Chem. Soc.* **86**, 2050-2059.

14.   Aune, K. C. & Tanford, C. (1969). Thermodynamics of the denaturation of lysozyme by guanidine hydrochloride. II. Dependence on denaturant concentration at 25 degrees. *Biochemistry* **11**, 4586-4590.

15.   Schellman, J. A. (2002). Fifty years of solvent denaturation. *Biophys. Chem.* **96**, 91-101.

16.   Möglich, A., Krieger, F. & Kiefhaber, T. (2005). Molecular Basis for the Effect of Urea and Guanidinium Chloride on the Dynamics of Unfolded Polypeptide Chains. *J. Mol. Biol.* **345**, 153-162.

17.   Bieri, O., Wirz, J., Hellrung, B., Schutkowski, M., Drewello, M. & Kiefhaber, T. (1999). The speed limit for protein folding measured by triplet-triplet energy transfer. *Proc. Natl. Acad. Sci. USA* **96**, 9597-9601.

18.   Krieger, F., Fierz, B., Bieri, O., Drewello, M. & Kiefhaber, T. (2003). Dynamics of unfolded polypeptide chains as model for the earliest steps in protein folding. *J. Mol. Biol.* **332**, 265-274.

19.   Fierz, B. & Kiefhaber, T. (2005). Dynamics of Unfolded Polypeptide Chains. In *Protein Folding Handbook* (Buchner, J. & Kiefhaber, T., eds.), Vol. 2, pp. 809-855. 5 vols. Wiley-VCH, Weinheim, Germany.

20.   Navon, A., Ittah, V., Landsman, P., Scheraga, H. A. & Haas, E. (2001). Distributions of intramolecular distances in the reduced and denatured states of bovine pancreatic ribonuclease A. Folding initiation structures in the C-terminal portions of the reduced protein. *Biochemistry* **40**, 105-118.

21.   Krieger, F., Möglich, A. & Kiefhaber, T. (2005). Effect of proline and glycine residues on dynamics and barriers of loop formation in polypeptide chains. *J. Am. Chem. Soc.* **127**, 3346-3352.

22.   Landolt-Börnstein. (1987). *Structure Data of Free Polyatomic Molecules*, II/15, Springer, Heidelberg, Germany.

23.   Landolt-Börnstein. (1992). *Structure Data of Free Polyatomic Molecules*, II/21, Springer, Heidelberg, Germany.

24.   Förster, T. (1948). Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann. Phys.* **2**, 55-75.

25.   Stryer, L. & Haugland, R. P. (1967). Energy transfer: a spectroscopic ruler. *Proc. Natl. Acad. Sci. USA* **58**, 719-726.

26.   Edwards, S. F. (1965). The statistical mechanics of polymers with excluded volume. *Proc. Phys. Soc. London* **85**, 613-624.

27.   Haas, E., Katchalski-Katzir, E. & Steinberg, I. Z. (1978). Brownian Motion of the Ends of Oligopeptide Chains in Solution as Estimated by Energy Transfer Between the Chain Ends. *Biopolymers* **17**, 11-31.

28. Szabo, A., Schulten, K. & Schulten, Z. (1980). First passage time approach to diffusion controlled reactions. *J. Chem. Phys.* **72**, 4350-4357.

29. Schellman, J. A. (1975). Macromolecular binding. *Biopolymers* **14**, 999-1018.

30. Schellman, J. A. (1978). Solvent denaturation. *Biopolymers* **17**, 1305-1322.

31. Schellman, J. A. (1987). Selective binding and solvent denaturation. *Biopolymers* **26**, 549-559.

32. Damaschun, G., Damaschun, H., Gast, K. & Zirwer, D. (1998). Denatured states of yeast phosphoglycerate kinase. *Biochemistry (Mosc)* **63**, 259-275.

33. Kohn, J. E., Millett, I. S., Jacob, J., Zagrovic, B., Dillon, T. M., Cingel, N., Dothager, R. S., Seifert, S., Thiyagarajan, P., Sosnick, T. R., Hasan, M. Z., Pande, V. S., Ruczinski, I., Doniach, S. & Plaxco, K. W. (2004). Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci. USA* **101**, 12491-12496.

34. Fitzkee, N. C. & Rose, G. D. (2004). Reassessing random-coil statistics in unfolded proteins. *Proc. Natl. Acad. Sci. USA* **101**, 12497-12502.

35. Flory, P. J. (1988). *Statistical mechanics of chain molecules*, Hanser, Munich, Germany.

36. Cantor, C. R. & Schimmel, P. R. (1980). *Biophysical Chemistry, Part 3: The behavior of biological macromolecules*, 3. 3 vols.

37. Pappu, R. V., Srinivasan, R. & Rose, G. D. (2000). The Flory isolated-pair hypothesis is not valid for polypeptide chains: implications for protein folding. *Proc. Natl. Acad. Sci. USA* **97**, 12565-12570.

38. Krieger, F. (2004). Dynamics in Unfolded Polypeptide Chains as Model for Elementary Steps in Protein Folding, Biozentrum, University of Basel.

39. von Smoluchowski, M. (1906). Zur kientischen Theorie der Brownschen Molekularbewegung und der Suspensionen. *Ann. Phys.* **21**, 756-780.

40. von Smoluchowski, M. (1916). Drei Vorträge über Diffusion, Brownsche Molekularbewegung und Koagulation von Kolloidteilchen. *Phys. Z.* **17**, 557-571.

41. Lide, D. R. (2004). *CRC Handbook of Chemistry and Physics*, CRC Press, Boca Raton, FL, USA.

42. Sadqi, M., Lapidus, L. J. & Munoz, V. (2003). How fast is protein hydrophobic collapse? *Proc. Natl. Acad. Sci. USA* **100**, 12117-12122.

43.   Lakowicz, J. R., Kusba, J., Wiczk, W., Gryczynski, I. & Johnson, M. L. (1990). End-to-end diffusion of a flexible bichromophoric molecule observed by intramolecular energy transfer and frequency-domain fluorometry. *Chem. Phys. Lett.* **173**, 319-326.

44.   Lakowicz, J. R., Kusba, J., Gryczynski, I., Wiczk, W., Szmacinski, H. & Johnson, M. L. (1991). End-to-end Diffusion and Distance Distributions of Flexible Donor-Acceptor Systems Observed by Intramolecular Energy Transfer and Frequency-Domain Fluorometry; Enhanced Resolution by Global Analysis of Externally Quenched and Nonquenched Samples. *J. Phys. Chem.* **95**, 9654-9660.

45.   Lakowicz, J. R., Kusba, J., Szmacinski, H., Gryczynski, I., Eis, P. S., Wiczk, W. & Johnson, M. L. (1991). Resolution of end-to-end diffusion coefficients and distance distributions of flexible molecules using fluorescent donor-acceptor and donor-quencher pairs. *Biopolymers* **31**, 1363-1378.

46.   Wiczk, W., Eis, P. S., Fishman, M. N., Johnson, M. L. & Lakowicz, J. R. (1991). Distance Distributions Recovered from Steady-State Fluorescence Measurements on Thirteen Donor-Acceptor Pairs with Different Förster Distances. *J. Fluoresc.* **1**, 273-286.

47.   Cantor, C. R. & Pechukas, P. (1971). Determination of Distance Distribution Functions by Singlet-Singlet Energy Transfer. *Proc. Natl. Acad. Sci. USA* **68**, 2099-2101.

48.   Magg, C. & Schmid, F. X. (2004). Rapid collapse precedes the fast two-state folding of the cold shock protein. *J. Mol. Biol.* **335**, 1309-1323.

Figure Captions

Fig. 1 – (A) Time-resolved naphthalene fluorescence monitored at 325 nm after excitation with a 1.5 ns flashlamp pulse. The fluorescence lifetime observed in the donor-only peptide (black trace) is decreased in the $(GS)_{16}$ peptide which bears the FRET acceptor chromophore dansyl (blue). Fluorescence traces have been scaled to maximum signal. The red line represents a fit of the donor-only data to a single-exponential function. (B) Corresponding fluorescence decays of pyrene-labelled donor-only (black) and FRET peptides (blue). Time-resolved fluorescence was monitored at 377 nm. Again, the fluorescence lifetime in the $(GS)_{16}$ peptide bearing donor and acceptor chromophore is markedly shorter than in the corresponding donor-only peptide. Note, however, that the relative decrease in fluorescence lifetime is more pronounced in the peptide bearing the longer-lived pyrene chromophore compared to naphthalene. The orange curves in (A) and (B) represent a global fit of both FRET traces to eqs. (6-8).

Fig. 2 – End-to-end distance probability distribution functions in the $(GS)_{16}$ peptide determined at different concentrations of GdmCl. From left to right results for GdmCl concentrations of 0, 1, 2, 3, 4, 5, 6, 7, 8 M are displayed. On average higher denaturant concentrations lead to an extension of the peptide chain.

Fig. 3 – (A) Root mean square end-to-end distances (RMSD) as a function of the mole fraction of GdmCl. RMSD values were calculated from the probability distributions shown in Fig. 2. The solid line represents a fit to a weak binding model (eq. (10)). (B) GdmCl dependence of the end-to-end diffusion constants (closed circles). Diffusion constants were corrected for solvent viscosity according to eq. (11) (open circles). The solid line describes a fit of the corrected diffusion constants to a weak binding model (eq. (12)).

Fig. 4 – Correlation of the viscosity-corrected end-to-end contact formation rate constants, k_c', for a $(GS)_{16}$ peptide at different concentrations of GdmCl with the corresponding mean square end-to-end distances and diffusion constants determined by FRET. (A) On the double-logarithmic plot a clear correlation between k_c' and $<r^2>$ is observed. A fit to eq. (17) shown as a solid line yields a slope of -0.63 ± 0.04. (B) Corresponding correlation of k_c' with D'. The correlation fit has a slope of -0.62 ± 0.08.

Fig. 5 – Comparison of contact formation rate constants determined by triplet-triplet energy transfer, k_c, with corresponding rate constants calculated from the end-to-end probability distributions and diffusion constants determined by FRET experiments, k_{SSS}. Rate constants k_{SSS} were calculated according to eq. (19) using a value of 4.2 Å for the reactive boundary, r_b. The solid line shows the linear correlation with a slope of 1.01 ± 0.05 and an intercept of virtually zero $[(-7.2 \cdot 10^4 \pm 2.4 \cdot 10^5)\ s^{-1}]$. The linear correlation coefficient amounts to 0.994.
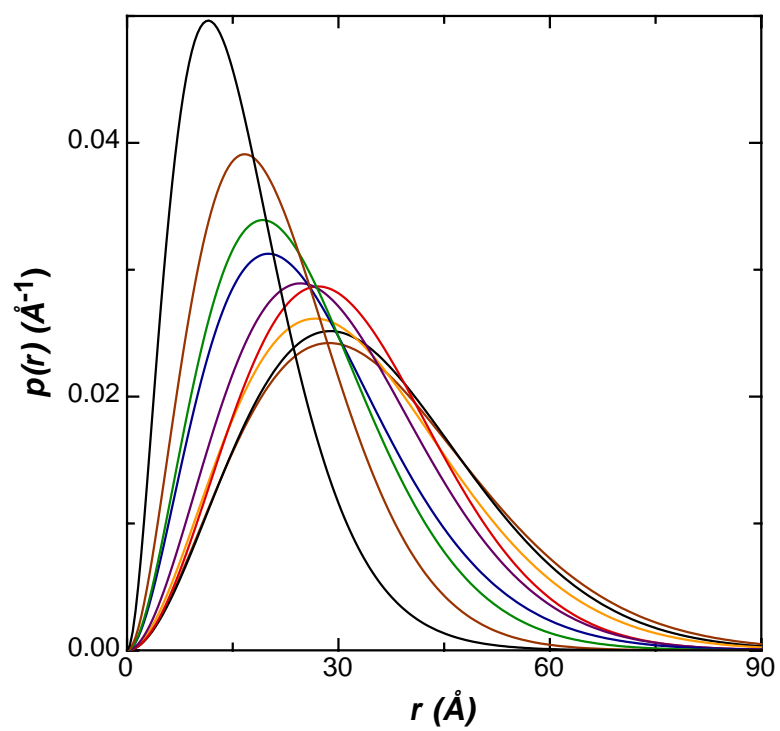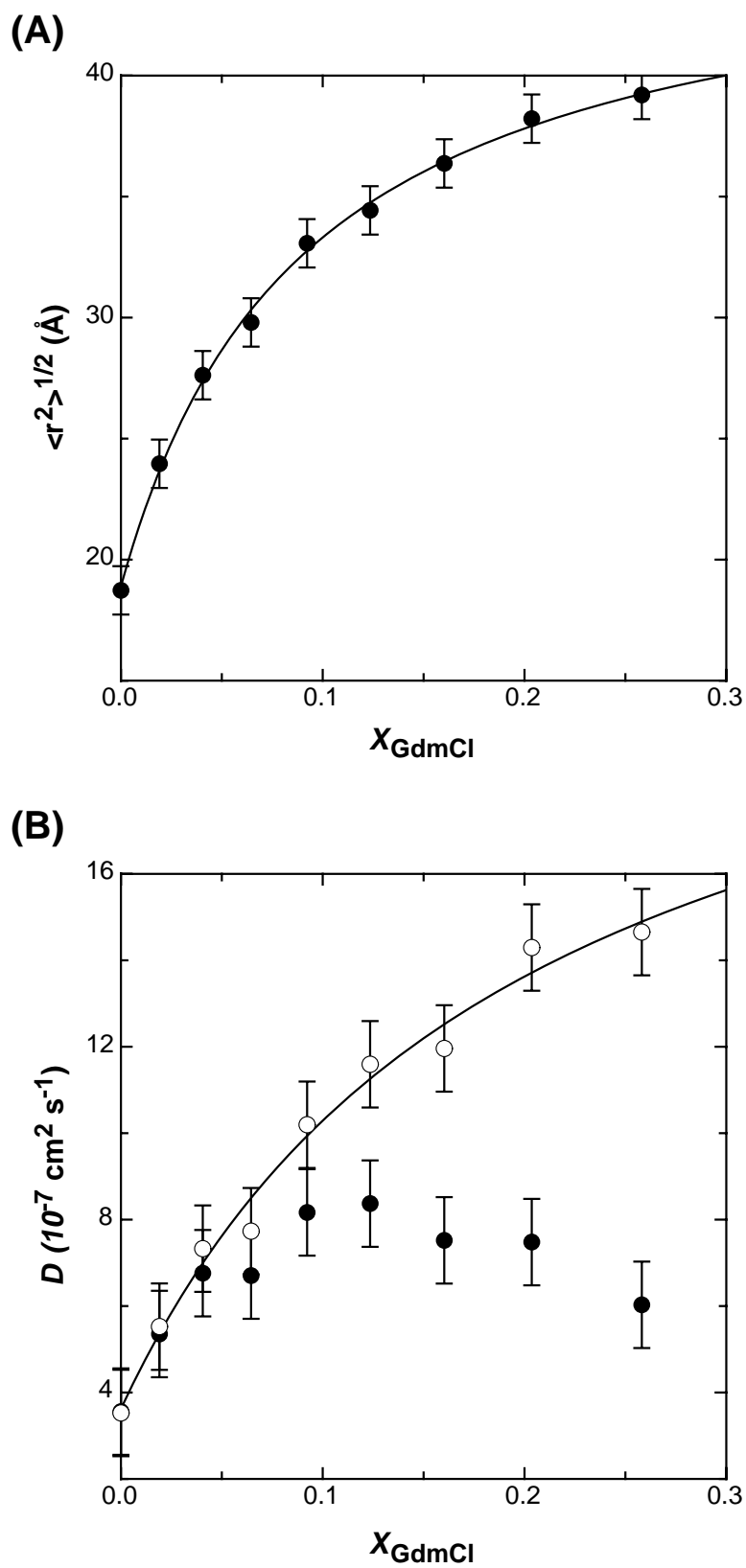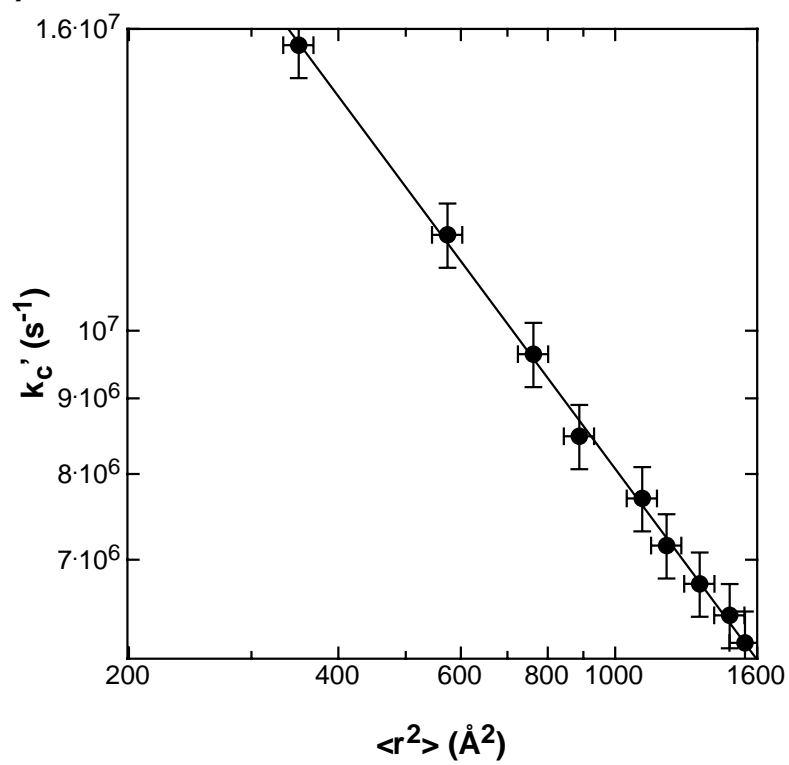
Fig. 1

Fig. 2
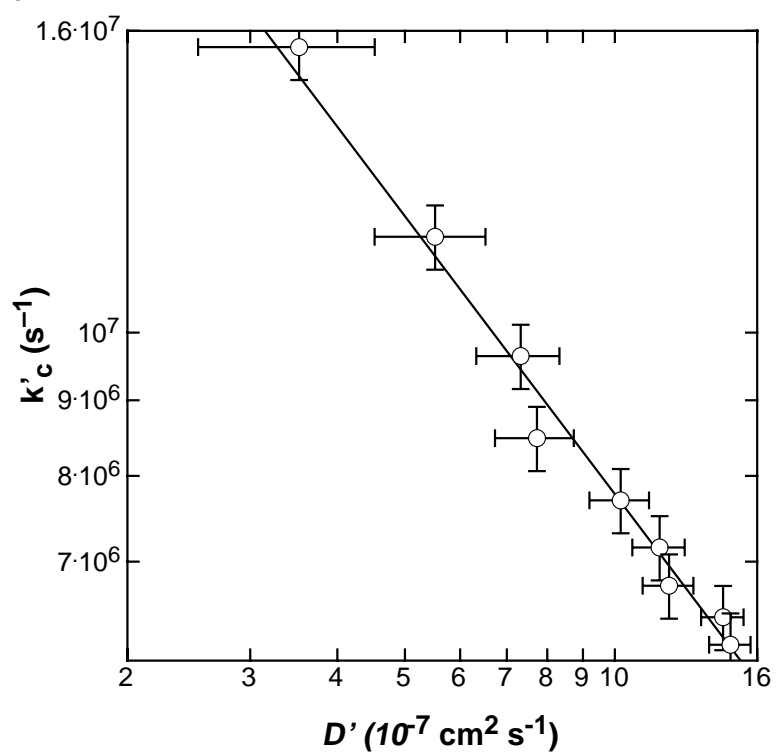
(A)

(B)
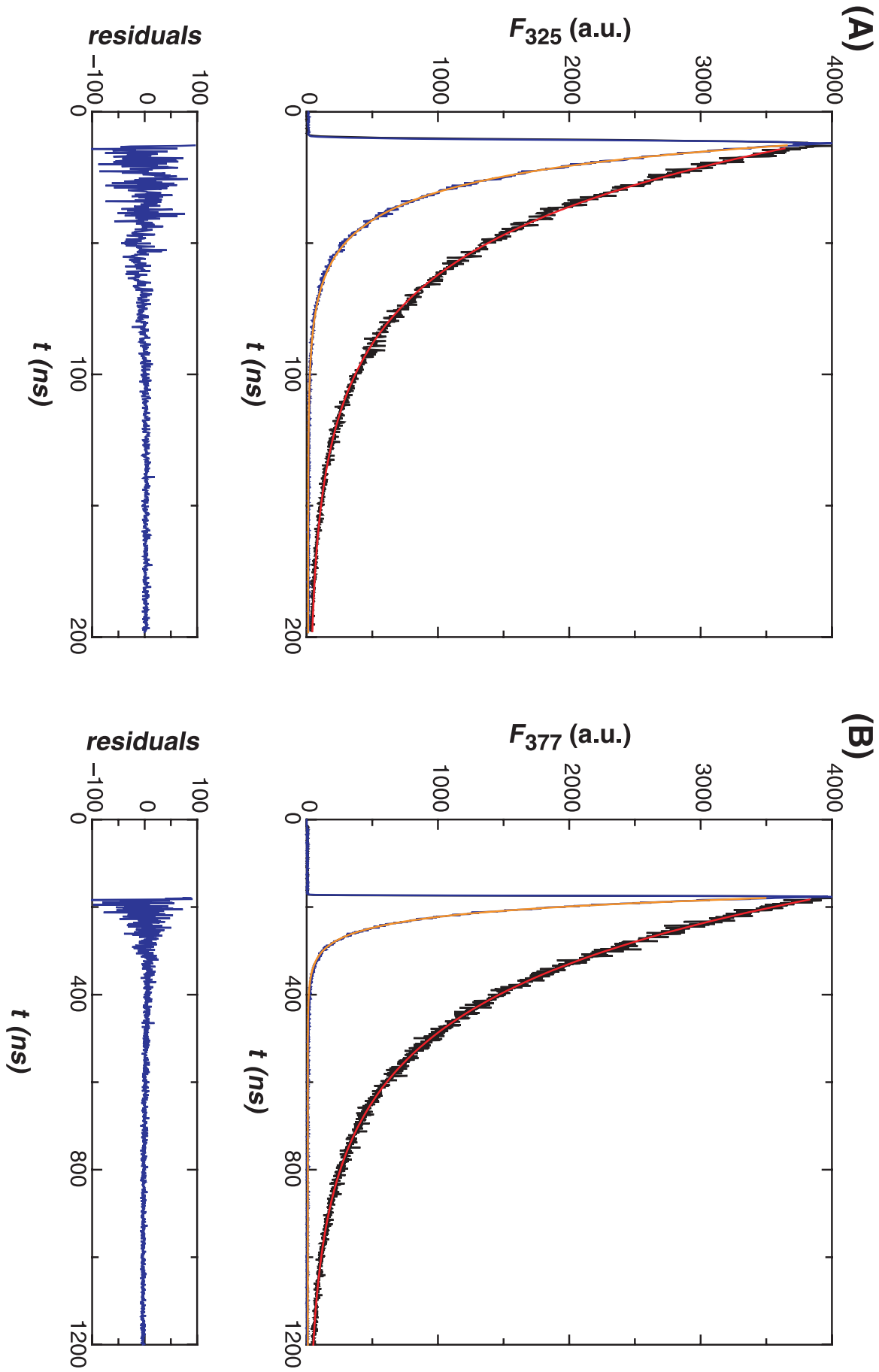
Fig. 3
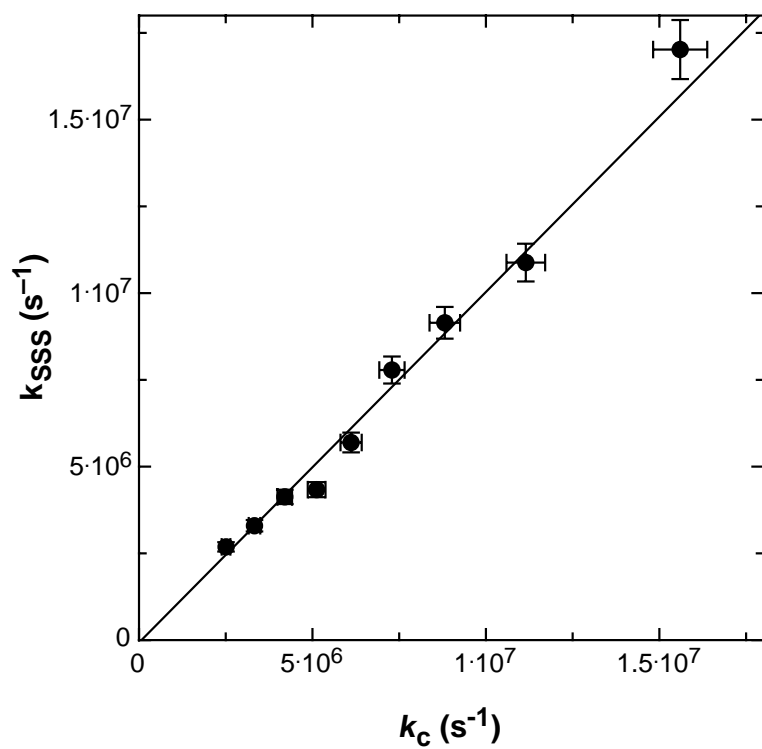
(A)

(B)

Fig. 4

Fig. 5

The Peptide Backbone is the Major Interaction Site

for Urea and Guanidinium Chloride

Andreas Möglich and Thomas Kiefhaber[*]

Biozentrum der Universität Basel, Department of Biophysical Chemistry, Klingelbergstr. 70, CH-4056 Basel, Switzerland.

[*]: corresponding author; phone: ++41 61 267 21 94, fax: ++41 61 267 21 89, email: t.kiefhaber@unibas.ch

Abstract

Despite the frequent use of chemical denaturants in protein folding, the exact mechanism by which these compounds exert their denaturing effect is not well characterized. Previous studies using triplet-triplet energy transfer (TTET) showed that intrachain contact formation in unstructured polypeptides is decelerated by denaturants. Further analysis revealed that the effect of the denaturants urea and guanidinium chloride (GdmCl) on chain dynamics results from a superposition of two components. First, denaturant addition causes an increase of solvent viscosity which impedes motions of the peptide chain. Second, denaturants bind to polypeptides which further slows down these dynamics.

Here, we report corresponding TTET studies of loop closure dynamics and their dependence on denaturant concentration for a series of host-guest peptides which contained different guest amino acids. In all peptides the logarithm of the contact formation rate constant, $\ln k_c$, decreased linearly with the concentrations of urea and GdmCl. The strongest denaturant dependencies were observed for peptides containing polar or charged residues, the weakest for peptides with aliphatic guest amino acids. For most peptides the effect of GdmCl on chain dynamics was on average 1.8 times larger than that of urea. Only the basic amino acids lysine and arginine showed remarkably weaker interactions with GdmCl. Studies of the viscosity dependence of intrachain contact formation showed that for all peptides the effect of denaturants on chain dynamics results from a combination of increased viscosity and denaturant binding. Despite different properties of their guest amino acids, all host-guest peptides displayed similar denaturant binding constants which indicates that the peptide backbone is the main interaction site for denaturants. The effect of denaturant binding was also similar for most peptides except for the ones containing basic amino acids which showed unusually weak interactions with GdmCl.

Introduction

Urea and guanidinium chloride (GdmCl) are invaluable in the study of protein stability and folding[1,2,3]. As chemical denaturants they selectively stabilize the unfolded state relative to the native state of proteins and thereby cause protein denaturation[4,5]. The free energy of protein unfolding (ΔG^0) was found to decrease linearly with the denaturant concentration [D] according to $m_{eq} = \partial \Delta G^0 / \partial [D]$.[6,7,8,9] The equilibrium m-value (m_{eq}) quantifies the sensitivity of protein stability with respect to denaturant concentration and was shown to correlate with protein size and the difference in solvent-accessible surface area between the native and the unfolded state[10]. Linear relationships were also reported for the dependence of the activation free energies of the folding ($\Delta G_f^{0\ddagger}$) and unfolding ($\Delta G_u^{0\ddagger}$) reactions on denaturant concentration[5,7], thus allowing the determination of the corresponding kinetic m-values m_f and m_u. In spite of the vast body of data on the denaturant effect on protein folding and stability, the molecular mechanism of protein denaturation by urea and GdmCl has remained elusive. Two mechanistic models for denaturant action were proposed which could describe the experimental data equally well[11]. The transfer model suggested by Tanford and coworkers postulates that denaturants increase the solubility of parts of the proteins compared to water[5,11,12]. Thus, high denaturant concentrations favour the exposure of inner parts of the protein and thereby lead to its unfolding. The alternative binding model accounts for denaturant-induced protein unfolding by direct binding of denaturant molecules to parts of the protein[11,13,14,15]. Net stabilization of the unfolded state would be effected by a larger number of binding sites and/or higher binding affinity than in the native state. Neither model specifies which parts of the protein denaturants mainly interact with. Due to the structural similarity of urea and GdmCl with the amide bond it has long been suggested that the peptide backbone is a possible interaction site for these denaturants.

We have recently studied the effect of urea and GdmCl on the dynamics of intrachain loop formation in unstructured peptides[16]. Using the triplet-triplet energy transfer (TTET) technique, we measured absolute rate constants for intramolecular contact formation in model peptides at different denaturant concentrations[16,17,18]. Analogous to the effect of denaturants on protein stability and folding rate constants, the logarithm of the rate constant of end-to-end contact formation, k_c, was found to linearly decrease with denaturant concentration. Further analysis revealed that this behaviour can be dissected into two components. First, denaturant addition leads to a substantial increase in solvent viscosity which impedes end-to-end diffusion and contact formation. Secondly, denaturant molecules can bind to the peptide, thereby further slowing down intramolecular contact formation. In a subsequent study fluorescence resonance energy transfer was employed to further characterize the interaction of denaturants with polypeptides (A. M. and T. K., in preparation). Denaturant binding was found to have two main effects. On the one hand the average dimensions of the polypeptide chain increase with denaturant concentration which accounts for the observed deceleration of end-to-end contact formation. On the other hand denaturant binding increases the internal flexibility of the polypeptide chains thereby facilitating configurational rearrangements and end-to-end diffusion.

These studies showed that a binding model can quantitatively account for the effects of urea and GdmCl on the dynamics and dimensions of unstructured polypeptides. However, it remained unclear which parts of the peptide chain denaturants mainly bind to. To test the influence of different amino acid sidechains on denaturant binding, in the present study we conducted TTET experiments on host-guest peptides. Different guest amino acids were incorporated into common host peptide scaffolds and end-to-end contact formation was measured at different concentrations of urea and GdmCl.

Materials and Methods

Peptide substrates were synthesized using standard fluorenylmethoxycarbonyl (F-moc) chemistry on an Applied Biosystems 433A peptide synthesizer (Foster City, CA, USA) or an ABIMED EPS 221 synthesizer (Langenfeld, Germany) as described[18]. As shown in Scheme 1, two series of peptides containing either three (A) or five (B) amino acids between the spectroscopic labels xanthone and naphthalene were synthesized. 9-oxoxanthen-2-carboxylic acid was coupled to the N-terminus and the nonnatural amino acid 1-naphthylalanine was introduced at the indicated positions. The letters '**X**' in Scheme 1 designate different guest amino acids which were incorporated in the common peptide scaffold. The corresponding substrates are referred to as SXS resp. SXSXS peptides according to the peptide sequence between the spectroscopic labels. Peptides were purified via reversed-phase HPLC on a RP-8-HPLC column (Merck, Darmstadt, Germany) using either 0.1 % trifluoroacetic acid (TFA)/H_2O and 0.1 % TFA/acetonitrile, or 50 mM ammonium acetate buffer pH 6.0 and acetonitrile as eluents. Correct peptide mass and higher than 95 % purity were confirmed by mass spectrometry and analytical HPLC, respectively.

Triplet-triplet energy transfer measurements were performed with a Laser Flash Photolysis Reaction Analyzer (LKS.60) from Applied Photophysics (Leatherhead, UK) as described[18]. Following a 4 ns laser flash ($\lambda = 355$ nm, ~ 50 mJ pulse energy), transient absorbance of the xanthone triplet state was monitored at 590 nm. Contact formation rate constants were determined by fitting the absorption decays to exponential functions. All experiments were conducted in solutions containing 10 mM potassium phosphate buffer (pH 7.0) and between 20 and 100 μM peptide. Prior to measurement samples were filtered and equilibrated at 22.5 °C. Concentrations of guanidinium chloride and urea were determined by refractive index measurements according to Pace[11,19]. To increase solvent viscosity, glycerol was added as a cosolvent. Viscosities of these solutions were determined with a falling-ball viscosimeter

(Haake, Germany). The viscosities of denaturant solutions were calculated according to the empirical formula provided by Perl *et al.*[20]

Data evaluation was carried out as before[16] using the computer programs ProFit (QuantumSoft, Zürich, Switzerland) and Matlab (The MathWorks, Natick, NA, USA).

Results

Intramolecular contact formation in host-guest peptides

Using triplet-triplet energy transfer (TTET), we studied intramolecular loop formation and its sensitivity towards denaturant concentration in short host-guest peptides containing different amino acids. Guest amino acids were incorporated at distinct positions in either one of two common host peptide contexts. In one series the two spectroscopic labels xanthonic acid and naphthalene were separated by the canonical amino acid sequence Ser-**Xaa**-Ser-**Xaa**-Ser (SXSXS series), where **Xaa** denotes the guest amino acid (Scheme 1). Due to limited solubility of some substrates and for control purposes, we also studied end-to-end contact formation in a second host-guest peptide series. Here, the two triplet labels were separated by the canonical amino acid sequence Ser-**Xaa**-Ser (SXS series).

In TTET experiments the triplet donor xanthonic acid is excited to its long-lived triplet state by irradiation with a short laser pulse[17,18]. Upon formation of van-der-Waals contact with the triplet acceptor naphthalene, the excited state energy is irreversibly transferred via a two-electron exchange process[21,22]. Both the formation of the xanthone triplet state and the transfer to naphthalene were shown to occur on the low picosecond timescale[23], thus rendering the TTET reaction diffusion-controlled[24] and allowing the measurement of absolute contact formation rate constants[25] (Krieger and Kiefhaber, in preparation; Krieger Ph. D. thesis[26]).

Due to the intense absorption of the xanthone and naphthalene triplet states, the transfer process can be monitored by time-resolved absorption measurements at 590 and 420 nm, respectively.

Denaturant dependence of end-to-end contact formation

Loop closure kinetics in the host-guest peptides were studied in the host-guest peptides in dependence of denaturant concentration. Fig. 1 shows transient absorption of the xanthone triplet state acquired for a SRSRS peptide at different concentrations of GdmCl. Measurements of the concomitant absorbance increase of the naphthalene triplet state at 420 nm confirmed triplet-triplet energy transfer (not shown). At higher concentrations of the denaturant the lifetime of the xanthone triplet state was clearly prolonged which is indicative of slower loop closure dynamics. Absorption decays at 590 nm were fitted to exponential functions to determine rate constants of end-to-end contact formation k_c.

As shown in Fig. 2, the logarithm of k_c linearly decreased with the concentrations of urea and GdmCl. Slight deviations from linearity were only seen in the GdmCl-dependence for the SGS, SIS and SVS peptides which can be ascribed to experimental problems. In the case of the SGS peptide the measured rate constants are close to the limit of time resolution of our flash photolysis instrument (~ 5 ns). The hydrophobic peptides SIS and SVS were only poorly soluble. Experimental data determined at different denaturant concentrations $[D]$ were fitted to eq. (1) to determine the rate constant of end-to-end contact formation in the absence of denaturant, k_c^0, and the m_c-value which quantifies the sensitivity of the loop closure kinetics towards denaturant concentration.

$$\ln k_c = \ln k_c^0 - m_c \frac{[D]}{RT} \tag{1}$$

All fit results are listed in Table 1. The rate constants of contact formation in the absence of denaturant, k_c^0, obtained for the different peptides are summarized in Fig. 3. In agreement with previous findings[18], the glycine-containing peptides SGS and SGSGS showed by far the fastest end-to-end contact formation reaction on the order of $1.1 \cdot 10^8$ s^{-1}. Within the SXS context almost the same rate constants k_c^0 of about $7.0 \cdot 10^7$ s^{-1} were measured for the guest amino acids aspartic acid, glutamic acid, lysine, arginine and serine. In contrast, peptides containing the hydrophobic residues isoleucine and valine, which are branched at the C$^\beta$-atom, displayed substantially lower end-to-end contact formation rate constants of around $4.6 \cdot 10^7$ s^{-1}. In the SXSXS context the small polar resp. charged amino acids serine, asparagine and aspartic acid led to the fastest contact formation except for glycine (about $6.0 \cdot 10^7$ s^{-1}). For the larger (glutamic acid, lysine, glutamine and arginine) or less polar (alanine) side chains slower contact formation reactions between $3.7 \cdot 10^7$ s^{-1} and $5.3 \cdot 10^7$ s^{-1} were observed.

The m_c-values determined for the different host-guest peptides are shown in Fig. 4. For GdmCl m_c-values between 310 and 450 J·mol^{-1}·M^{-1} were found. The highest values were observed within the SXSXS context for the charged and polar amino acids such as aspartic acid and asparagine. The shorter SXS peptides and especially peptides with nonpolar amino acids showed low m_c-values. A largely similar pattern was also observed for the effect of urea on intramolecular end-to-end contact formation. Again, the highest m_c-values of up to 330 J·mol^{-1}·M^{-1} were found for charged and polar guest amino acids, whereas the nonpolar amino acids displayed lower m_c-values between 210 and 250 J·mol^{-1}·M^{-1}. In contrast to the data obtained for the effect of GdmCl on intramolecular contact formation, the basic amino acids lysine and arginine showed by far the highest m_c-values for urea.

Viscosity dependence of end-to-end contact formation

We have shown previously[16] that the increase of solvent viscosity η with increasing denaturant concentration makes an important contribution to the m_c-values. To account for this, we determined the effect of η on k_c. Glycerol was used as an additive to increase solvent viscosity. As shown by Krieger *et al.* (Krieger and Kiefhaber, in preparation; Krieger Ph.D. thesis[26]), within error the same viscosity dependence of contact formation is obtained when other small viscogenic agents are employed. TTET experiments were performed for all peptides at solvent viscosities of up to 6.3 cP and rate constants of end-to-end contact formation, k_c, were determined (Fig. 5). For all peptides the viscosity dependence of k_c could be described by the empirical relationship given in eq. (2).

$$k_c = k_c^0 \left(\frac{\eta}{\eta_0} \right)^{\beta}$$
(2)

The coefficient β quantifies the response of the contact formation rate constant towards viscosity changes and η_0 denotes the reference viscosity of pure buffer. Results obtained for the different peptides are shown in Fig. 6 and are listed in Table 1. Analogous to the m_c-values for the denaturant effect, the lowest absolute β-values were seen for the short peptides with hydrophobic guest amino acids, e. g. $\beta = -0.67$ in the case of SVS. Considerably higher absolute β-values (up to $\beta = -0.97$) were found for charged and polar amino acids. Interestingly, the basic amino acids lysine and arginine, which displayed the highest m_c-values for the effect of urea on contact formation, also show the highest absolute viscosity coefficients.

Viscosity-corrected chain dynamics

We used the above determined viscosity coefficients β to correct the chain dynamics measured at different denaturant concentrations for the corresponding solvent viscosities according to eq. (3).

$$k_c^{'} = k_c \left(\frac{\eta}{\eta_0} \right)^{-\beta} \qquad\qquad (3)$$

For all peptides the viscosity-corrected end-to-end contact formation rate constants, k_c', showed a hyperbolic decrease with denaturant concentration (not shown). As in our previous studies[16], we analyzed these data in terms of a weak binding interaction of denaturant molecules with the peptide employing the Schellman 'site exchange' model[14,27,28,29]. For this analysis it is favourable to convert denaturant concentrations $[D]$ to the mole fraction scale, X_D (Fig. 7). Viscosity-corrected rate constants were fitted to eq. (4).

$$k_c^{'} = k_c^0 \left(1 - \gamma \frac{(K_{ex} - 1)X_D}{(K_{ex} - 1)X_D + 1} \right) \qquad\qquad (4)$$

K_{ex} signifies a dimensionless equilibrium constant for denaturant binding on the mole fraction scale and γ a factor which quantifies the effect of denaturant binding. The viscosity-corrected end-to-end contact formation rate constants for all peptides and both denaturants could be well described by binding isotherms according to eq. (4). As stated above, the data obtained for the peptides SGS, SIS and SVS had larger errors due to experimental problems. Consequently, small deviations of the fitted binding isotherms from the experimental data were observed for these peptides.

Results from the data analysis are summarized in Table 2. Taking into account the parameter errors, very similar binding constants are determined for all peptides although they contain a variety of amino acids with very different properties. For urea K_{ex} varies between 6.8 and 16.0

and for GdmCl values between 9.6 and 21.5 are found. With the exception of the relatively poor data for the SVS peptide, both for urea and GdmCl the binding constants observed for the different peptides vary by less than a factor of two. Less clear results are obtained for the parameters γ (Table 2). Values between 0.42 and 0.68 are observed for urea, and between 0.48 and 0.65 for GdmCl. Consistent with our previous findings[16], on average similar γ-values are obtained for both denaturants which indicates that urea and GdmCl mainly differ in their relative binding strength but not in their effect on chain dynamics when bound to peptides. The basic amino acids lysine and arginine, which showed unusually weak GdmCl dependencies of k_c, have relatively low γ^{GdmCl}-values in comparison to the γ^{urea}-values as well. Due to correlation of the parameters γ and K_{ex} within the fitting procedure, they could only be determined with limited accuracy.

Discussion

Comparison of the effects of urea and GdmCl on peptide chain dynamics

Our results on the effects of urea and GdmCl on intrachain loop formation in host-guest peptides showed that the relative order of the m_c-values found for the different peptides was largely the same for both denaturants. The only exception were the peptides containing the basic amino acids lysine and arginine, which showed exceptionally large m_c-values for urea. This becomes also evident from a correlation plot of the m_c-values determined for the two denaturants (Fig. 8). All peptides except the basic ones (SKS, SRS, SKSKS and SRSRS) show a linear correlation of the m_c-values for the two denaturants with a correlation coefficient of 0.98. This correlation of m_c^{GdmCl} versus m_c^{urea} yields a slope of 1.82 ± 0.33 and an intercept of virtually zero $((-75.3 \pm 87.6)$ J·mol^{-1}·M^{-1}). It is interesting to note that these

results are similar to those reported by Myers *et al.*[10] for the linear correlation between the

m_{eq}-values for protein unfolding by urea and GdmCl ($m_{eq} = \partial\Delta G^0 / \partial[D]$). To facilitate better

comparison to our results, we reevaluated these data and found a slope of 2.18 ± 0.11 and an

intercept of (10.5 ± 94.8) J·mol^{-1}·M^{-1} for the correlation of m_{eq}^{GdmCl} versus m_{eq}^{urea}. These

findings imply that the effects of denaturants on intrachain contact formation and on protein

stability are due to similar molecular mechanisms. We further examined whether the deviating

denaturant dependence of contact formation observed for the basic amino acid residues is also

reflected in the protein stability data reported by Myers *et al.*[10] We compared the fraction of

basic amino acids in the different proteins to the corresponding ratios $m_{eq}^{GdmCl} / m_{eq}^{urea}$ and

found that these quantities were uncorrelated. There are several reasons why a possibly

different behaviour of basic amino acids with respect to the denaturant effect on protein

stability might not be detectable. First, proteins are composed of a heterogeneous mixture of

different amino acids and accordingly the fraction of basic amino acid side chains is low and

relatively invariant between several proteins. Secondly, protein stability data report on the

conformational transition between the native and unfolded state of proteins. Specific

interactions between protein residues in either state will influence the observed denaturant

dependence and might mask a possible difference of the basic amino acids.

The low ratio of m_c^{GdmCl} over m_c^{urea} seen for peptides containing lysine or arginine raises the

question whether the interaction of these amino acids with urea is particularly strong or

whether the interaction with GdmCl is comparatively weak. An answer to that problem is

afforded by the viscosity dependence observed for intrachain contact formation. We have

shown before[16] that part of the denaturant effect on loop formation kinetics is due to increased

solvent viscosity. Accordingly, the denaturant and viscosity coefficients, β and m_c, determined

for loop formation in the host-guest peptides are linearly correlated as shown in Fig. 9.

However, the peptides containing basic amino acids significantly deviate from the linear

correlation between the m_c-value for GdmCl and β (Fig. 9B). These data show that the m_c^{GdmCl}-values for basic peptides are unusually low for a given viscosity coefficient β. Thus, in comparison to other residues, the interaction of lysine and arginine with GdmCl is remarkably weak, while the interaction with urea corresponds to that of the other amino acids.

Main interaction site for denaturants

Although the studied peptides contained amino acid side chains with very different properties, they all displayed remarkably similar effects of urea and GdmCl on end-to-end contact formation. With the exception of the peptides containing lysine or arginine, all peptides showed the same correlation of the two denaturant m_c-values with each other and with the viscosity coefficient β (Figs. 8 and 9). Differences in the m_c-values between the host-guest peptides are mainly due to differences in their viscosity coefficients β. Furthermore, the constants obtained for denaturant binding were similar for all peptides. In addition, we were able to describe the dependence of the viscosity-corrected end-to-end contact formation rate constants, k_c', on the mole fraction of denaturant, X_D, by a weak binding interaction with the same global binding constants for all peptides. Corresponding fits of the experimental data yielded low reduced χ^2-values of 0.19 and 0.52 for the urea and GdmCl dependence, respectively. This indicates that all data can be well described with global binding constants which amounted to 9.8 ± 1.3 and 16.9 ± 1.4 for urea and GdmCl, respectively. Global analysis of the viscosity-corrected chain dynamics also allowed to determine the γ-values with improved accuracy (Table 2 and Fig. 10). Except for the basic peptides the γ-values adopted similar values for the effect of urea and GdmCl. In the case of the peptides containing lysine or arginine γ^{GdmCl}-values were lower than the γ^{urea}-values. These findings suggest a common mechanism for the effect of the two denaturants on the chain dynamics of all the peptides.

Thus, denaturants are most likely interacting with the peptide backbone which was invariant within the series of the different substrates in study. The somewhat different behaviour of the basic amino acids might be due to repulsive interactions of the like charged side chain groups and the guanidinium cation.

Recently, related findings were reported by Whittington *et al.*[30] who studied the conformational properties of different unstructured model peptides at varying denaturant concentrations by circular dichroism (CD). They found a hyperbolic increase of a CD signal around 220 nm with denaturant concentration. This signal arises from the absorption of the peptide group and is indicative of polyproline II structure formation[31,32,33]. As clearly the conformation of the backbone undergoes changes in these experiments, it is reasonable to assume that the denaturant molecules interact with the peptide backbone. This is also underlined by the fact that in the study of Whittington *et al.* similar observations were made for several peptides of different amino acid sequence.

Implications for protein folding

As discussed above, we found a ratio of 1.82 ± 0.33 between the m_c-values for the effects of GdmCl and urea on peptide loop closure dynamics. This value is close to the ratio of $m_{eq}^{GdmCl} / m_{eq}^{urea} = 2.18 \pm 0.11$ previously observed for the effects of these two denaturants on protein stability[10]. These data suggest that the effects of denaturants on protein stability, folding and unfolding rate constants follow a mechanism closely related to the one presently observed and characterized for peptide chain dynamics. In particular, the peptide backbone seems to be the main interaction site for denaturants. Molecular dynamics simulations of unstructured peptides in urea solutions[34] and of urea-induced protein unfolding[35] also implicate the protein backbone as the main interaction site for denaturants. Furthermore,

Bachmann *et al.*[36] recently studied the folding of natural collagen fragments. m_{eq}-values obtained for the effect of GdmCl on protein stability were in good agreement with m_{eq}-values expected for globular proteins of corresponding size[10]. In contrast to globular proteins, in collagen the main fraction of solvent-accessible surface area which is buried upon folding derives from the peptide backbone, while the amino acid sidechains remain solvated. The agreement with the m_{eq}-values expected for globular proteins implies that the backbone is the main interaction site for denaturants.

It remains to be studied whether the slightly deviating response of intramolecular contact formation towards GdmCl concentration observed for the basic peptides has consequences for protein folding and stability as well. As pointed out above, corresponding studies are hampered by the intrinsic complexity of even the simplest protein folding reactions. Further, it would be interesting to examine to what extent the denaturant dependencies of protein stability, folding and unfolding reactions contain contributions from solvent viscosity. As denaturant-induced unfolding of proteins is the method of choice to study protein folding reactions, this might have far-reaching consequences for the interpretation of data on protein stability and folding kinetics.

Acknowledgements

Tables

Table 1 – End-to-end contact formation rate constants and denaturant and viscosity dependencies of host-guest peptides.

Peptide[a]	Symbol[b]	k_c^0 (s^{-1})[c]	m_c^{urea} (J mol^{-1} M^{-1})[d]	m_c^{GdmCl} (J mol^{-1} M^{-1})[d]	β[e]	m_c^{GdmCl} / m_c^{urea}
SDS	□	$(4.7 \pm 0.4) \cdot 10^7$	268.6 ± 11.3	406.8 ± 13.2	-0.80 ± 0.03	1.51 ± 0.08
SES	□	$(7.3 \pm 0.4) \cdot 10^7$	256.4 ± 11.2	389.8 ± 13.1	-0.78 ± 0.02	1.52 ± 0.08
SGS	◇	$(1.14 \pm 0.1) \cdot 10^8$	218.4 ± 13.0	322.6 ± 15.3	-0.73 ± 0.03	1.48 ± 0.11
SIS	▽	$(4.5 \pm 0.3) \cdot 10^7$	230.8 ± 11.3	346.8 ± 13.1	-0.76 ± 0.02	1.50 ± 0.09
SKS	○	$(6.7 \pm 0.5) \cdot 10^7$	309.0 ± 12.7	398.0 ± 15.0	-0.97 ± 0.03	1.29 ± 0.07
SRS	○	$(6.8 \pm 0.5) \cdot 10^7$	301.9 ± 13.1	405.0 ± 15.7	-0.93 ± 0.03	1.34 ± 0.08
SSS	◇	$(7.2 \pm 0.5) \cdot 10^7$	282.3 ± 11.6	420.0 ± 13.3	-0.81 ± 0.02	1.49 ± 0.08
SVS	▽	$(4.7 \pm 0.3) \cdot 10^7$	218.0 ± 12.8	316.8 ± 15.0	-0.67 ± 0.03	1.45 ± 0.11
SASAS	▼	$(5.3 \pm 0.3) \cdot 10^7$	247.6 ± 12.1	382.6 ± 14.6	-0.80 ± 0.03	1.55 ± 0.10

SDSDS	■	$(6.3 \pm 0.4) \cdot 10^7$	276.2 ± 11.5	446.0 ± 13.4	-0.85 ± 0.02	1.62 ± 0.08
SESES	■	$(4.7 \pm 0.3) \cdot 10^7$	279.8 ± 11.6	423.0 ± 13.5	-0.79 ± 0.02	1.51 ± 0.08
SGSGS	◆	$(1.09 \pm 0.1) \cdot 10^8$	281.2 ± 12.9	436.2 ± 15.6	-0.87 ± 0.03	1.55 ± 0.09
SKSKS	●	$(3.8 \pm 0.2) \cdot 10^7$	332.7 ± 11.5	433.5 ± 13.2	-0.90 ± 0.03	1.30 ± 0.06
SNSNS	▲	$(5.9 \pm 0.4) \cdot 10^7$	283.2 ± 11.3	439.3 ± 13.3	-0.81 ± 0.02	1.55 ± 0.08
SQSQS	▲	$(3.7 \pm 0.3) \cdot 10^7$	258.2 ± 13.0	397.5 ± 16.6	-0.79 ± 0.03	1.54 ± 0.10
SRSRS	●	$(3.9 \pm 0.2) \cdot 10^7$	326.5 ± 11.2	419.9 ± 13.2	-0.95 ± 0.03	1.29 ± 0.06
SSSSS	◆	$(6.0 \pm 0.4) \cdot 10^7$	266.3 ± 12.4	414.5 ± 13.9	-0.86 ± 0.03	1.56 ± 0.09

[a]: Amino acid sequence of the peptides between the TTET labels in one-letter code (see Scheme 1).

[b]: Symbols used to denote data for the different peptides in Figs. 2, 5, 7, 8, 9, 10.

[c]: End-to-end contact formation rate constants in pure buffer determined from a global fit of the denaturant and viscosity dependencies according to eqs. (1) and (2).

[d]: Determined by fitting the denaturant dependence of the rate constant of end-to-end contact formation according to eq. (1).

[e]: Determined by fitting the corresponding viscosity dependence of intramolecular contact formation (eq. (2)).

Table 2 – Analysis of the viscosity-corrected end-to-end contact formation rate constants in terms of denaturant binding.

Peptide[a]	individual analysis[b]				global analysis[c]	
	K_{ex}^{urea}	γ^{urea}	K_{ex}^{GdmCl}	γ^{GdmCl}	γ^{urea}	γ^{GdmCl}
SDS	8.3 ± 4.3	0.62 ± 0.14	14.3 ± 4.5	0.64 ± 0.06	0.56 ± 0.06	0.61 ± 0.03
SES	7.8 ± 4.3	0.62 ± 0.15	16.8 ± 5.6	0.59 ± 0.05	0.56 ± 0.06	0.59 ± 0.04
SGS	13.8 ± 8.8	0.42 ± 0.09	11.2 ± 7.5	0.48 ± 0.11	0.45 ± 0.06	0.44 ± 0.04
SIS	10.5 ± 7.3	0.44 ± 0.11	15.3 ± 6.6	0.50 ± 0.06	0.45 ± 0.06	0.50 ± 0.04
SKS	9.9 ± 3.9	0.63 ± 0.11	15.6 ± 4.8	0.50 ± 0.05	0.63 ± 0.06	0.52 ± 0.04
SRS	12.5 ± 4.9	0.58 ± 0.08	21.5 ± 7.1	0.53 ± 0.05	0.62 ± 0.06	0.54 ± 0.04
SSS	8.1 ± 3.9	0.68 ± 0.15	18.5 ± 5.6	0.61 ± 0.04	0.62 ± 0.06	0.62 ± 0.03
SVS	6.8 ± 4.4	0.61 ± 0.19	9.6 ± 4.5	0.56 ± 0.10	0.53 ± 0.06	0.48 ± 0.04
SASAS	11.5 ± 6.8	0.48 ± 0.10	18.8 ± 6.6	0.55 ± 0.05	0.48 ± 0.06	0.56 ± 0.04
SDSDS	11.4 ± 5.5	0.57 ± 0.10	20.6 ± 5.6	0.65 ± 0.04	0.58 ± 0.06	0.66 ± 0.03
SESES	13.7 ± 6.1	0.56 ± 0.08	20.4 ± 5.9	0.62 ± 0.04	0.59 ± 0.06	0.64 ± 0.03
SGSGS	10.6 ± 4.8	0.60 ± 0.10	16.2 ± 5.0	0.63 ± 0.05	0.60 ± 0.06	0.62 ± 0.03
SKSKS	14.0 ± 4.9	0.65 ± 0.07	15.6 ± 4.8	0.65 ± 0.05	0.71 ± 0.05	0.64 ± 0.03
SNSNS	8.6 ± 3.9	0.68 ± 0.13	19.7 ± 5.4	0.65 ± 0.04	0.64 ± 0.06	0.67 ± 0.03
SQSQS	11.1 ± 6.3	0.52 ± 0.10	10.7 ± 5.9	0.63 ± 0.13	0.54 ± 0.06	0.57 ± 0.04
SRSRS	10.5 ± 4.4	0.66 ± 0.10	17.2 ± 6.1	0.56 ± 0.05	0.66 ± 0.05	0.56 ± 0.04
SSSSS	16.0 ± 9.0	0.45 ± 0.07	16.9 ± 6.2	0.60 ± 0.06	0.51 ± 0.06	0.60 ± 0.04

[a]: Amino acid sequence of the peptides in one-letter code (Scheme 1).

[b]: Results from fitting the viscosity-corrected end-to-end contact formation rate constants to a Schellman binding model (eq. (4)). K_{ex} denotes the binding constant for denaturants on the mole fraction scale and γ is a factor which quantifies the effect of denaturant binding.

[c]: Corresponding results from a global fit where global constants for binding of urea and GdmCl common to all peptides were assumed. Global binding constants amounted to 9.8 ± 1.3 and 16.9 ± 1.4 for urea and GdmCl, respectively.

Schemes

Scheme 1 – Covalent structure of the host-guest peptide substrates.

(A)
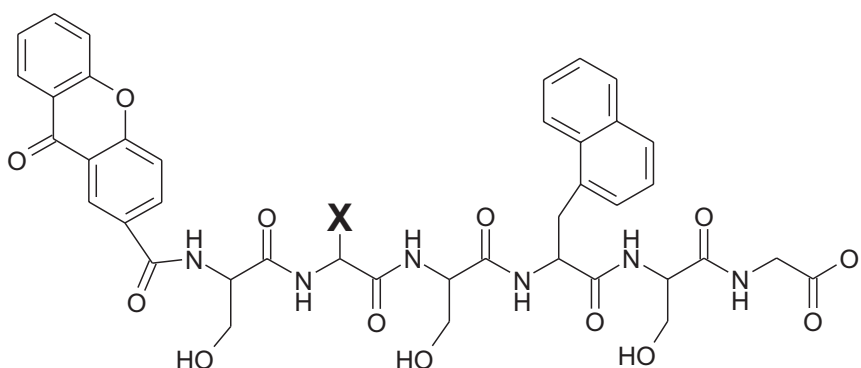
(B)
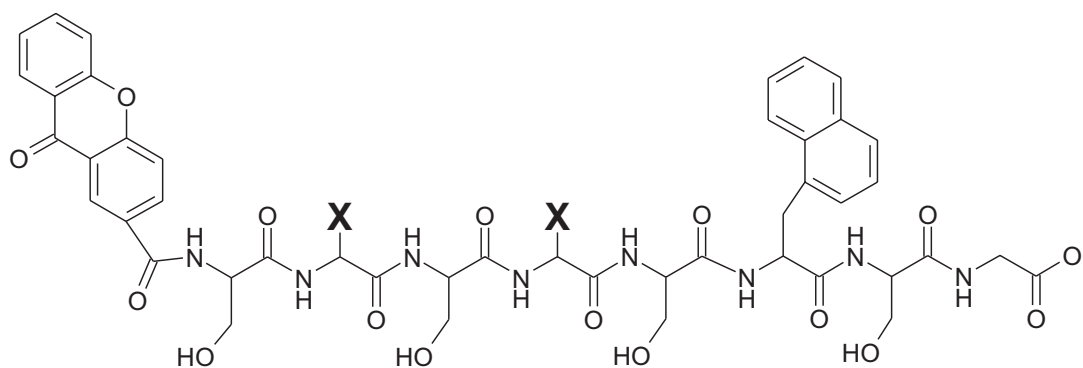
The letters '**X**' denote the guest amino acids. Peptides of series (A) and (B) are referred to as SXS and SXSXS peptides, respectively.

References

1. Spiro, K. (1900). Über die Beeinflußung der Eiweißcoagulation durch stickstoffhaltige Substanzen. *Z. Physiol. Chem.* **30**, 182-199.

2. Ramsden, W. (1902). Some new properties of urea. *J. Physiol.* **28**, 23-27.

3. Greenstein, J. P. (1938). Sulfhydryl groups in proteins. *J. Biol. Chem.* **125**, 501-513.

4. Tanford, C. (1968). Protein denaturation. *Adv. Prot. Chem.* **23**, 121-282.

5. Tanford, C. (1970). Protein denaturation. Part C. Theoretical models for the mechanism of denaturation. *Adv. Prot. Chem.* **24**, 1-95.

6. Pace, C. N. & Tanford, C. (1968). Thermodynamics of the unfolding of β-lactoglobulin A in aqueous urea solutions between 5 and 55°. *Biochemistry* **7**, 198-208.

7. Alexander, S. S. & Pace, C. N. (1971). A comparison of the denaturantion of bovine-lactoglobulins A and B and goat-lactoglobulin. *Biochemistry* **10**, 2738-2743.

8. Greene, R. F. J. & Pace, C. N. (1974). Urea and guanidine-hydrochloride denaturation of ribonuclease, lysozyme, alpha-chymotrypsin and beta-lactoglobulin. *J. Biol. Chem.* **249**, 5388-5393.

9. 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**, 8063-8068.

10. 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.

11. Pace, C. N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Meth. Enzymol.* **131**, 266-280.

12. Tanford, C. (1964). Isothermal unfolding of globular proteins in aqueous urea solutions. *J. Am. Chem. Soc.* **86**, 2050-2059.

13. Aune, K. C. & Tanford, C. (1969). Thermodynamics of the denaturation of lysozyme by guanidine hydrochloride. II. Dependence on denaturant concentration at 25 degrees. *Biochemistry* **11**, 4586-4590.

14. Schellman, J. A. (2002). Fifty years of solvent denaturation. *Biophys. Chem.* **96**, 91-101.

15.   Timasheff, S. N. (2002). Thermodynamic binding and site occupancy in the light of the Schellman exchange concept. *Biophys. Chem.* **101-102**, 99-111.

16.   Möglich, A., Krieger, F. & Kiefhaber, T. (2005). Molecular Basis for the Effect of Urea and Guanidinium Chloride on the Dynamics of Unfolded Polypeptide Chains. *J. Mol. Biol.* **345**, 153-162.

17.   Bieri, O., Wirz, J., Hellrung, B., Schutkowski, M., Drewello, M. & Kiefhaber, T. (1999). The speed limit for protein folding measured by triplet-triplet energy transfer. *Proc. Natl. Acad. Sci. USA* **96**, 9597-9601.

18.   Krieger, F., Fierz, B., Bieri, O., Drewello, M. & Kiefhaber, T. (2003). Dynamics of unfolded polypeptide chains as model for the earliest steps in protein folding. *J. Mol. Biol.* **332**, 265-274.

19.   Pace, C. N., Grimsley, G. R. & Scholtz, J. M. (2005). Denaturation of Proteins by Urea and Guanidine Hydrochloride. In *Protein Folding Handbook* (Buchner, J. & Kiefhaber, T., eds.), Vol. 1, pp. 45-69. 5 vols. Wiley-VCH, Weinheim, Germany.

20.   Perl, D., Jacob, M., Bano, M., Stupak, M., Antalik, M. & Schmid, F. X. (2002). Thermodynamics of a diffusional protein folding reaction. *Biophys. Chem.* **96**, 173-190.

21.   Dexter, D. L. (1953). A Theory of Sensitized Luminescence in Solids. *J. Chem. Phys.* **21**, 836-850.

22.   Turro, N. J. (1991). *Modern Molecular Photochemistry*, University Science Books, Sausalito, CA.

23.   Satzger, H., Schmidt, B., Root, C., Zinth, W., Fierz, B., Krieger, F., Kiefhaber, T. & Gilch, P. (2004). Ultrafast Quenching of the Xanthone Triplet by Energy Transfer: New Insight into the Intersystem Crossing Kinetics. *J. Phys. Chem. A* **108**, 10072-10079.

24.   Krieger, F., Fierz, B., Axthelm, F., Joder, K., Meyer, D. & Kiefhaber, T. (2004). Intrachain diffusion in a protein loop fragment from carp parvalbumin. *Chem. Phys.* **307**, 209-215.

25.   Fierz, B. & Kiefhaber, T. (2005). Dynamics of Unfolded Polypeptide Chains. In *Protein Folding Handbook* (Buchner, J. & Kiefhaber, T., eds.), Vol. 2, pp. 809-855. 5 vols. Wiley-VCH, Weinheim, Germany.

26.   Krieger, F. (2004). Dynamics in Unfolded Polypeptide Chains as Model for Elementary Steps in Protein Folding, Biozentrum, University of Basel.

27.   Schellman, J. A. (1975). Macromolecular binding. *Biopolymers* **14**, 999-1018.

28.   Schellman, J. A. (1978). Solvent denaturation. *Biopolymers* **17**, 1305-1322.

29.   Schellman, J. A. (1987). Selective binding and solvent denaturation. *Biopolymers* **26**, 549-559.

30.   Whittington, S. J., Chellgren, B. W., Hermann, V. M. & Creamer, T. P. (2005). Urea promotes polyproline II helix formation: implications for protein denatured states. *Biochemistry* **44**, 6269-6275.

31.   Tiffany, M. L. & Krimm, S. (1968). New chain conformations of poly(glutamic acid) and polylysine. *Biopolymers* **6**, 1379-1382.

32.   Tiffany, M. L. & Krimm, S. (1968). Circular dichroism of poly-L-proline in an unordered conformation. *Biopolymers* **6**, 1767-1770.

33.   Shi, Z., Olson, C. A., Rose, G. D., Baldwin, R. L. & Kallenbach, N. R. (2002). Polyproline II structure in a sequence of seven alanine residues. *Proc. Natl. Acad. Sci. USA* **99**, 9190-9195.

34.   Tobi, D., Elber, R. & Thirumalai, D. (2003). The dominant interaction between peptide and urea is electrostatic in nature: a molecular dynamics simulation study. *Biopolymers* **68**, 359-369.

35.   Bennion, B. J. & Daggett, V. (2003). The molecular basis for the chemical denaturation of proteins by urea. *Proc. Natl. Acad. Sci. USA* **100**, 5142-5147.

36.   Bachmann, A., Kiefhaber, T., Boudko, S., Engel, J. & Bächinger, H. P. (2005). Collagen triple-helix formation in all-trans chains proceeds by a nucleation/growth mechanism with a purely entropic barrier. *Proc. Natl. Acad. Sci. USA* **102**, 13897-13902.

Figure Captions

Fig. 1 – Time-resolved xanthone absorption at 590 nm in a SRSRS peptide as a function of GdmCl concentration (from left to right 1, 2, 3, 4, 5, 6, 7, 8 M GdmCl). Data have been normalized.

Fig. 2 – Denaturant dependence of end-to-end contact formation in host-guest peptides. Rate constants of contact formation, k_c, were determined at different concentrations of urea (left column) and GdmCl (right). The top row shows data obtained for the SXS peptides and the bottom one for the SXSXS peptides. The solid lines represent fits of the experimental data to eq. (1). Open symbols denote the peptides from the SXS series as follows: SDS (□), SES (□), SGS (◇), SIS (▽), SKS (○), SRS (○), SSS (◇) and SVS (▽). Closed symbols represent the SXSXS peptides: SASAS (▼), SDSDS (■), SESES (■), SGSGS (◆), SKSKS (●), SNSNS (▲), SQSQS (▲), SRSRS (●) and SSSSS (◆).

Fig. 3 – Rate constants of end-to-end contact formation in buffer, k_c^0, for SXS (top) and SXSXS (bottom) peptides. Values of k_c^0 were determined from global fits of denaturant and viscosity dependencies of contact formation according to eqs. (1) and (2) and are listed in Table 1.

Fig. 4 – m_c-values for the effect of urea (left) and GdmCl (right) on end-to-end contact formation in SXS (top) and SXSXS (bottom) peptides. m_c-values were determined by fitting the data in Fig. 2 to eq. (1) and are listed in Table 1.

Fig. 5 – Viscosity dependence of end-to-end contact formation in SXS (top) and SXSXS (bottom) peptides. Solid lines are fits of the experimental data to eq. (2). Symbols are as in Fig. 2 and are also given in Table 1.

Fig. 6 – Viscosity coefficients β quantifying the sensitivity of the intramolecular contact formation reaction towards solvent viscosity. In the top panel values determined for the SXS peptides and in the lower one data for the SXSXS peptides are shown. β-values were determined by fitting the experimental data shown in Fig. 5 to eq. (2) and are given in Table 1.

Fig. 7 – Viscosity-corrected rate constants of end-to-end contact formation, k_c', for SXS (top) and SXSXS (bottom) peptides as a function of the mole fraction of urea, X_{urea} (left), and GdmCl, X_{GdmCl} (right). Colored symbols denote the different peptides and are explained in Fig. 2 and Table 1. The solid lines are fits of the data to a weak binding isotherm according to eq. (4).

Fig. 8 – Correlation of the denaturant coefficients m_c observed for the effect of urea and GdmCl on intramolecular contact formation in the host-guest peptides. Colored symbols correspond to the different peptides as indicated in Fig. 2 and Table 1. The solid line represents a linear correlation fit of all data points except for the peptides SKS, SRS, SKSKS and SRSRS, which contain basic amino acids. The correlation has a slope of 1.82 ± 0.33 and an intercept of (-75.3 ± 87.6) J·mol^{-1}·M^{-1}. The correlation coefficient amounts to 0.98.

Fig. 9 – Correlation of the viscosity and denaturant coefficients for end-to-end contact formation, β and m_c. Colored symbols indicate the different peptides as shown in Fig. 2 and Table 1. (A) The m_c-values for the effect of urea are clearly correlated to the viscosity

coefficients β. The solid line shows a correlation fit yielding a correlation coefficient of 0.89.

(B) In the corresponding correlation plot for the m_c-values of GdmCl systematic deviations

for the peptides containing basic amino acids are observed (SKS, SRS, SKSKS and SRSRS).

The other data are well correlated with a coefficient of 0.88 (solid line).

Fig. 10 – Correlation of γ-values for the effect of urea and GdmCl binding on peptide chain

dynamics. Colored shapes symbolize the different peptides as before (cp. Fig. 2 and Table 1).
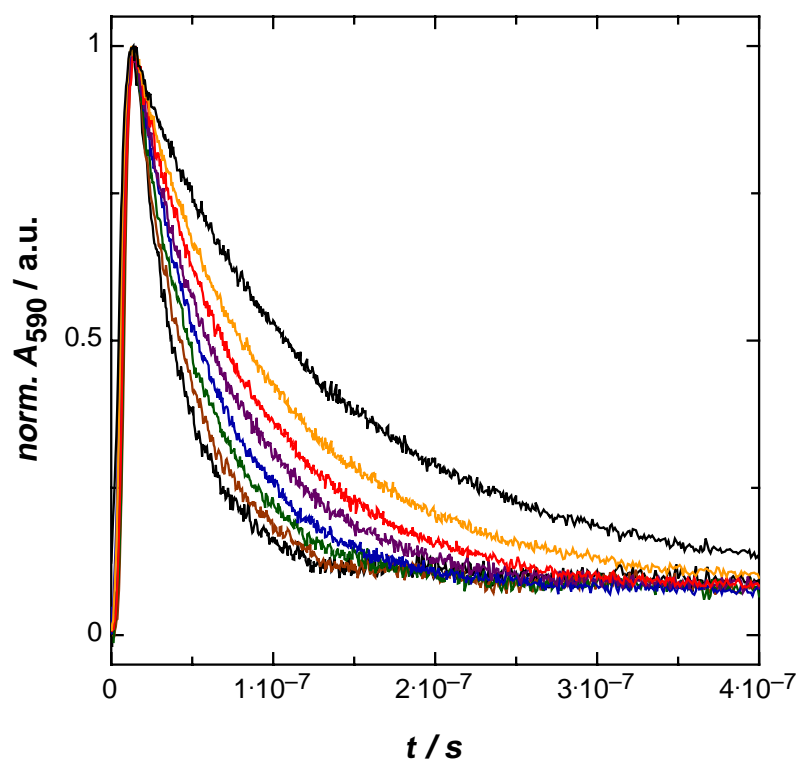
Fig. 1
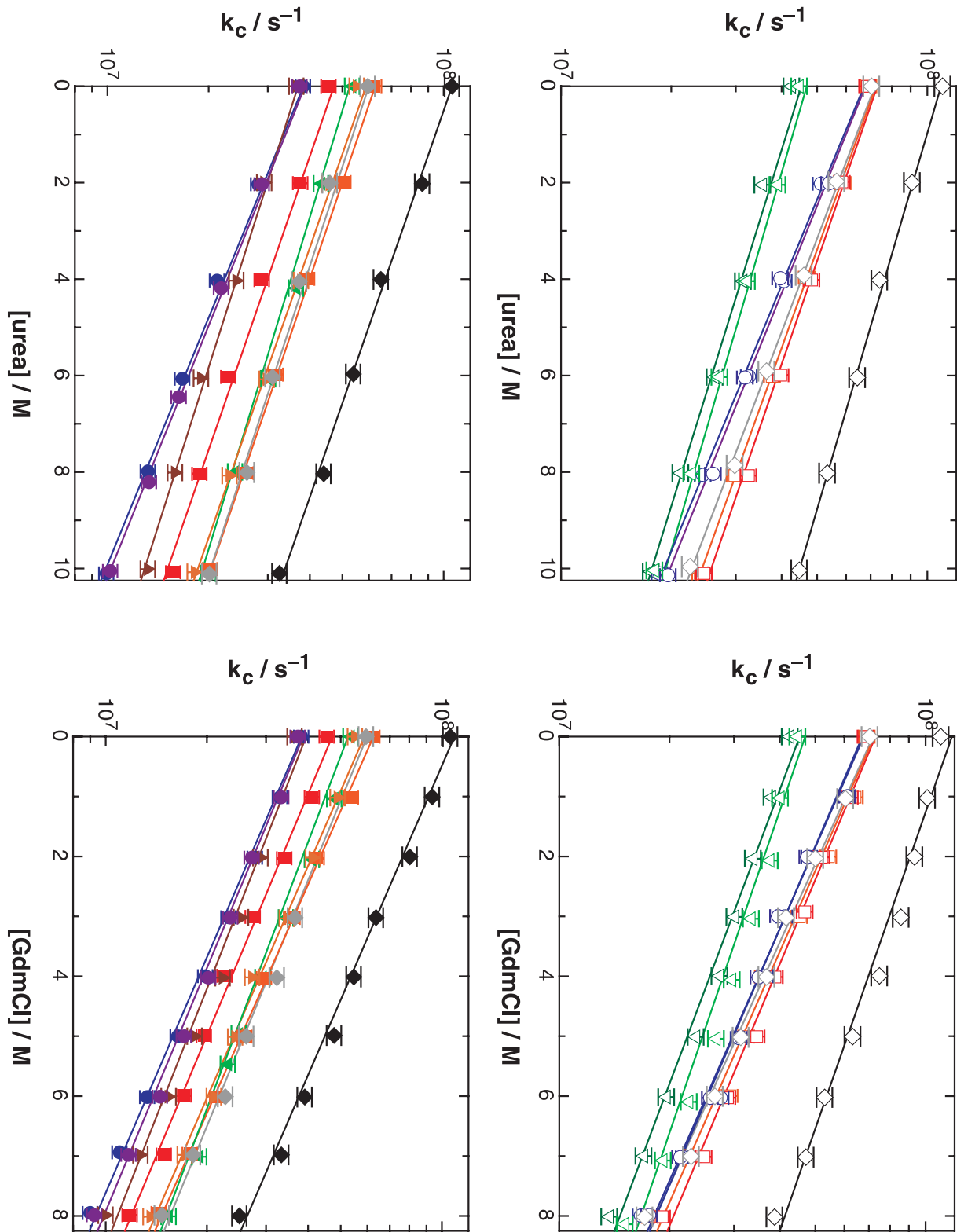
Fig. 2
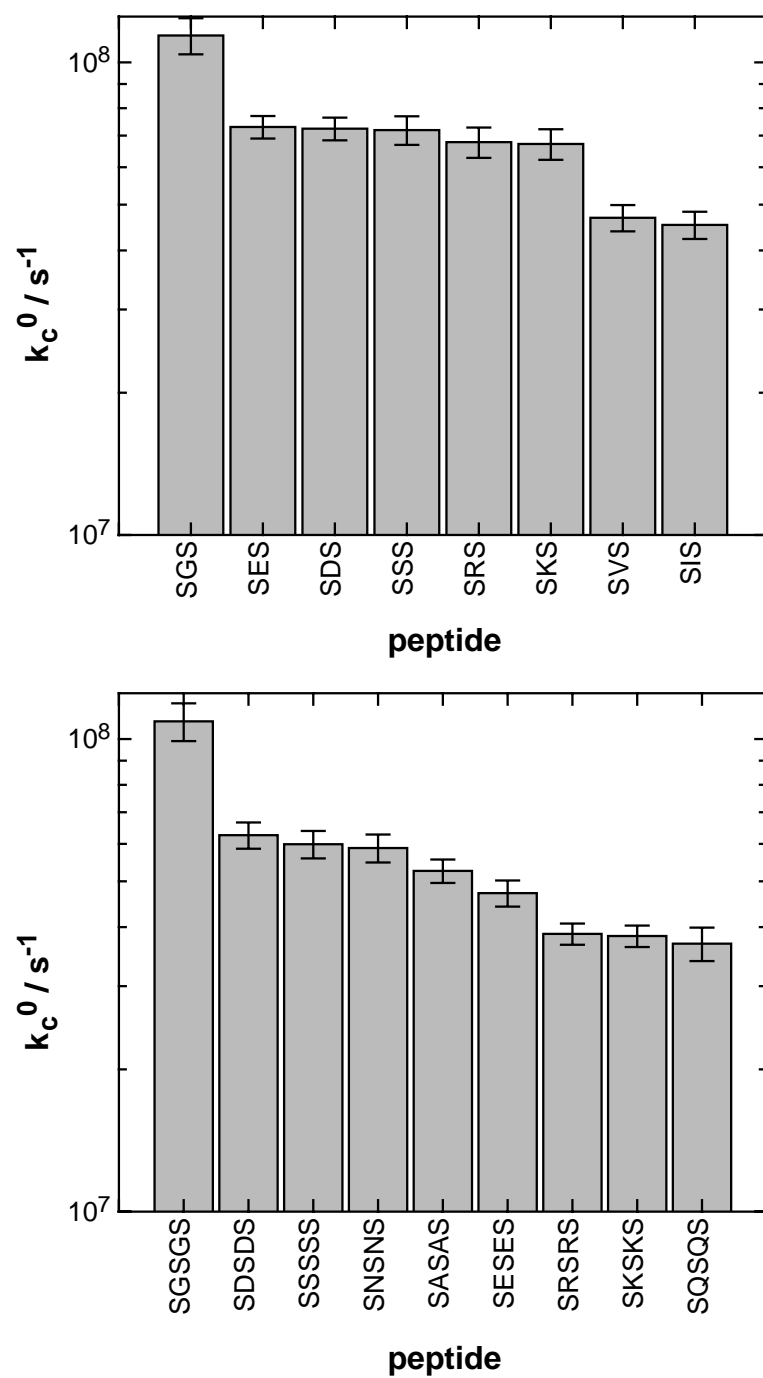
Fig. 3
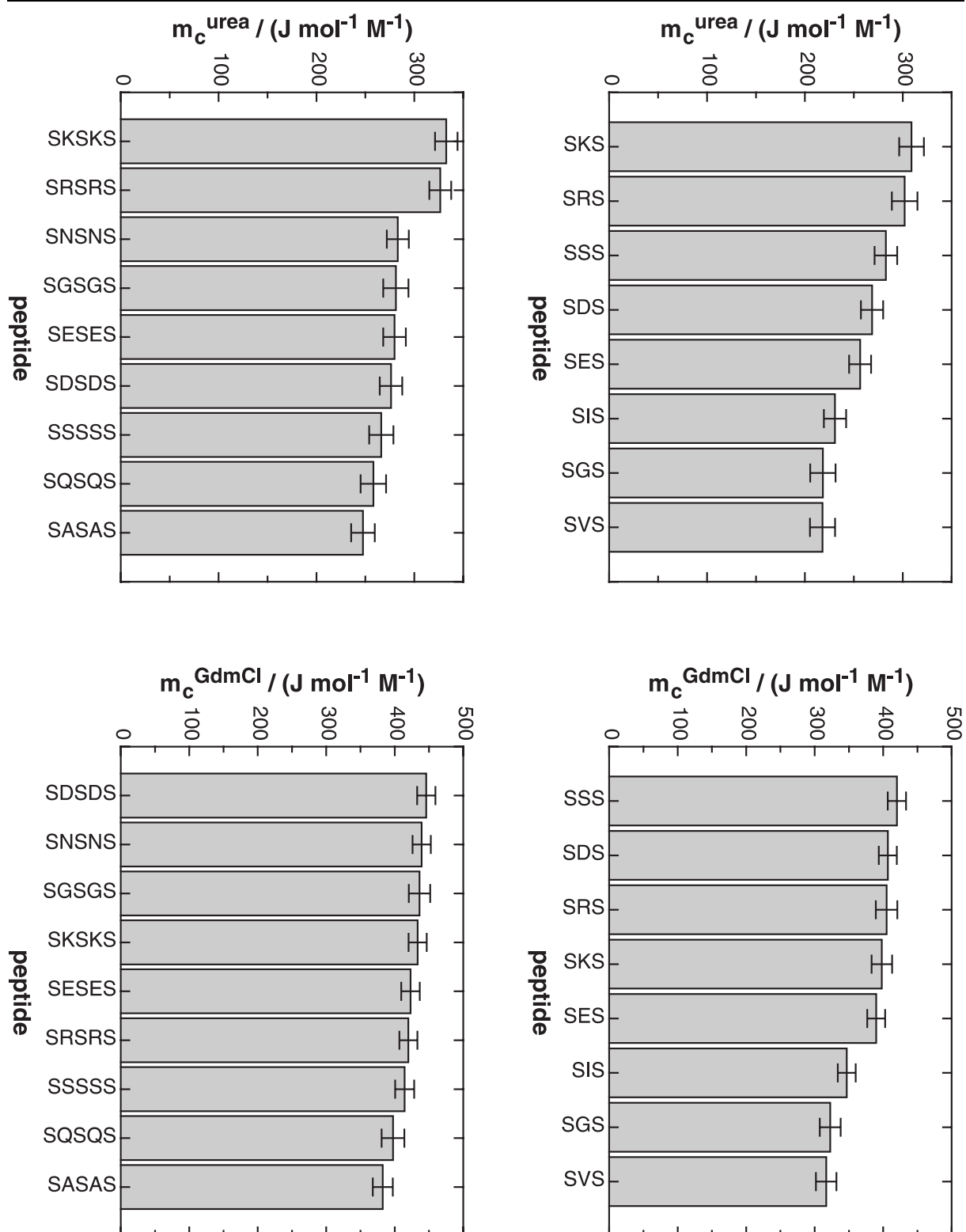
Fig. 4
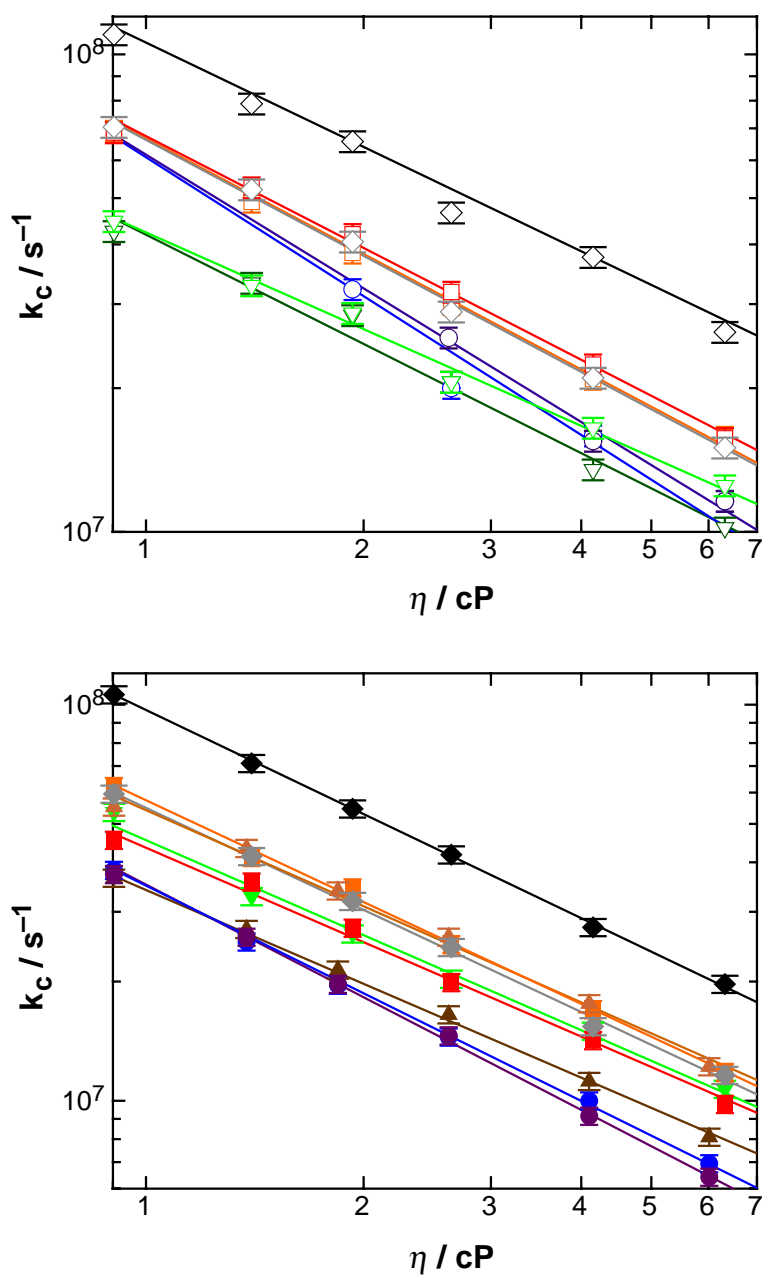
Fig. 5
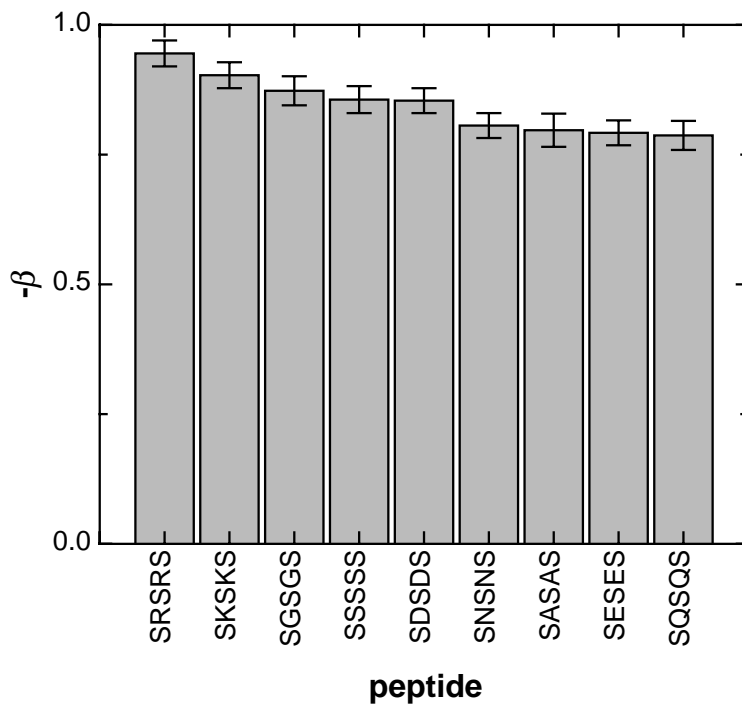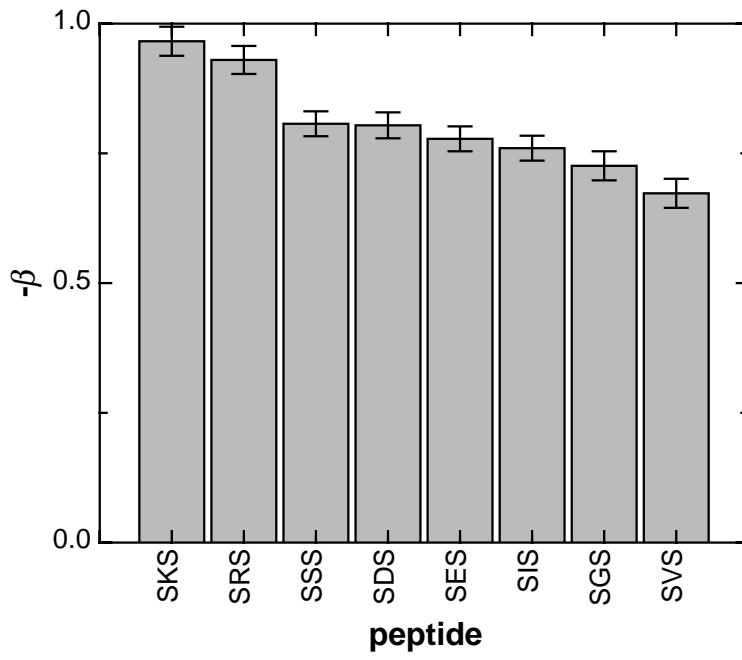
Fig. 6
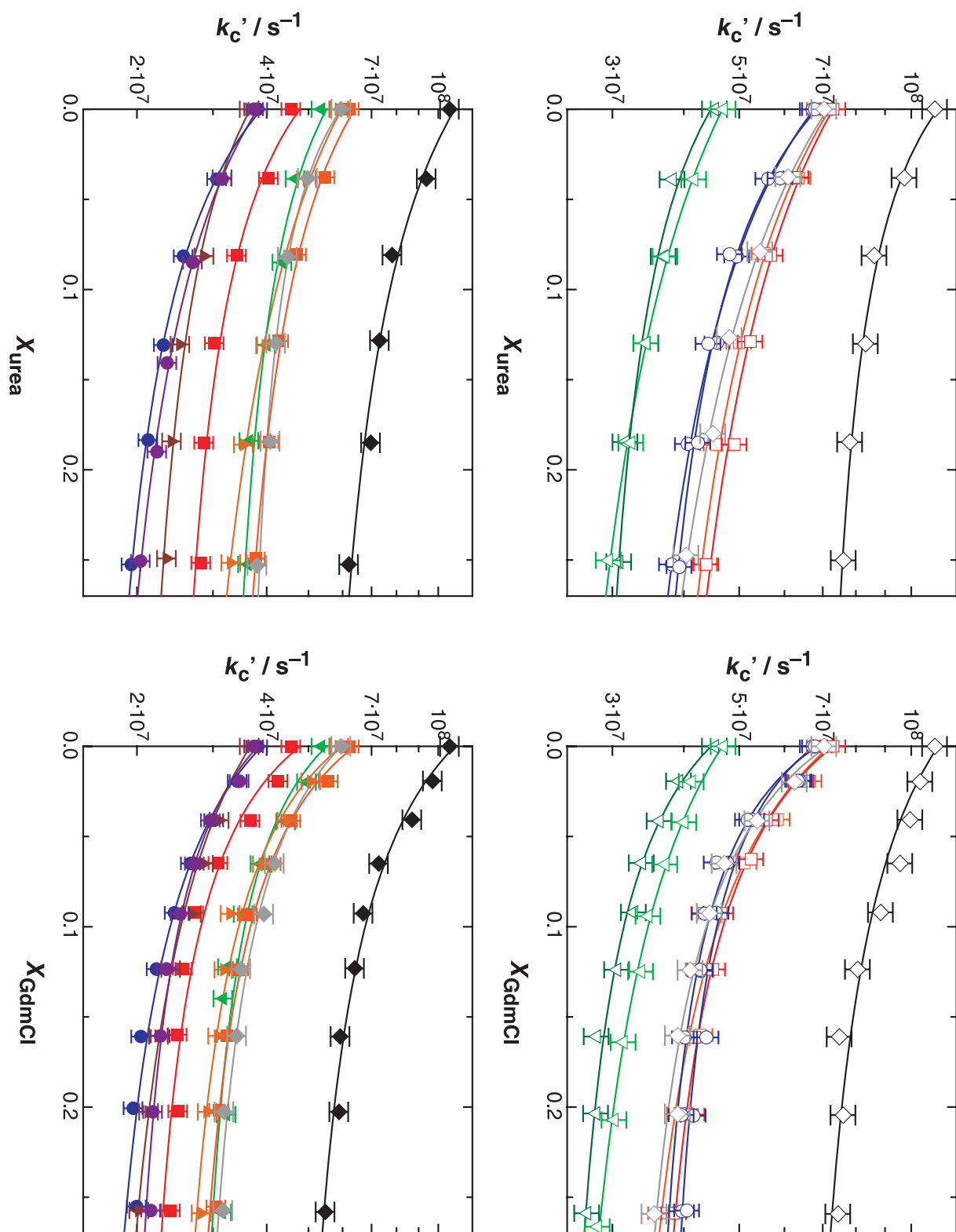
Fig. 7
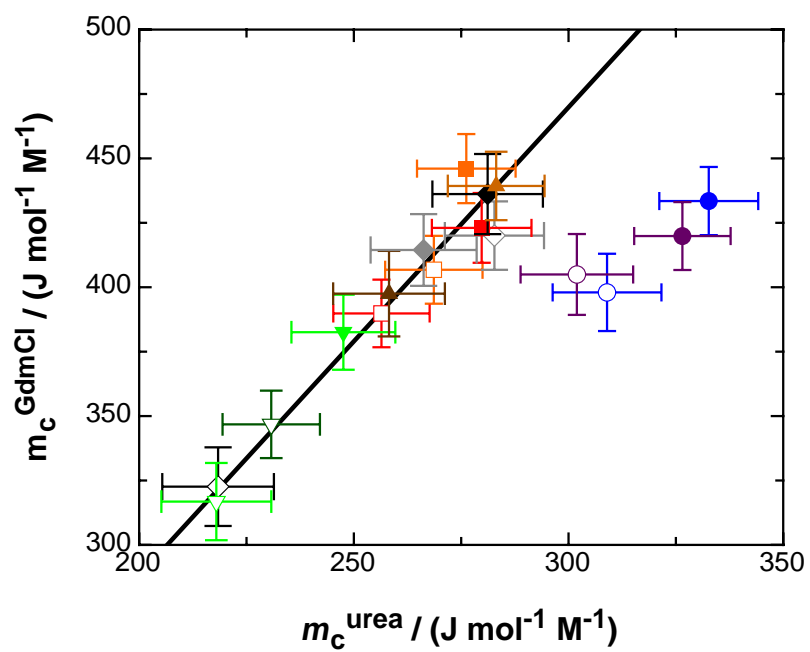
Fig. 8
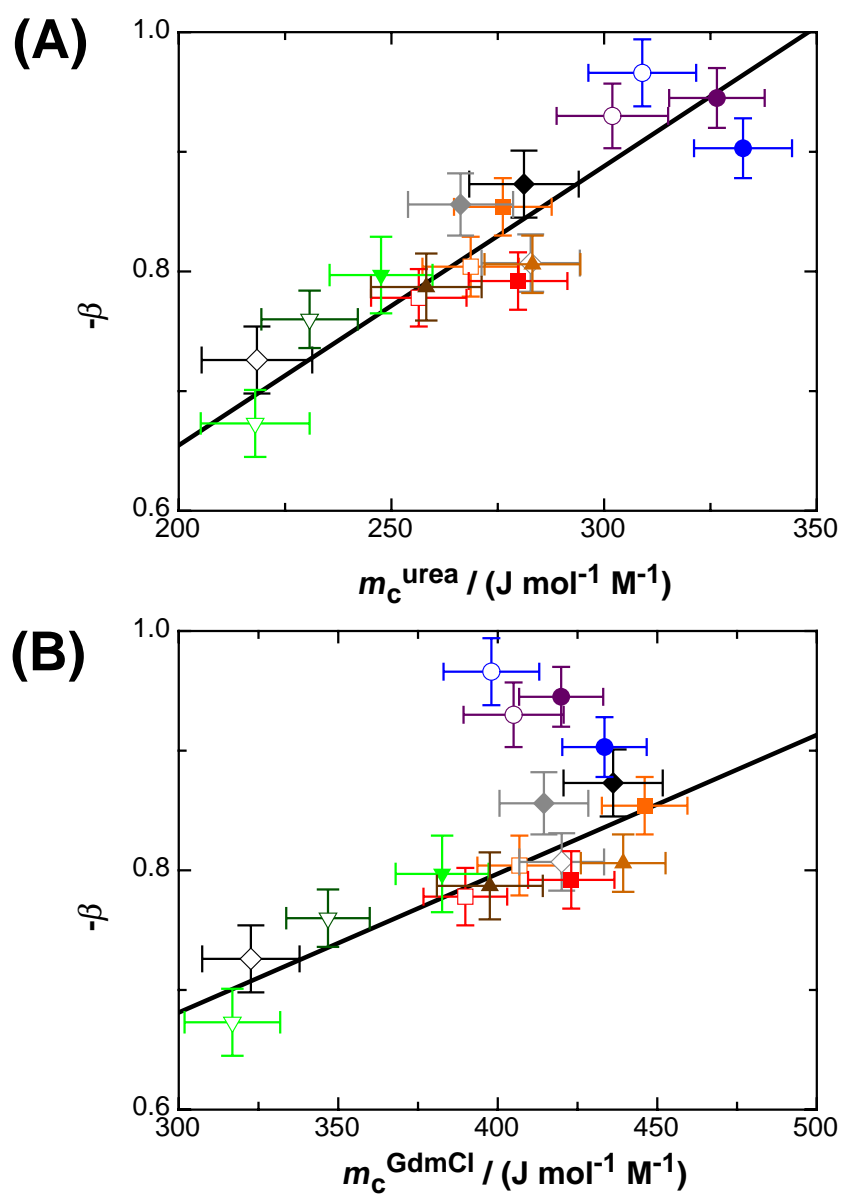
Fig. 9
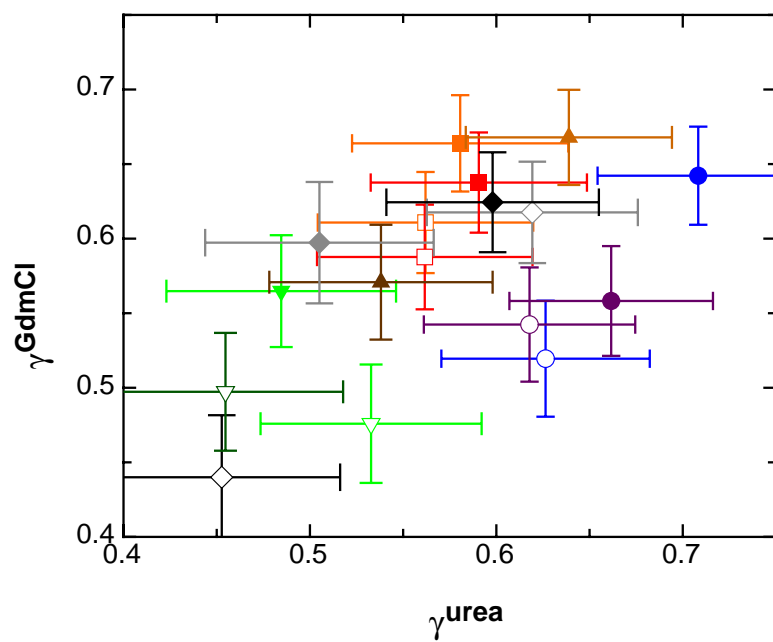
Fig. 10

Effect of Proline and Glycine Residues on Dynamics and Barriers of Loop Formation in Polypeptide Chains

Florian Krieger, Andreas Möglich, and Thomas Kiefhaber*

Contribution from the Division of Biophysical Chemistry, Biozentrum der Universität Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

Received November 30, 2004; E-mail: t.kiefhaber@unibas.ch

Abstract: Glycine and proline residues are frequently found in turn and loop structures of proteins and are believed to play an important role during chain compaction early in folding. We investigated their effect on the dynamics of intrachain loop formation in various unstructured polypeptide chains. Loop formation is significantly slower around trans prolyl peptide bonds and faster around glycine residues compared to any other amino acid. However, short loops are formed fastest around cis prolyl bonds with a time constant of 6 ns for end-to-end contact formation in a four-residue loop. Formation of short loops encounters activation energies in the range of 15 to 30 kJ/mol. The altered dynamics around glycine and trans prolyl bonds can be mainly ascribed to their effects on the activation energy. The fast dynamics around cis prolyl bonds, in contrast, originate in a higher Arrhenius pre-exponential factor, which compensates for an increased activation energy for loop formation compared to trans isomers. All-atom simulations of proline-containing peptides indicate that the conformational space for cis prolyl isomers is largely restricted compared to trans isomers. This leads to decreased average end-to-end distances and to a smaller loss in conformational entropy upon loop formation in cis isomers. The results further show that glycine and proline residues only influence formation of short loops containing between 2 and 10 residues, which is the typical loop size in native proteins. Formation of larger loops is not affected by the presence of a single glycine or proline residue.

Introduction

A major field of study in biophysical chemistry is the elucidation of the protein folding process. Starting from an ensemble of unfolded states a polypeptide chain has to form a large number of specific noncovalent intramolecular interactions during folding. To understand the complete folding reaction at the molecular level it is essential to characterize the structural and dynamic properties of the free energy landscape between the unfolded and the native state. Of particular interest is the ensemble of unfolded conformations as the starting point of the reaction. Several experimental techniques such as nuclear magnetic resonance spectroscopy[1-4] and small-angle X-ray scattering[5-7] have been applied to analyze structural properties of unfolded proteins. Less is known about the dynamics in unfolded polypeptide chains, which determine the formation of interactions during the early stages of the folding process. Especially the formation of loop and turn structures is important

early in folding, since it allows the polypeptide chain to explore energetically favorable interactions and leads to chain compaction. Here we investigate the effect of glycyl and prolyl residues on intrachain contact formation in unfolded polypeptide chains. These residues frequently occur in turn and loop sequences[8,9] and may thus play an important role early in folding.

In recent years several experimental systems have been applied to obtain information on the time scale of loop formation in polypeptides.[10-16] We have used triplet−triplet energy transfer (TTET) from xanthone (Xan) to naphthylalanine (NAla) to directly measure the kinetics of intrachain contact formation in polypeptide chains.[12,15,17] TTET from a triplet donor to a triplet acceptor group involves transfer of two electrons (Dexter mechanism) and requires van der Waals contact between the

(1) Neri, D.; Billeter, M.; Wider, G.; Wüthrich, K. *Science* **1992**, *257*, 1559−1563.
(2) Logan, T. M.; Theriault, Y.; Fesik, S. W. *J. Mol. Biol.* **1994**, *236*, 637−648.
(3) Kortemme, T.; Kelly, M. J.; Kay, L. E.; Forman-Kay, J. D.; Serrano, L. *J. Mol. Biol.* **2000**, *297*, 1217−1229.
(4) Klein-Seetharaman, J.; Oikawa, M.; Grimshaw, S. B.; Wirmer, J.; Duchardt, E.; Ueda, T.; Imoto, T.; Smith, L. J.; Dobson, C. M.; Schwalbe, H. *Science* **2002**, *295*, 1719−1722.
(5) Damaschun, G.; Damaschun, H.; Gast, K.; Zirwer, D. *Biochemistry (Moscow)* **1998**, *63*, 259−275.
(6) Bachmann, A.; Kiefhaber, T. *Biophys. Chem.* **2002**, *96*, 141−151.
(7) Millett, I. S.; Doniach, S.; Plaxco, K. W. *Adv. Protein Chem.* **2002**, *62*, 241−262.
(8) Richardson, J. S. *Adv. Protein Chem.* **1981**, *34*, 167−339.
(9) Wilmot, C. M.; Thornton, J. M. *J. Mol. Biol.* **1988**, *203*, 221−232.
(10) Haas, E.; Katchalski-Katzir, E.; Steinberg, I. Z. *Biopolymers* **1978**, *17*, 11−31.
(11) Hagen, S. J.; Hofrichter, J.; Szabo, A.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 11615−11617.
(12) Bieri, O.; Wirz, J.; Hellrung, B.; Schutkowski, M.; Drewello, M.; Kiefhaber, T. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9597−9601.
(13) Lapidus, L. J.; Eaton, W. A.; Hofrichter, J. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 7220−7225.
(14) Hudgins, R. R.; Huang, F.; Gramlich, G.; Nau, W. M. *J. Am. Chem. Soc.* **2002**, *124*, 556−564.
(15) Krieger, F.; Fierz, B.; Bieri, O.; Drewello, M.; Kiefhaber, T. *J. Mol. Biol.* **2003**, *332*, 265−274.
(16) Chang, I.-J.; Lee, J. C.; Winkler, J. R.; Gray, H. B. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3838−3840.
(17) Krieger, F.; Fierz, B.; Axthelm, F.; Joder, K.; Meyer, D.; Kiefhaber, T. *Chem. Phys.* **2004**, *307*, 209−215.

Scheme 1. Schematic Representation of Triplet−Triplet Energy Transfer Measurements in Unfolded Polypeptide Chains
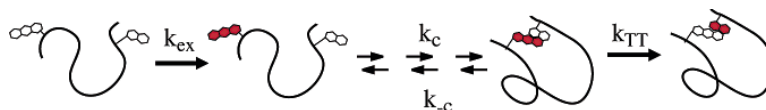
two groups.[18] Formation of the xanthone triplet state and electron transfer between xanthone and naphthalene both occur on the 1−2 ps time scale,[19] and the transfer process is diffusion-controlled.[15,17] TTET between xanthone and naphthalene thus allows measurements of absolute rate constants for diffusional processes slower than about 10 ps (Scheme 1).[19,20] In earlier experiments the triplet donor and acceptor groups were introduced near the ends of various repetitive sequences or of fragments from natural proteins.[12,15,17] This enabled us to determine the effects of chain length and amino acid sequence on the dynamics. Single-exponential kinetics on the nanoseconds time scale were observed for contact formation in all peptides. In host−guest studies on short peptides only small effects of the amino acid sequence on local chain dynamics were observed. The only exceptions were glycine and proline, which show significantly different dynamics.[15] Glycine showed faster rate constants for contact formation than any other amino acid as expected from its increased backbone flexibility due to the lack of a C_β-atom. The presence of a proline residue leads to more complex dynamics with a slow and a very fast process of contact formation.[15] Here we investigate the molecular origins of the effects of prolyl and glycyl residues on the kinetics of loop formation. We show that the heterogeneity in the prolyl containing peptides can be assigned to dynamics of cis and trans Xaa-Pro isomers. Loops of varying length are used to assess the distance dependence of the effect exerted by a single glycine, cis proline, or trans proline residue. Measuring the temperature dependence of loop formation allows us to evaluate changes in the entropy and enthalpy of activation caused by the different residues. These results are compared to chain properties of the various peptides derived from all-atom simulations.

Materials and Methods

9-Oxoxanthen-2-carboxylic acid (xanthone acid) synthesis and peptide synthesis were carried out as described previously.[15] Pseudo-proline (ΨPro) peptides were synthesized using a Fmoc-Val-[Ψ^{Me,Me}Ser] building block (Novabiochem) on an acid sensitive preloaded trityl resin (TentaGel S, Rapp Polymere) and trityl protecting groups. Cleavage conditions were 1% (v/v) trifluoroacetic acid (Fluka) in dichloromethane (Fluka). All peptides were purified by preparative HPLC. Purity and mass of the peptides were tested by analytical HPLC and mass spectrometry, respectively. Laserflash experiments were performed as described.[12,15] At the end of the fast exponential decay corresponding to TTET through intrachain contact formation a small amount of xanthone triplets (<10%) remains in the triplet state. These triplets decay on a much slower time scale with a rate corresponding to the intrinsic lifetime of xanthone triplets under the given conditions ($\tau > 20$ μs) and may be due to oligomeric structures which do not allow TTET.[15] NMR spectra were recorded on a Bruker ARX600

spectrometer. The cis content of the proline-containing peptides was determined according to the method described by Reimer et al.[21]

Viscosity Dependence. All measurements were carried out in aqeuous solutions. The viscosity was varied by adding 0−70% (v/v) glycerol (Aldrich). For all solutions, the solvent viscosity, η, was measured by using a falling sphere viscometer (Haake) at 22.5 °C. The viscosity dependencies of the rate constants for loop formation were analyzed using the empirical equation

$$k_c = k_c^0 \left(\frac{\eta}{\eta_0} \right)^\beta \tag{1}$$

where k_c is the observed contact formation rate constant, η_0 is the reference viscosity of water at 22.5 °C (0.94 cP), k_c^0 is the rate constant for intrachain contact formation at η_0, and the exponent β reflects the sensitivity of the reaction to solvent viscosity. The β-values were independent of the cosolutes used to modify solvent viscosities. We typically used water/glycerol solutions to modify solvent viscosity. However, identical β-values were obtained for glycerol, glucose, sucrose, and ethylene glycol, indicating a genuine effect of solvent viscosity on the dynamics.

Temperature Dependence. The rate constants measured at different temperatures were corrected for the effect of temperature on water viscosity. Equation 2 was used to determine the viscosity corrected activation energies (E_a):

$$k_c^0(T) = k_c(T) \left(\frac{\eta_0}{\eta(T)} \right)^\beta = A \exp\left(\frac{-E_a}{RT} \right) \tag{2}$$

where A is the Arrhenius pre-exponential factor at 22.5 °C ($\eta_0 = 0.94$ cP). The β-values were taken from the viscosity dependence of k_c (eq 1; see Table 1). The viscosity of water at various temperatures was taken from ref 22.

Simulations of Chain Conformations. All-atom simulations of oligopeptides were carried out using a hard-sphere model. Contact radii, bond lengths, and bond angles corresponded to the values described by Rose and co-workers.[23] Atom contact radii were taken from Hopfinger[24] and softened by multiplying with a factor of 0.95. Bond lengths and angles were obtained from Engh and Huber.[25] The dihedral angle ω was fixed at 179.5° for trans peptide bonds and at −0.5° for a cis prolyl peptide bond. For proline residues the dihedral angle ϕ was fixed at −60°.[26] Peptide conformations were generated by randomly varying the values of the backbone dihedral angles ϕ and ψ and the side-chain dihedral angle χ_1 of serine residues. Calculated peptide conformations were checked for steric overlaps using the above hard-sphere contact radii. Following Pappu et al.[23] only pairs of atoms separated by at least four covalent bonds were tested for steric clashes. In addition, N-terminal amide proton and C-terminal carbonyl oxygen atoms were not checked for steric interactions. For each peptide at least 10^5 valid conformations without steric overlaps were generated by this procedure. End-to-end distances reflect the distance between the N-terminal amide nitrogen and the C-terminal carbonyl carbon atom.

(18) Klessinger, M.; Michl, J. *Excited states and photochemistry of organic molecules*; VCH: Weinheim, 1995.
(19) Satzger, H.; Schmidt, B.; Root, C.; Zinth, W.; Fierz, B.; Krieger, F.; Kiefhaber, T.; Gilch, P. *J. Phys. Chem. A* **2004**, *108*, 10072−10079.
(20) Fierz, B.; Kiefhaber, T. In *Protein Folding Handbook*; Buchner, J., Kiefhaber, T., Eds.; WILEY-VCH Verlag GmbH & Co KGaA: Weinheim, 2005; pp 805−851.
(21) Reimer, U.; Scherer, G.; Drewello, M.; Kruber, S.; Schutkowski, M.; Fischer, G. *J. Mol. Biol.* **1998**, *279*, 449−460.
(22) *Handbook of Chemistry and Physics*, 53rd ed.; CRC: Cleveland, OH, 1972.
(23) Pappu, R. V.; Srinivasan, R.; Rose, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 12565−12570.
(24) Hopfinger, A. J. *Conformational Properties of Macromolecules*; Academic Press: New York, 1973.
(25) Engh, R. A.; Huber, R. *Acta Crystallogr.* **1991**, *47*, 392−400.
(26) Schimmel, P. R.; Flory, P. J. *Proc. Natl. Acad. Sci. U.S.A.* **1967**, *58*, 52−59.

Table 1. Rate Constants and Activation Parameters for Intrachain Contact Formation in Various Peptides

sequence[a]	k_1[b] (10^6 s^{-1})	A_1[c] (s^{-1})	E_a^1[d] (kJ/mol)	β_1[e]	k_2[b] (10^6 s^{-1})	A_2[c] (s^{-1})	E_a^2[a] (kJ/mol)	β_2[e]	% cis[f]
SP	9.5 ± 0.5	(1.7 ± 0.8) × 10^{10}	18.3 ± 1.2	−0.75 ± 0.03	84 ± 8	(2.7 ± 4.0) × 10^{15}	42.9 ± 3.4	−0.69 ± 0.04	15 ± 1
SS	75 ± 4	(5.1 ± 1.0) × 10^{10}	16.1 ± 0.3	−0.83 ± 0.03					
SPS	20 ± 1	(4.1 ± 1.8) × 10^{10}	18.8 ± 1.2	−0.75 ± 0.02	120 ± 20	nd	nd	−0.77 ± 0.04	17 ± 2
SSS	72 ± 8	(2.8 ± 1.0) × 10^{10}	14.7 ± 0.8	−0.86 ± 0.06					
SGS	120 ± 10	(1.1 ± 1.3) × 10^{10}	11.4 ± 3.0	−0.83 ± 0.03					
VPS	9.5 ± 1.0	(2.9 ± 1.4) × 10^{10}	20.0 ± 1.2	−0.81 ± 0.05	92 ± 8	(1.6 ± 1.3) × 10^{13}	29.8 ± 3.4	−0.77 ± 0.04	16 ± 2
V(ΨP)S	9.7 ± 0.4	nd	nd	−0.80 ± 0.05	98 ± 5	nd	nd	−0.78 ± 0.06	72 ± 5
SSPS	28 ± 2	nd	nd	−0.83 ± 0.04	150 ± 20	nd	nd	−0.87 ± 0.06	18 ± 4
SVPS	16 ± 1	nd	nd	−0.86 ± 0.02	71 ± 5	nd	nd	−0.96 ± 0.04	18 ± 2
SV(ΨP)S	15 ± 1	nd	nd	−0.83 ± 0.06	75 ± 5	nd	nd	−0.98 ± 0.03	78 ± 6
S$_4$PS$_4$	37 ± 2	(2.4 ± 0.7) × 10^9	10.3 ± 0.7	−0.93 ± 0.03					
S$_9$	40 ± 2	(1.7 ± 0.8) × 10^9	9.2 ± 1.0	−0.92 ± 0.03					
S$_4$GS$_4$	48 ± 3	(1.6 ± 0.8) × 10^9	8.7 ± 1.2	−0.93 ± 0.02					

[a] Loop sequence, X, between the labels for TTET in peptides of the canonical sequence Xan-X-NAla-Ser-Gly. [b] k_1 and k_2 are the rate constants of the slower and the faster kinetic phase, respectively, at 22.5 °C in water ($\eta = 0.94$ cP). [c] Arrhenius pre-exponential factor in water at 22.5 °C ($\eta = 0.94$ cP) corrected for temperature effects on solvent viscosity according to eq 2. [d] Arrhenius activation energy corrected for temperature effects on solvent viscosity according to eq 2. The apparent activation energies before viscosity correction can be calculated by adding ($-\beta \times 17.7$) kJ/mol to the given values. [e] β-values reflect the sensitivity of the respective kinetic phase to solvent viscosity and were determined by fitting the viscosity dependence of the rate constants according to eq 1. The origin of the fractional β-values will be discussed elsewhere (F.K. and T.K., in preparation). [f] The cis content reflects the relative amplitude of the fast kinetic phases. The values obtained from measurements at different viscosities were averaged. [g] Loop formation in the cis isomer of the SPS peptide was too fast to reliably determine activation energies.

Ramachandran maps for cis and trans prolyl isomers were calculated for the central serine residue in Ser-Ser-Pro peptides[27] as described above. Conformational free energies were calculated by dividing the resulting Ramachandran maps into bins of 5×5 degrees. The number of conformations per bin was used to calculate relative free energies according to the Boltzmann equation. The most frequent conformation was arbitrarily assigned a value of 0; all other free energies are relative to that value.

The simulation program was written in ANSI C and compiled using the Bloodshed Dev-C++ development environment (http://www.bloodshed.net). Compiled programs were executed on IBM-compatible computers running Windows.

Results and Discussion

Kinetics of Loop Formation in Short Proline- and Glycine-Containing Peptides. To investigate the effect of prolyl and glycyl residues on chain dynamics in more detail we measured the kinetics of loop formation in host−guest peptides of the structure Xan-(Ser)$_x$-**Xaa**-(Ser)$_y$-NAla-Ser-Gly with $x,y = 0−7$ and Xaa = Gly, Ser, or Pro. Figure 1 compares TTET kinetics in the Xan-Ser-**Ser**-Ser-NAla-Ser-Gly peptide (SSS peptide)[28] with Xan-Ser-**Pro**-Ser-NAla-Ser-Gly (SPS peptide) and Xan-Ser-**Gly**-Ser-NAla-Ser-Gly (SGS peptide). These peptides serve as good models for tight protein turns with i, $i+3$ interactions which frequently occur at the end of β-hairpins.[29] Intrachain loop formation in the SSS and SGS peptides can be described by single-exponential kinetics with rate constants of $(1.6 \pm 0.1) \times 10^7$ s^{-1} and $(2.5 \pm 0.1) \times 10^7$ s^{-1}, respectively, in the presence of 52% glycerol. Under the same conditions, double exponential kinetics are observed for loop formation in the SPS peptide with rate constants of $(4.7 \pm 0.2) \times 10^6$ s^{-1} and $(2.7 \pm 0.3) \times 10^7$ s^{-1} and respective amplitudes of 80 ± 5% and 20 ± 5%. The faster kinetic phase in the SPS peptides is slightly faster than loop formation in the SGS peptide, whereas the slower phase is the slowest reaction observed in the three
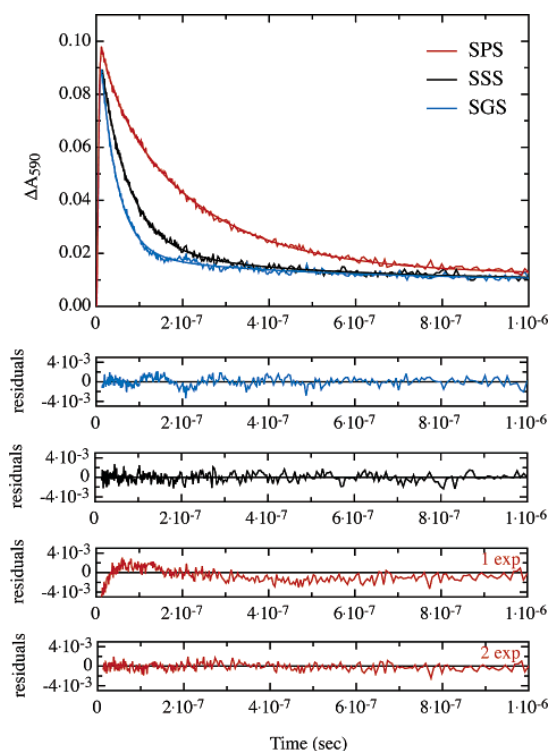
(27) Schimmel, P. R.; Flory, P. J. *J. Mol. Biol.* **1968**, *34*, 105−120.
(28) The peptides are abbreviated using one letter codes for the amino acid sequence (X) between the xanthone (Xan) and the naphthylalanine (NAla) labels. The peptides have the general sequence Xan-X-NAla-Ser-Gly. Thus, an SPS peptide denotes the peptide sequence Xan-Ser-Pro-Ser-NAla-Ser-Gly.
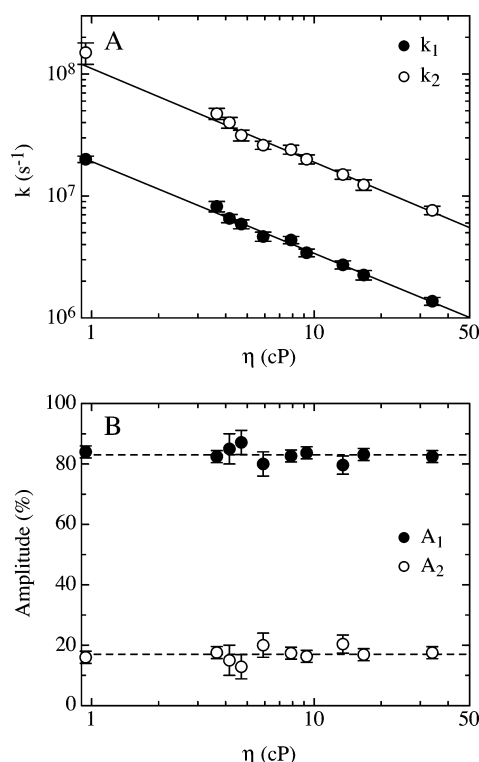(29) Sibanda, B. L.; Thornton, J. M. *Nature* **1985**, *316*, 170−174.

Figure 1. Effect of proline and glycine residues on the kinetics of loop formation in peptides of the sequence Xan-Ser-Xaa-Ser-NAla-Ser-Gly with Xaa = Ser, Gly, or Pro.

peptides. Since the faster kinetic phase of the proline peptide in water is near the resolution limit of the measurements, the experiments were carried out in the presence of increasing amounts of glycerol between 0% and 70% to slow chain dynamics due to increased solvent viscosity. The experiments on the water/glycerol mixtures allow a more reliable determination of the rate constant and amplitude of the fastest reaction via extrapolation to pure water (Figure 2A). The rate constants for loop formation in the SPS peptide in water at 22.5 °C are

Figure 2. Viscosity dependence of the rate constants (A) and their respective amplitudes (B) for the kinetics of loop formation in a Xan-Ser-Pro-Ser-NAla-Ser-Gly peptide at 22.5 °C. Glycerol was used as cosolute to adjust solvent viscosity. The dashed lines in part B indicate the average amplitudes.

$(2.0 \pm 0.1) \times 10^7$ s^{-1} and $(1.2 \pm 0.2) \times 10^8$ s^{-1} (see Figure 2A and Table 1). For the SSS and SGS peptides rate constants of $(7.2 \pm 0.8) \times 10^7$ s^{-1} and $(1.2 \pm 0.1) \times 10^8$ s^{-1}, respectively, were determined (Table 1). Glycerol has no effect on the amplitudes of the two kinetic phases. The average amplitudes for the faster and slower process measured at various glycerol concentrations are $17.0 \pm 2.2\%$ and $83.0 \pm 2.2\%$, respectively (Figure 2B).

A likely origin for the double exponential kinetics in the SPS peptide are different dynamics in cis and trans isomers of the Ser-Pro peptide bond. Xaa-Pro peptide bonds populate a large fraction of cis isomer due to the cyclic structure of the prolyl moiety. Studies on model peptides showed that the cis content depends on the preceding amino acid and varies between 7% and 38% [21] in contrast to about 0.15%−0.5% cis isomer at nonprolyl peptide bonds.[30] The observed 17% fast phase in the loop closure dynamics in the SPS peptide essentially reflects the cis−trans ratio at the Ser-Pro peptide bond in the SPS peptide, which has a cis content of $16 \pm 2\%$ as determined by 1D ^{1}H NMR spectroscopy (data not shown). Since cis−trans isomerization is slow ($1/k_{app} \sim 20$ s at 22.5 °C),[21,31,32] no equilibration between the two isomers occurs on the time scale of the TTET experiments. Thus loop formation in both isomers

(30) Scherer, G.; Kramer, M. L.; Schutkowski, M.; Reimer, U.; Fischer, G. *J. Am. Chem. Soc.* **1998**, *120*, 5568−5574.
(31) Brandts, J. F.; Halvorson, H. R.; Brennan, M. *Biochemistry* **1975**, *14*, 4953−4963.
(32) Keller, M.; Sager, C.; Dumy, P.; Schutkowski, M.; Fischer, G. S.; Mutter, M. *J. Am. Chem. Soc.* **1998**, *120*, 2714−2720.

should be observable if they have different dynamics. However, also rotation around the ψ-angle of proline was shown to be slow with a rotation barrier around 60 kJ/mol.[33] This might give rise to the observed heterogeneity in the kinetics of loop formation in prolyl-containing peptides. To discriminate between cis/trans isomers of Xaa-Pro bonds and slow rotation around the proline ψ-angle as the molecular origin for the complex dynamics, we investigated the effect of the cis content on the kinetics of loop formation. Pseudo-proline (ΨPro) is known to increase the cis content of a Val-Pro peptide bond to about 80% due to steric effects induced by the methyl groups in the trans conformation.[32,34] We introduced ΨPro into two different peptides of the sequence Xan-Val-ΨPro-Ser-NAla-Ser-Gly (VΨPS-peptide **A**) and Xan-Ser-Val-ΨPro-Ser-NAla-Ser-Gly (SVΨPS-peptide) and compared the resulting kinetics of loop formation to the kinetics of the corresponding proline-containing peptides VPS (peptide **B**) and SVPS.
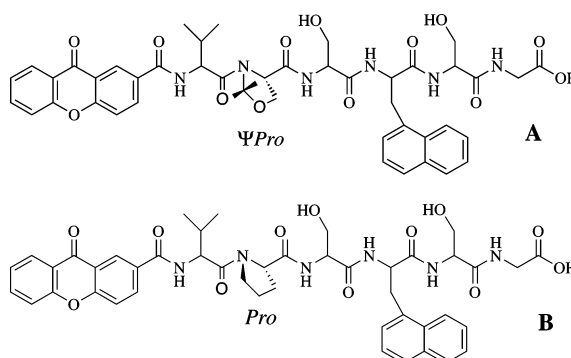
Both ΨPro-containing peptides show double exponential kinetics of contact formation with significantly increased amplitude of the faster reaction (70%−80%) compared to the corresponding prolyl-containing peptides (Figure 3A and Table 1). This shows that the faster reaction corresponds to the dynamics of peptides with a cis isomer and is not caused by slow rotation around the proline ψ-angle.

Temperature Dependence of Loop Formation. Information on the origin of the faster dynamics in chains with cis prolyl isomers was obtained from the temperature dependence of loop formation in the Xan-Val-Pro-Ser-NAla-Ser-Pro peptide (VPS peptide). Fitting the data to the Arrhenius equation yielded apparent activation energies of 43.2 ± 2.0 kJ/mol for the cis isomer and 35.9 ± 1.2 kJ/mol for the trans isomer. Since the dynamics were shown to be sensitive to solvent viscosity (Figure 2), the rate constants measured at different temperatures had to be corrected for contributions from the effect of temperature on water viscosity to obtain the actual activation energies. Since, the kinetics of loop formation do not correlate with $1/\eta$ but show fractional viscosity dependencies ($k \sim 1/(\eta)^\beta$; see eq 1 and Table 1), we used eq 2 to determine viscosity-corrected activation energies. The resulting activation energies (E_a) were 29.8 ± 3.4 and 20.0 ± 1.2 kJ/mol for loop formation in peptides with a cis and a trans Val-Pro peptide bond, respectively (Figure 3B, Table 1). The results further revealed largely different

(33) Nagaraj, R.; Venkatachalapathi, Y. V.; Balaram, P. *Int. J. Pept. Protein Res.* **1980**, *16*, 291−298.
(34) Dumy, P.; Keller, M.; D. E., R.; Rohwedder, B.; Wohr, T.; Mutter, M. *J. Am. Chem. Soc.* **1997**, *119*, 918−925.
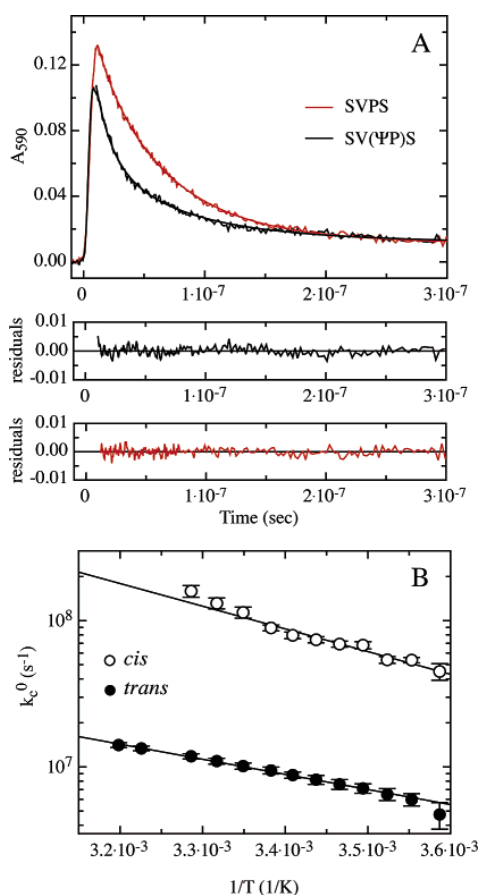
Figure 3. (A) Kinetics of intrachain contact formation in Xan-Val-**Xaa**-Ser-NAla-Ser-Gly peptides with Xaa = ΨPro (peptide **A**) or Pro (peptide **B**). (B) Temperature dependence of intrachain contact formation in cis and trans isomers of the Xan-Val-Pro-Ser-NAla-Ser-Gly peptide. A fit to the Arrhenius equation (eq 2) gives viscosity-corrected activation energies and pre-exponential factors (*A*) indicated in Table 1.

Arrhenius pre-exponential factors (*A*) of $(1.6 \pm 1.3) \times 10^{13}$ s^{-1} and $(2.9 \pm 1.4) \times 10^{10}$ s^{-1} for loop formation in the cis and trans isomer, respectively. Obviously, the fast loop formation of the cis isomer is the result of a significantly higher pre-exponential factor despite a higher activation energy compared to the trans peptide bonds. This indicates favorable entropic contributions to the rate constant for loop closure in the cis isomer. A significantly higher activation energy of 42.9 ± 3.4 kJ/mol and a higher pre-exponential factor of $(2.7 \pm 4) \times 10^{15}$ s^{-1} are found for the cis isomer in the shorter Xan-Ser-Pro-NAla-Ser-Gly peptide (SP peptide; Table 1), whereas the activation parameters for dynamics in the trans isomer of this peptide are similar to the respective values in the trans isomer of the VPS peptide (Table 1). For all proline-containing peptides the pre-exponential factors for the loop closure dynamics in the trans Pro isomers are similar to the values in polyserine peptides of similar length, whereas the activation energies in the polyserine peptides are significantly lower (Table 1). This indicates that the presence of a trans prolyl isomer mainly influences the height of the energy barrier for loop formation but not the entropic contributions. The activation parameters for loop formation in the SGS peptide reveal that
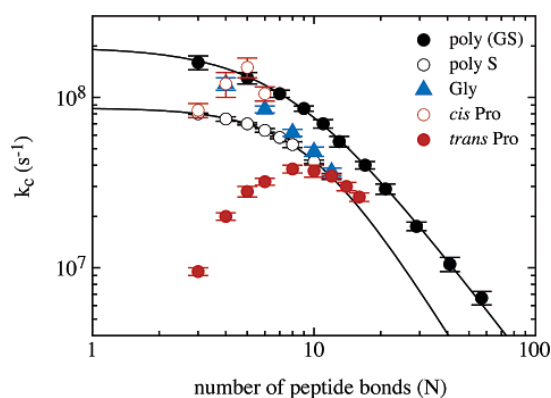
Figure 4. Effect of a single proline or glycine residue on the rate constant of loop formation in peptides of varying length compared to polyserine and poly(glycine-serine) peptides. Data for polyserine and poly(Gly-Ser) were taken from ref 15. The lines describe the loop size dependence of intrachain loop formation in the polyserine and poly(glycine-serine) series as described in ref 15.

glycine decreases the activation energy for loop formation compared to serine- and proline-containing peptides (Table 1). A likely origin of this effect is the lower rotation barrier around the ϕ, ψ angles in glycine. This accelerates loop formation despite a decreased pre-exponential factor, which can be explained by increased conformational space around glycyl residues (Table 1).

Linear Arrhenius plots were observed for all peptides, which shows that loop formation is not associated with significant changes in heat capacity ($\Delta C_p \approx 0$). This argues against major changes in solvation during the rate-limiting motions for loop formation. The results on the dynamics in the different proline-containing peptides further reveal that the amino acid preceding the prolyl residue significantly influences the dynamics of both the cis and the trans isomer. In all short peptides Val-Pro bonds exhibit about 2-fold slower loop formation kinetics compared to Ser-Pro bonds (Table 1). This indicates steric effects on loop formation imposed by the amino acid preceding the prolyl residue. Interestingly, the bulkier ΨPro residue gives virtually the same rate constant for loop formation as Pro in all peptides.

Effect of Loop Size. The results show that proline and glycine residues strongly influence the kinetics of formation of short loops. To test whether these effects are also observed over longer distances we measured loop formation in peptides containing a single prolyl or glycyl residue at a central position and varied the loop size. Figure 4 compares the kinetics of loop formation in peptides of the sequence Xan-(Ser)$_x$-**Xaa**-(Ser)$_y$-NAla-Ser-Gly (Xaa = Gly or Pro, $x,y = 0-7$) with the results from polyserine and poly(glycine-serine) loops. The effect of a central prolyl and glycyl residue decreases with increasing loop size. A glycyl residue has no effect on loop formation if the sites of contact formation are further than five amino acids from both sides of the glycin. Peptides with $N \geq 11$ (N = number of peptide bonds between the two sites of contact formation) show the same rate constants for contact formation as polyserine chains of the same length (Figure 4). A more complex behavior is observed for prolyl residues. Only for $x,y \leq 2$ the dynamics of peptides with cis and trans isomers can be resolved, suggesting that the dynamics of two isomers are similar in longer peptides (Figure 4). In short peptides the cis isomer always shows faster
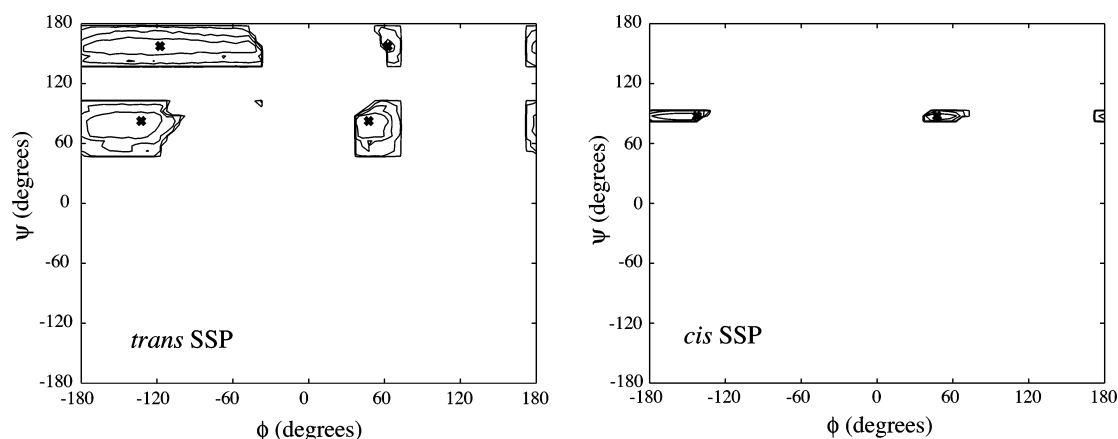
Figure 5. Ramachandran maps for the serine residue in either the cis or the trans Ser-Pro peptide bond in Ser-Ser-Pro peptides. The contour levels denote free energies of 0.5, 1, 2, and $4k_BT$. The crosses indicate the conformation of lowest free energy.

loop formation than the trans isomer, irrespective of whether donor and acceptor are separated by an even or an uneven number of peptide bonds (Figure 4 and Table 1). Formation of short loops becomes faster with increasing loop size in chains with cis Pro isomers and the rate constant reaches a maximum of $k_c = 1.5 \times 10^8 \text{ s}^{-1}$ for $N = 5$, i.e., if two amino acids are located on both sides of the cis prolyl bond. The rate constant for loop formation in peptides with trans prolyl isomers also increases with increasing chain length until a loop size of 9–10 residues and matches dynamics of a polyserine chain at $N = 11$. Formation of single proline-containing loops with $N \geq 11$ behave similar to polyserine chains and show decreasing rate constants with increasing loop size (Figure 4). These results show that single glycine and proline residues affect chain dynamics only very locally and do not influence formation of loops if the contact sites are further than five amino acids on either side of the glycyl or prolyl residue.

Effect of Cis and Trans Prolyl Peptide Bonds on Chain Dimensions. The temperature dependence of the loop closure reaction showed that the faster dynamics in cis prolyl isomers are due to a significantly increased pre-exponential factor compared to trans isomers indicating entropically favored loop formation in cis isomers. A likely origin of this effect are differences in the equilibrium properties of chains containing cis or trans isomers. This prompted us to evaluate the sterically allowed conformations for the model polypeptide chains. Using all-atom Monte Carlo simulations of conformational space introduced by Rose and co-workers[23,35] we calculated the probability distribution of end-to-end distances for different chains containing either cis or trans Ser-Pro isomers and compared the results to the dimensions of polyserine chains. In these calculations all bond lengths, bond angles, and peptide bond torsion angles, ω, were held at the values given by Pappu et al.[23] Sterically allowed ϕ,ψ pairs were evaluated for each amino acid, and the resulting allowed regions in the ϕ,ψ space were used to generate energetically favorable conformations for each peptide.[23] The results show that the energetically favorable conformational space of a cis prolyl isomer in a Ser-Pro bond is drastically reduced compared to the trans isomer (Figure 5). This is in agreement with an analysis of nonprolyl cis peptide

bonds in protein structures, which revealed that also in native proteins the conformational space of cis peptide bonds is significantly restricted compared to trans peptide bonds.[36] Our calculations further revealed that a central cis Ser-Pro peptide bond leads to a distance distribution function that is significantly shifted to shorter end-to-end distances compared to a trans bond in all peptides (Figure 6). This is consistent with faster rate constants for intrachain contact formation in chains with a cis peptide bond due to a smaller average distance between the sites of contact formation. However, the relative difference in end-to-end distance between cis and trans isomers is small for longer chains, which might explain the inability to resolve both kinetic phases for contact formation in longer peptides (Figure 6A). These results show that a single cis Xaa-Pro peptide bond significantly influences the available conformational space of a polypeptide chain and leads to reduced chain dimensions. This is compatible with earlier calculations on poly(L-proline), which also showed that the presence of a small fraction of cis prolyl isomers drastically reduces the chain dimensions.[37] A similar behavior was observed for stereoirregular vinyl polymers[38] and in stereoirregular 1,4-polybutadiene and 1,4-polyisoprene,[39] where the presence of a small fraction of stereoirregularity significantly reduces chain dimension.

The restricted conformational space and the shorter end-to-end distances found for cis Xaa-Pro peptide bonds explain the higher pre-exponential factor for loop closure reactions due to a smaller loss of conformational entropy. The results from the simulations are further compatible with higher activation energies for chain motions in cis isomers due to increased barriers for bond rotation caused by steric clashes during bond rotations on a restricted conformational space (Figure 6B). The simulations also show that trans prolyl isomers have similar chain dimensions as the corresponding serine-conatining peptides, which is in agreement with virtually identical pre-exponential factors for loop formation for these peptides.

(35) Fitzkee, N. C.; Rose, G. D. *Protein Sci.* **2004**, *13*, 633–639.

(36) Jabs, A.; Weiss, M. S.; Hilgenfeld, R. *J. Mol. Biol.* **1999**, *286*, 291–304.
(37) Tanaka, S.; Scheraga, H. A. *Macromolecules* **1975**, *8*, 623–631.
(38) Flory, P. J. *Statistical Mechanics of Chain Molecules*; Hanser Publishers: Munich, 1969.
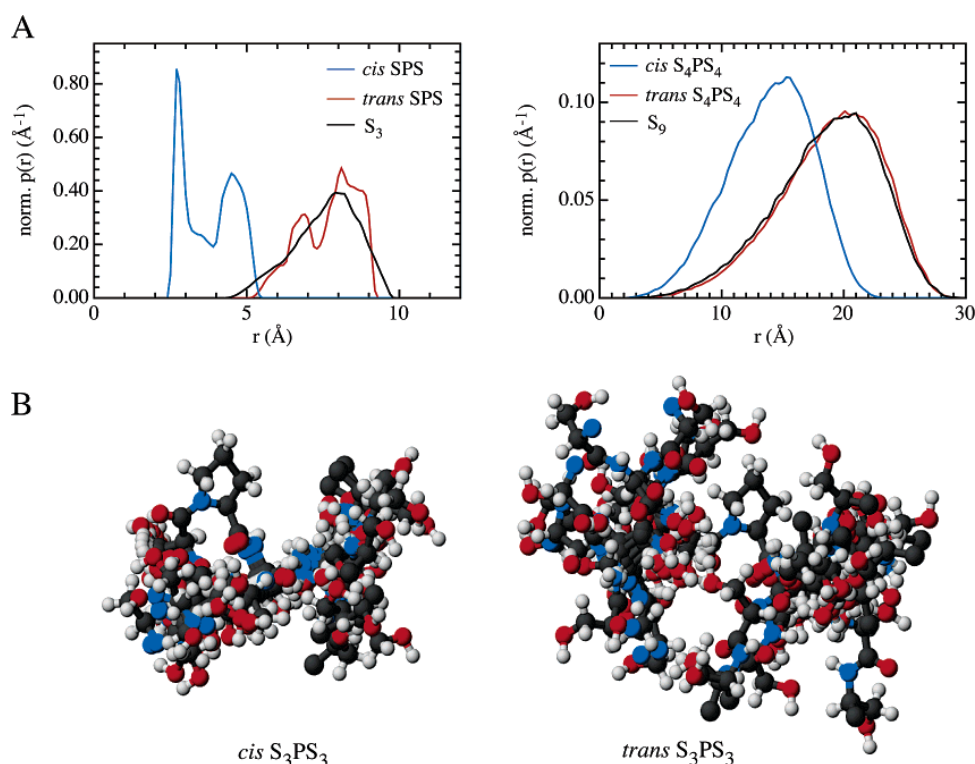(39) Tanaka, S.; Nakajima, A. *Polym. J.* **1972**, *3*, 600–506.

Figure 6. (A) Normalized end-to-end distance distribution in cis and trans prolyl isomers in various peptides compared to polyserine peptides of the same length (B) Representative structures (10) of cis and trans prolyl isomers of a Ser$_3$-Pro-Ser$_3$ peptide. The structures are overlaid with respect to identical orientation of the central prolyl residue. The structures are displayed using the program MOLMOL.[41]

Conclusions

Glycine and proline residues have a major influence on the kinetics of loop formation in proteins. Glycine accelerates loop formation by decreasing the activation energy, whereas trans prolyl bonds slow loop formation by increasing the barrier height. The most interesting effect is observed for loop formation around cis prolyl bonds, which show the fastest kinetics of all sequences despite an increased activation energy. The fast loop formation around cis prolyl isomers is due to largely restricted conformational space and shorter end-to-end distances compared to trans isomers which results in a drastically increased pre-exponential factor for loop formation. A single prolyl or glycyl residue influences chain dynamics only locally and has no effect on formation of loops longer than 11 residues. The most abundant loop and turn structures in natural proteins have average sizes of 6−10 residues (ω-loops)[40] and of 2−7 residues (β-hairpin loops)[29] and are especially rich in proline and glycine.

Our results show that the rate constants for formation of these loops during protein folding can be modulated by a factor of 10 by the presence of either glycine/cis proline or trans proline, which might help to direct the folding polypeptide chain to productive folding routes. The significantly different dynamic and conformational properties of loops containing cis and trans prolyl peptide bonds may further be important for the function of prolines as molecular switches in regulation of cellular processes.

JA042798I

(40) Leszcynski, J. F.; Rose, G. D. *Science* **1986**, *234*, 849−855.
(41) Koradi, R.; Billeter, M.; Wüthrich, K. *J. Mol. Graphics* **1996**, *14*, 51−55.

JMB

Very Fast Folding and Association of a Trimerization Domain from Bacteriophage T4 Fibritin

Sarah Güthe[1]†, Larisa Kapinos[1]†, Andreas Möglich[1]†
Sebastian Meier[2], Stephan Grzesiek[2] and Thomas Kiefhaber[1]*

[1]*Division of Biophysical Chemistry, Biozentrum der Universität Basel Klingelbergstrasse 70, CH-4056 Basel, Switzerland*

[2]*Division of Structural Biology Biozentrum der Universität Basel, Klingelbergstrasse 70 CH-4056 Basel, Switzerland*

*Corresponding author

The foldon domain constitutes the C-terminal 30 amino acid residues of the trimeric protein fibritin from bacteriophage T4. Its function is to promote folding and trimerization of fibritin. We investigated structure, stability and folding mechanism of the isolated foldon domain. The domain folds into the same trimeric β-propeller structure as in fibritin and undergoes a two-state unfolding transition from folded trimer to unfolded monomers. The folding kinetics involve several consecutive reactions. Structure formation in the region of the single β-hairpin of each monomer occurs on the submillisecond timescale. This reaction is followed by two consecutive association steps with rate constants of $1.9(\pm 0.5) \times 10^6 \, M^{-1} \, s^{-1}$ and $5.4(\pm 0.3) \times 10^6 \, M^{-1} \, s^{-1}$ at 0.58 M GdmCl, respectively. This is similar to the fastest reported bimolecular association reactions for folding of dimeric proteins. At low concentrations of protein, folding shows apparent third-order kinetics. At high concentrations of protein, the reaction becomes almost independent of protein concentrations with a half-time of about 3 ms, indicating that a first-order folding step from a partially folded trimer to the native protein ($k = 210(\pm 20) \, s^{-1}$) becomes rate-limiting. Our results suggest that all steps on the folding/trimerization pathway of the foldon domain are evolutionarily optimized for rapid and specific initiation of trimer formation during fibritin assembly. The results further show that β-hairpins allow efficient and rapid protein–protein interactions during folding.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: protein folding; protein association; trimeric proteins; prolyl isomerization; fast folding

Introduction

Fibritin is a rod-like structural protein of bacteriophage T4, which is attached to the neck of the virion *via* its N-terminal domain to form the collar structures ("whiskers"). Fibritin consists of an N-terminal anchor domain (residues 1–46), a large central coiled-coil part (residues 47–456) and a small C-terminal globular domain (residues 457–486).[1] The 30 amino acid residue C-terminal domain was termed foldon, since it was shown to be essential for fibritin trimerization and folding *in vivo* and *in vitro*.[1–3] Each subunit of the foldon domain consists of a single β-hairpin, which assemble into a β-propeller-like structure in the trimer.[1] The trimer is stabilized by hydrophobic interactions involving Trp476 of each subunit, intermolecular salt-bridges between Glu461 and Arg471, and intermolecular backbone hydrogen bonds between Tyr469 and Arg471 (Figure 1). Expression of the isolated foldon domain (residues 457–483) yields a stable trimer, which shows a cooperative two-state thermal unfolding transition.[4] Residues 484–486 were omitted from this study, since this region is unordered in the X-ray structure of fibritin.[1]

The foldon domain was proposed to be an evolutionarily optimized trimerization/folding
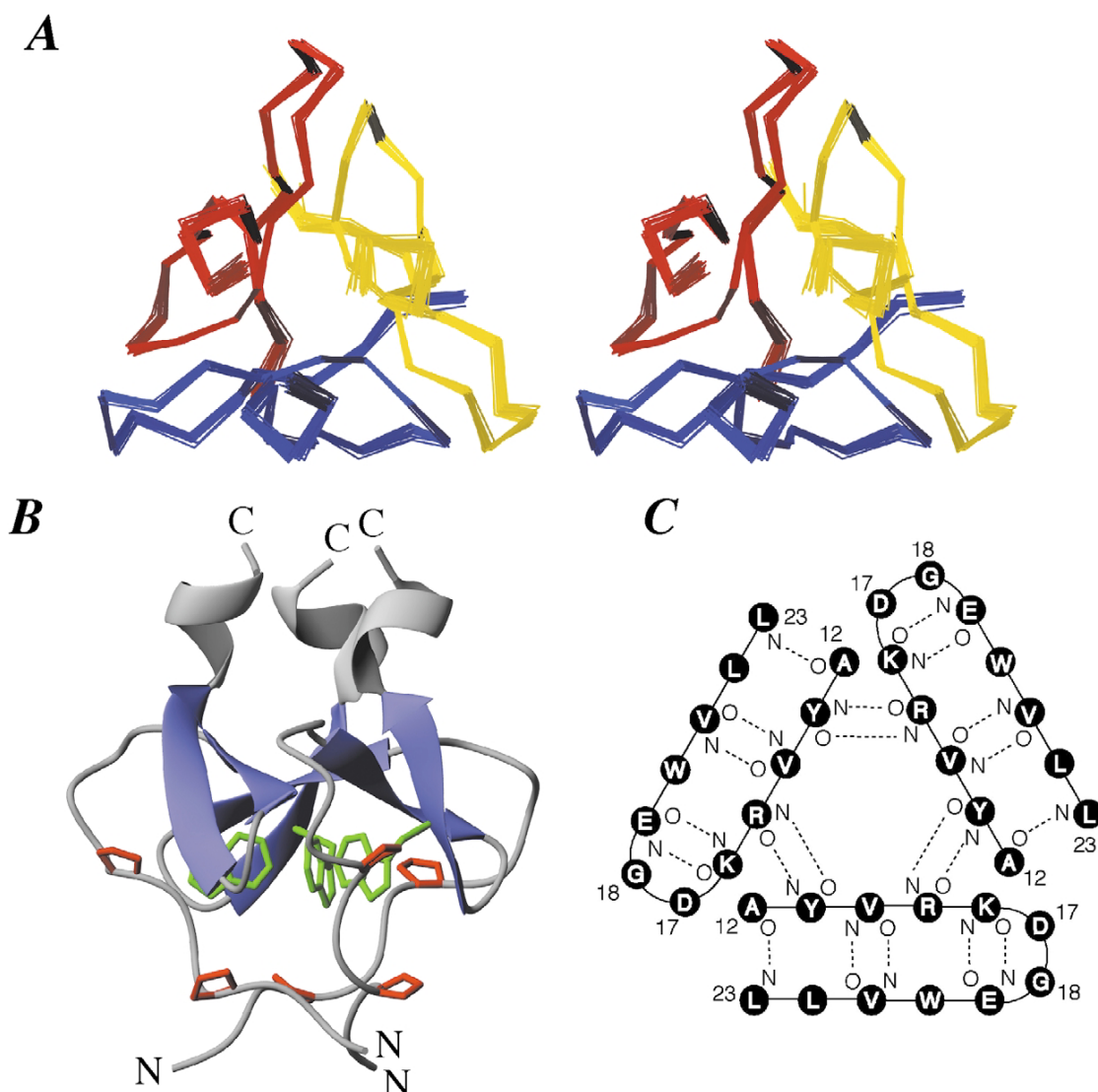

Figure 1. A, Stereo view of a bundle of the 20 lowest-energy structures of the trimeric foldon domain determined by NMR spectroscopy. Each subunit is displayed in a different color. B, Side view of the foldon structure with the single Trp residues at position 20 of each chain highlighted in green and the two prolyl residues at positions 4 and 7 highlighted in red. C, Topology of the interactions of the three β-hairpins in the native foldon domain. The figures in A and B were prepared using the program MOLMOL[37] and rendered with PovRay.

motif, as its only known function is to promote folding of fibritin.[3] The small size of its structured part (27 amino acid residues) and its simple fold make the foldon domain a perfect system for a detailed study on the mechanism of a folding reaction linked to intermolecular association steps. All previous folding studies on trimeric proteins investigated large filamentous proteins, which show extremely slow and complex folding kinetics, usually accompanied by irreversible aggregation reactions.[5]

We expressed the foldon domain in *Escherichia coli* and synthesized it by solid-phase *N*-(9-fluorenyl)methoxycarbonyl (F-moc) chemistry to investigate its structure, stability and folding mechanism. For clarity, we are numbering the

foldon sequence from residues 1 to 27

```
 1           11          21
GYIPEAPRDG QAYVRKDGEW VLLSTFL
```

corresponding to residues 457–483 in fibritin. All kinetic and stability data presented here were obtained using the chemically synthesized foldon domain, whereas the recombinant *E. coli* product was used for structural analysis. The *E. coli* product and the synthetic foldon domain showed identical stability and folding behavior. Further, the additional C-terminal amino acid residues Ser-Pro-Ala, which are present in the wild-type fibritin sequence, do not affect any thermodynamic or kinetic properties of the foldon domain.

Results and Discussion

Structure and stability of the foldon domain

To test whether the 27 amino acid residue foldon domain adopts the same fold as in fibritin, we solved its solution structure to a backbone rmsd of 0.31 Å with 28 experimental restraints per residue (Figure 1 and Table 1). The structure largely resembles the crystal structure in constructs carrying the 75 and 120 C-terminal amino acid residues of fibritin, in which the foldon domain constitutes only a minor part of the total construct.[1,6] Only at the immediate N terminus (residues 1–3) the isolated foldon domain assumes a slightly different and presumably more relaxed structure compared

to fibritin. The trimer consists of an N-terminal hydrophobic stretch in left-handed polyproline II helix conformation between Pro4 and Pro7, which is connected to a β-hairpin (residues 12–23) and forms a hydrophobic cap of the hairpin on the N-terminal side (Figure 1). The hairpin terminates in a 3_{10} helix at the C terminus with homophilic interactions of hydrophobic residues (Tyr2, Ile3, Val14, Leu23, Leu27) between the monomers along the symmetry axis. Large-scale nanosecond dynamics as evidenced by ^{15}N relaxation occur only at the most N-terminal residue, Tyr2. All other residues in the highly rigid foldon domain exhibit order parameters $S^2 > 0.77$ at 25 °C as determined by the program TENSOR.[7]

The equilibrium unfolding properties of the foldon domain were measured by guanidinium chloride (GdmCl)-induced unfolding transitions at various concentrations of protein. Figure 2 shows unfolding curves at monomer concentrations of 5 μM and 30 μM. The coincidence of fluorescence and CD-monitored transition curves demonstrates that the trimer unfolds in a cooperative two-state transition at both concentrations of protein. Two-state unfolding is observed for all concentrations of protein between 2 μM and 100 μM. The sensitivity of the unfolding transitions to changes in protein concentration is expected for unfolding of a native trimer (N) to unfolded monomers (U):

$$N \rightleftharpoons 3U \qquad (1)$$

The transitions at 5 μM and 30 μM can be fit globally to equation (1) (continuous line in Figure 2) by using:

$$K_{eq} = \frac{3f_U^3[M]_0^2}{1 - f_U} \qquad (2)$$

where $[M]_0$ indicates the total monomer concentration ($[M]_0 = [U] + 3[N]$), f_u is the fraction of

Table 1. Statistics of the foldon NMR structure

rmsd from experimental distance constraints (Å)	
All (607)[a]	0.046 ± 0.002
rmsd from NMR data	
NMR quality factor Q[b]	0.199 ± 0.0062
rmsd (Hz) between measured and calculated dipolar couplings (81)[c]	1.91 ± 0.07
Experimental dihedral constraints (deg.)[d] (43)	1.73 ± 0.29
$^3J_{HNHA}$ coupling constants (Hz) (22)	0.89 ± 0.05
Total number of restraints per monomer	753
Deviation from the idealized covalent geometry	
Bonds (Å)	0.0079 ± 0.0004
Angles (deg.)	0.92 ± 0.03
Impropers[e] (deg.)	0.75 ± 0.06
Coordinate precision[f] (Å)	
Backbone non-hydrogen atoms	0.273
All non-hydrogen atoms	0.636
Non-Gly, non-Pro residues in Ramachandran regions[g]	
Most favored (%)	91.7
Allowed (%)	8.3
Generously allowed (%)	0.0
Disallowed (%)	0.0

The statistics were obtained from a subset of the 40 best energy structures out of 100 following a standard simulated annealing protocol with dipolar restraints incorporated. Individual simulated annealing structures are fitted to each other using residues 2–27 of all subunits. The number of the various constraints per monomer is given in parentheses.
[a] Distance restraints comprise: 111 intraresidual NOEs; 139 sequential NOEs ($|i − j| = 1$); 73 short range NOEs ($1 < |i − j| ≤ 5$); 128 long-range NOEs ($|i − j| ≤ 5$); 156 intermolecular NOEs; 11 H-bonds (eight intramolecular, three intermolecular). For each backbone hydrogen bond constraint, there are two distance restraints: $r_{NH–O}$, 1.7–2.5 Å, $r_{N–O}$, 2.3–3.5 Å.
[b] The NMR quality factor Q is defined as the ratio of the rmsd between observed and calculated couplings and the rmsd of the observed couplings.[35]
[c] The 81 RDCs comprise 22 $^1D_{HN}$, 20 $^1D_{HαCα}$, 12 $^1D_{CαCβ}$, 13 $^1D_{NC'}$ (0.231), 14 $^1D_{CH3}$. Ramping the force constant for RDCs in the structure calculation from 0.001 kcal mol^{-1} Hz^{-2} to 0.5 kcal mol^{-1} Hz^{-2} was determined as optimal.
[d] The dihedral angle constraints comprise 69 φ and 60 ψ angles.
[e] The improper torsion restraints serve to maintain planarity and chirality.
[f] The coordinate precision is defined as the average rms difference between the individual simulated annealing structures and the mean coordinates. Values are reported for residues 2–27.
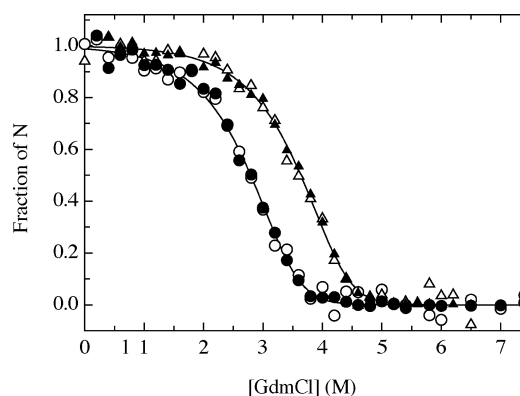[g] These values are calculated with the program PROCHECK-NMR.[36] Values are reported for all residues.

Figure 2. GdmCl-induced unfolding transition of the foldon domain at pH 7.1, 20 °C. Transitions at 5 μM (○,●) and 30 μM (▲,△) total monomer concentration ($[M]_0$) were measured by changes in Trp fluorescence (●,▲) and in far-UV CD at 228 nm (○,△). The data were normalized to fraction of native molecules using the result of a global fit of all data according to equations (14a) and (14b) (continuous lines).

unfolded monomer ($f_U = [U]/[M_0]$) and K_{eq} is the equilibrium constant (for details, see Materials and Methods). The global fit yields a free energy of unfolding of $\Delta G^0(H_2O) = 89.2(\pm 0.6)$ kJ/mol, which is unusually high compared to stabilities of small, single-domain proteins of similar size. However, this value applies to standard conditions of 1 M total monomer concentration. At typical physiological protein concentrations around 5 μM this corresponds to $\Delta G = 29.7$ kJ, which is similar to the stabilities found for small monomeric proteins. The change in free energy with GdmCl ($m_{eq} = \partial \Delta G^0 / \partial [\text{GdmCl}]$) is $-10.4(\pm 0.2)$ (kJ/mol)/ M, which is the value expected for a monomeric globular protein of the size of the folded trimer.[8] This shows that native foldon has properties comparable to those of small monomeric proteins with a compact hydrophobic core and a cooperative two-state unfolding transition.

Burst phase fluorescence changes

To investigate the folding kinetics of the foldon domain we performed stopped-flow refolding experiments starting from GdmCl-unfolded protein. Figure 3A shows a refolding trace at a residual denaturant concentration of 0.58 M and $[M]_0$ of 5 μM. The kinetics were monitored by the change in intrinsic tryptophan fluorescence above 320 nm. Within the first millisecond of refolding, a major burst phase reaction occurs, which leads to a significant increase in fluorescence intensity above the signals of both the unfolded and the native protein. This indicates very rapid structural changes in the dead-time of stopped-flow mixing (about 1 ms). The fluorescence intensity decreases slowly and reaches the value of the native protein after about 300 seconds. The burst phase increase in fluorescence is observed for all measured concentrations of protein (0.5 μM to 200 μM), indicating that the reaction occurs within the monomer. Even the fastest, diffusion-controlled association reaction to a partially folded dimer could not be complete within 1 ms at a monomer concentration of 1 μM and below, if we assume a maximum second-order rate constant[9] of about 1×10^9 M^{-1} s^{-1}. To further investigate the structural changes occurring in the burst phase, we monitored the folding kinetics at single wavelengths between 290 nm and 430 nm. Extrapolating the kinetic traces at the individual wavelengths to time zero allows the determination of the fluorescence spectrum of the burst phase intermediate (Figure 3B). Comparison of the fluorescence spectra of native and unfolded protein with the zero timepoint spectrum shows that the burst phase intermediate has a fluorescence emission maximum around 330 nm, which is between the emission maximum of the native protein ($\lambda_{max} = 317$ nm) and the unfolded state ($\lambda_{max} = 345$ nm). The significantly blue-shifted fluorescence maximum and the largely increased fluorescence intensity in the intermediate relative
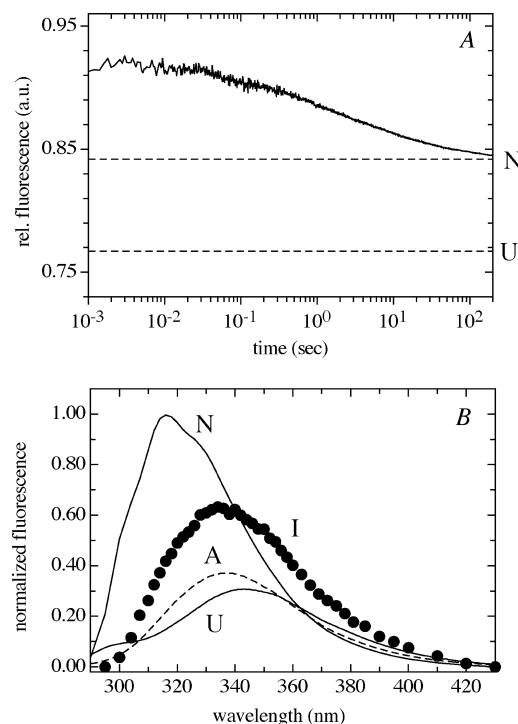

Figure 3. A, Refolding of the foldon domain in 0.58 M GdmCl, pH 7.1 ($[M]_0$ 5 μM) measured by the change in Trp fluorescence using a 320 nm emission cut-off filter. The broken lines represent the signals of the native and of the unfolded state at 0.58 M GdmCl, as indicated. The signal of the unfolded state is extrapolated from the unfolded baseline at high concentrations of GdmCl to 0.58 M GdmCl (see Materials and Methods). B, Comparison of the fluorescence spectrum of the kinetic burst phase intermediate (I) with the spectra of native (N) protein in 0.58 M GdmCl, the unfolded protein (U) in 8.2 M GdmCl and the monomeric A-state formed at pH 2 (A). The spectrum of I was determined in single-wavelength detection stopped-flow experiments. The fluorescence intensity extrapolated to $t = 0$ is shown. $[M]_0$ was 5 μM for all spectra.

to the unfolded state suggest that the burst phase intermediate has a significantly more hydrophobic environment around the single tryptophan residue at position 20 in each β-hairpin (see Figure 1). The absence of a tyrosine fluorescence band at 303 nm in the burst phase intermediate further indicates significant chain compaction, which allows efficient energy transfer from the two tyrosine residues at positions 2 and 13 to Trp20. Similar fluorescence properties are observed for an acid-induced monomeric state (A-state) of the foldon domain, which shows virtually the same fluorescence emission spectrum as the burst phase intermediate but with reduced fluorescence intensity (Figure 3).

Fast and slow steps during association of the foldon domain

To determine the nature of the rate-limiting steps

during folding and association of the foldon domain, we analyzed the concentration-dependence of the refolding kinetics at final concentrations of protein between 0.5 μM and 200 μM. Figure 4A shows that the kinetics are strongly concentration-dependent. A very slow reaction on the hundreds of seconds timescale is observed at low concentrations of protein. At higher concentrations of protein ($[M]_0 > 10$ μM) a concentration-dependent faster reaction with a half-time of about 4 ms at $[M]_0 = 102$ μM and a concentration-independent slow reaction ($\tau = 50$ s) are observed. The foldon domain contains two prolyl residues (Pro4 and Pro7) per subunit, which are in the *trans* conformation in the native state. This makes prolyl isomerization reactions a possible source for the slow, concentration-independent kinetics. Figure 5 shows that the slow reaction is catalyzed efficiently by human cyclophilin 18, a peptidyl-prolyl *cis–trans* isomerase,[10] which identifies this reaction as a *cis–trans* isomerization at one or both of the two Xaa-Pro peptide bonds per monomer. The faster
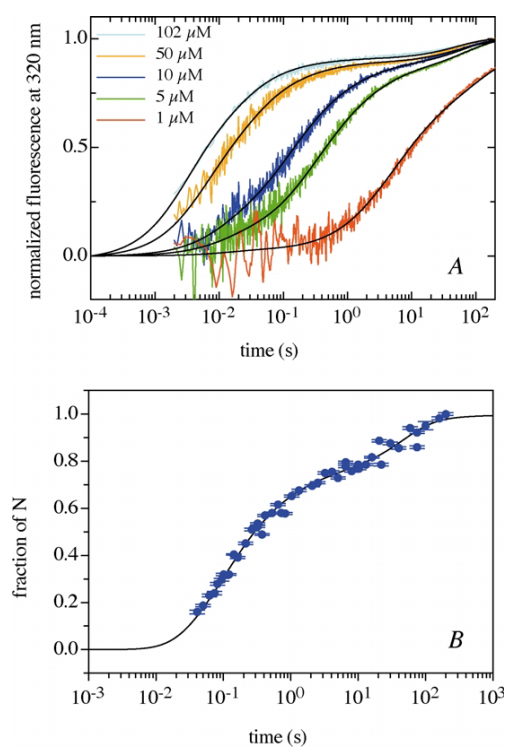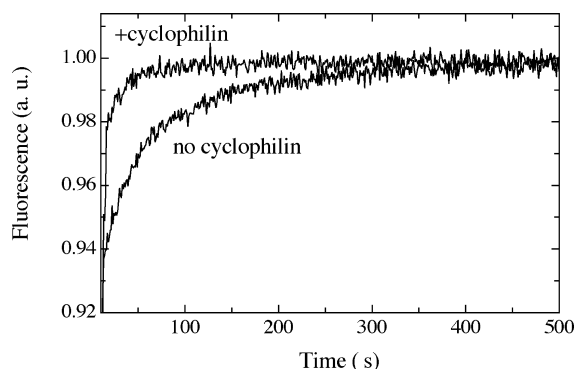

Figure 5. Effect of human cyclophilin 18 on the slow refolding reaction of the foldon domain. The plot compares refolding at pH 7.1, 0.58 M GdmCl, 20 °C in the absence and in the presence of 3.3 μM cyclophilin. The presence of cyclophilin increases the rate constant for the slow reaction from 7×10^{-3} s^{-1} to 4×10^{-2} s^{-1}.

reaction is not affected by the presence of cyclophilin (data not shown).

We tested whether the faster, concentration-dependent reaction produces native protein or a folding intermediate by performing interrupted refolding experiments.[11–13] In these experiments, the protein is allowed to refold for a certain time (t_i). The protein is then transferred to unfolding conditions and the resulting unfolding kinetics are monitored. The native state has a characteristic stability and barrier for unfolding, which results in a characteristic rate constant for its unfolding reaction. This distinguishes it from partially folded intermediates. Interrupted refolding experiments measure the increase in amplitude of the unfolding reaction of the native protein as a function of the refolding time, t_i. This corresponds to the time-course of formation of the native state. Thus, interrupted refolding experiments can distinguish whether a folding reaction produces native protein or a folding intermediate.[12,13] Figure 4B shows that native molecules are formed in a fast and a slow reaction, which occur on the same timescale as the two fluorescence-detected folding reactions (Figure 4B). Obviously, both the faster, concentration-dependent process and the prolyl-isomerization limited process produce native protein. This suggests that the fast reaction reflects formation of the native trimer for molecules with all prolyl peptide bonds in the native *trans* conformation (fast-folding molecules, U_F). The slow process is due to folding of molecules with at least one non-native *cis* isomer (slow-folding molecules, U_S), as shown by its catalysis by cyclophilin (Figure 5).

Mechanism of folding and association

The assignment of the different kinetic phases to direct folding and prolyl isomerization steps enables us to characterize the folding/association

Figure 4. A, Fluorescence detected folding kinetics in the presence of 0.58 M GdmCl (pH 7.1), 20 °C at the indicated values of $[M]_0$ measured after stopped-flow mixing at an emission wavelength of 320 nm. B, Time-course of formation of native molecules measured in interrupted refolding experiments at pH 7.1, 20 °C, $[M]_0 = 10$ μM. The continuous lines in both panels represent the results from a global fit of the data. For global fitting, ten fluorescence-detected refolding traces at $[M]_0$ between 0.5 μM and 102 μM and the time-course of formation of native molecules shown in B, were fitted simultaneously to the kinetic model shown in equation (4).
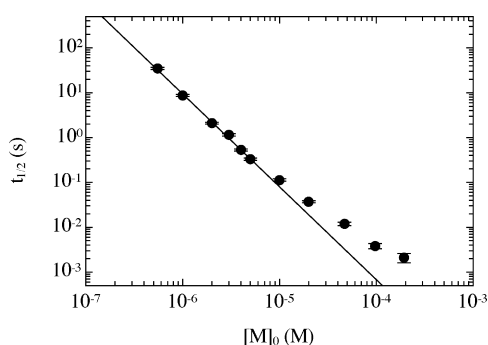
Figure 6. Effect of total monomer concentration ($[M]_0$) on the half-time of the fast-refolding reaction of the foldon domain at 0.58 M GdmCl (pH 7.1), 20 °C. The continuous line represents a fit of the data between 0.5 μM and 4 μM to equation (3). The fit gives a slope of -2.06 ± 0.05 indicating an apparent reaction order of 3 at low concentrations of protein.

process of the foldon domain in more detail. The analysis of the concentration-dependence of the half-lives ($t_{1/2}$) of the folding reaction starting from U_F allows us to determine the apparent reaction order for the fast-folding pathway according to:

$$\log t_{1/2} = c - (n - 1)\log [M]_0 \qquad (3)$$

where n is the reaction order and c is a reaction-specific constant.[14] Analysis of the concentration-dependence of folding thus yields information on contributions from concentration-dependent association reactions and from concentration-independent folding steps. We determined the half-time of the folding reaction from the concentration-dependence of the fluorescence-detected kinetics shown in Figure 4A. At high concentrations of protein, where folding and prolyl isomerization are well separated, we evaluated the half-lives of the fast reactions in order to obtain information on the direct folding reaction. Figure 6 shows that the slope of $\log t_{1/2}$ *versus* $\log[M]_0$

cating a change in the apparent reaction order. Between 100 μM and 200 μM initial monomer concentration there is only little effect of protein concentration on the folding kinetics. This shows that the fast-folding reaction approaches the first-order limit at high concentrations of protein.

The change from apparent third-order kinetics to first-order kinetics shows that folding is limited by association steps at low concentrations of protein and by a unimolecular folding reaction at high concentrations of protein. It is very unlikely that apparent third-order kinetics for a reaction in solution arise from a true trimolecular reaction.[15] Third-order reactions are usually caused by a rapid monomer–dimer pre-equilibrium followed by a second bimolecular association step to form the trimer. This mechanism gives rise to apparent third-order kinetics if the dimer is populated to only very low levels. The observation of apparent third-order kinetics at low concentrations of protein confirms the finding that the rapid collapse observed within the first millisecond of refolding is due to a conformational change in the monomer.

The weak concentration-dependence of the fluorescence-detected kinetics at high concentrations of protein implies that a unimolecular folding step becomes rate-limiting when all association reactions are fast. The half-time for formation of the native state of about 3–4 ms at high concentrations of protein indicates that the rate constant of the unimolecular step is around 200–300 s⁻¹. It is reasonable to assume that this unimolecular folding process is due to a structural rearrangement in the trimer and represents a late step in the formation of the native structure, and might be similar to the final steps during folding of a single-domain protein. However, we cannot exclude completely the possibility that the first-order reaction, which becomes rate-limiting at high concentrations of protein, occurs at the level of the burst phase intermediate or a partially folded dimer. These considerations lead to the minimal folding model for the foldon domain:
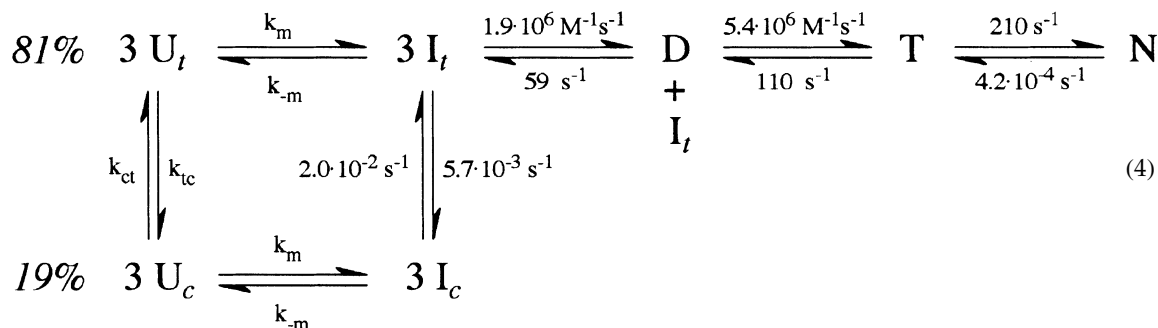
changes with protein concentration. At low concentrations of protein the slope is -2.06 ± 0.05, which corresponds to apparent third-order kinetics. At monomer concentrations above 5 μM the slope decreases significantly, indi-

In this mechanism, U denotes the completely unfolded monomer, I is the monomeric burst phase intermediate, D is a partially folded dimer, T is a partially folded trimer and N the native trimer. The subscripts c and t indicate monomers with a least

one *cis* Xaa-Pro peptide bond and monomers with all *trans* peptide bonds, respectively. k_m and k_{-m} represent rate constants for formation and unfolding of the burst phase intermediate, respectively, which occurs on the submillisecond timescale. To test whether this folding mechanism can describe all experimental data quantitatively, we fitted the folding kinetics at different concentrations of protein (Figure 4A) together with the time-course of formation of native foldon at 10 μM (Figure 4B) globally to the mechanism shown in equation (4). The results show that both the concentration-dependence of fluorescence kinetics and the time-course of formation of native molecules are well described by the model (continuous lines in Figure 4A and B). The global fit allows the determination of the rate constants for all reactions that occur after the submillisecond burst phase, since the apparent reaction order changes from third-order to almost first-order. The results of the global fit are given in equation (4). In agreement with the lower limit of the estimate from the limiting value of the half-lives for folding at high concentrations of protein (Figure 6) the fit gives a rate constant (k_f) of structural rearrangement in the trimer of $210(\pm 20)\,s^{-1}$. This is in the same range as the rate constants observed for the fastest folding monomeric proteins of similar size.[16] The bimolecular rate constants for formation of the dimer and of the trimer are $1.9(\pm 0.5) \times 10^6\,M^{-1}\,s^{-1}$ and $5.4(\pm 0.3) \times 10^6\,M^{-1}\,s^{-1}$, respectively, which is about 100 to 200-fold slower than the expected diffusion limit for bimolecular reactions of chains of this size.[9] However, both association reactions on the foldon trimerization pathway are significantly faster than most bimolecular steps during folding of dimeric proteins like the well-characterized GCN4[17] leucine zipper and of many globular dimeric proteins.[18] Rate constants similar to the two bimolecular steps during trimerization of foldon have been reported for wild-type arc repressor,[9] the fastest folding naturally occurring dimeric protein known, and for some designed leucine zippers.[19] However, an engineered arc repressor variant[20] and a designed fragment of trp repressor[21] show bimolecular rate constants of about $3 \times 10^8\,M^{-1}\,s^{-1}$, which are the fastest association reactions reported for folding of small dimeric proteins to date. In the case of the engineered arc repressor, the rate-enhancement relative to the wild-type protein was achieved by replacing the intermolecular salt-bridge/hydrogen bonding network by hydrophobic residues.[20] The fast-folding trp repressor fragment is stabilized mainly by intermolecular hydrophobic interactions. Similar to wild-type arc repressor, the foldon domain is stabilized by intermolecular hydrogen bonds and by an intermolecular salt-bridge, which contributes substantially (17 kJ/mol) to trimer stability (S.M. *et al.*, unpublished results). In this respect, it is interesting to note that the association steps in the foldon domain have virtually the same rate constants as the dimerization step of wild-type arc repressor.

The unfolding rate constants of foldon under native conditions (0.58 M GdmCl) obtained from the fit reveal that the major barrier for unfolding is represented by the reaction from the native trimer (N) to the partially folded trimer ($k = 4.2(\pm 0.5) \times 10^{-4}\,s^{-1}$). This value corresponds well to the rate constant for unfolding of native foldon measured at high concentrations of denaturant and extrapolated to 0.58 M GdmCl (S.Gü. & T.K., unpublished results). Unfolding of the partially folded trimer (T) and the dimer (D) are significantly faster with rate constants of $110(\pm 40)\,s^{-1}$ and $59(\pm 4)\,s^{-1}$, respectively, obtained from the fit.

In the folding mechanism shown in equation (4), we assumed that only the monomers with native *trans* prolyl isomers can form productive dimers and trimers. According to studies on model peptides the Ile3-Pro4 (12% *cis*) and Ala6-Pro7 (8% *cis*) peptide bonds in foldon should lead to 19% unfolded monomers with at least one *cis* prolyl peptide bond.[22] This agrees well with $20(\pm 1)$% of slow-folding molecules observed in the interrupted refolding experiments (equation (4) and Figure 4B), if we assume that only monomers with both prolyl peptide bonds *in trans* can enter the productive folding pathway. The presence of $20(\pm 1)$% slow-folding molecules would, however, be observed if only the Ala6–Pro7 peptide bond, which is in a highly structured region of foldon (Figure 1), was essential for folding and if a *cis* bond at this position could be incorporated into partially folded dimers and trimers. For this mechanism, the presence of 22% slow-folding molecules would be expected. Our data do not allow us to discriminate between these mechanisms. However, the rate constants for the fast-folding pathway are not influenced significantly by the folding mechanism of the slow-folding molecules.

The determination of all rate constants on the folding/association pathway of the foldon domain enables us to calculate the population of each species during folding at various concentrations of protein (Figure 7). At an initial concentration of 1 μM monomer, only the burst phase intermediate (I) and the native state become populated significantly (>10%) during folding. At concentrations of protein above 5 μM, the dimer becomes populated transiently, in agreement with a change in reaction order around this concentration of protein (Figure 6). As a consequence, the fast-folding pathway changes from apparent three-state to apparent four-state with the burst phase intermediate and the dimer populated to significant amounts. Above a concentration of 50 μM monomer, also the trimeric intermediate becomes populated significantly. The significant population of dimeric and trimeric intermediates explains the low apparent half-time of the reaction at high concentrations of protein, which would suggest a unimolecular folding reaction faster than the $210(\pm 20)\,s^{-1}$ obtained from the global fit. Since the apparent half-time was determined from the fluorescence measurements (Figure 6), it will be influenced by
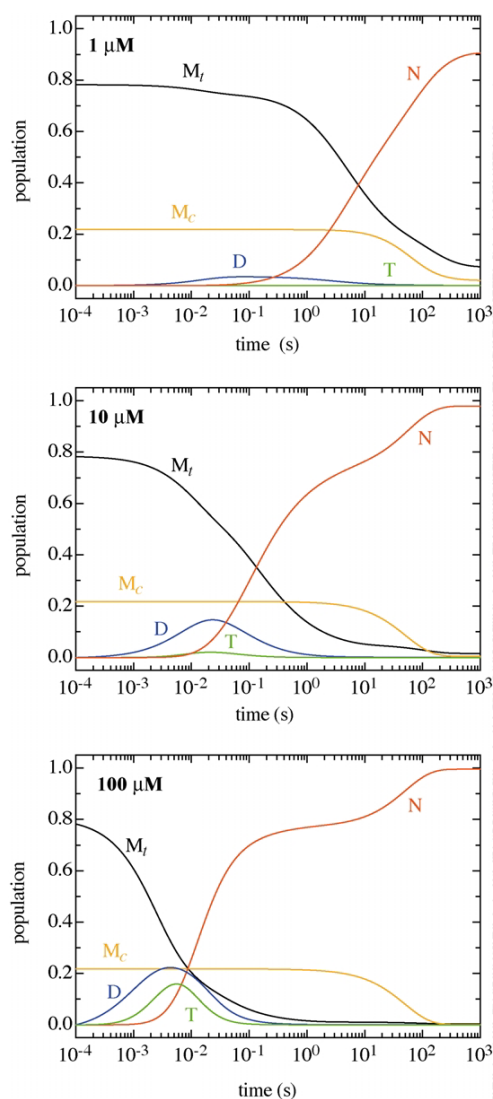
Figure 7. Simulation of the time-course of all kinetic species during folding of the foldon domain at 0.58 M GdmCl (pH 7.1), 20 °C. The results from the global fit of all kinetic data (see Figure 4) to the kinetic model shown in equation (4) were used to calculate the concentrations of the different species at the indicated values of $[M]_0$.

the kinetics of formation of dimers and trimers when these become populated significantly. These reactions are fast at high concentrations of protein, which results in apparently smaller and slightly concentration-dependent half-times for the folding reaction at high concentrations of protein. However, these half-times do not represent the true half-times for the first-order reactions.

During phage assembly in the cell, the estimated concentration of monomeric fibritin molecules is between 1 μM and 5 μM. Figure 7 shows that under these conditions neither the dimeric (D) nor the trimeric (T) intermediate becomes populated to significant amounts. At *in vivo* concentrations,

the half-time for folding is between one and ten seconds, which is very fast compared to the generation time of the phage of around 30 minutes.

Fast folding of multimeric proteins compared to monomeric proteins

The folding mechanism of the foldon domain has several interesting differences compared to fast-folding reactions of small monomeric proteins. Small monomeric proteins usually do not populate partially folded states to significant amounts during folding, although high-energy intermediates were shown to be obligatory for folding of many apparent two-state folders.[23,24] Obviously, small monomeric proteins are able to avoid transient high concentrations of intermediates, which minimizes the probability of aggregation side-reactions and optimizes the shape of the free energy barriers for rapid folding.[25] In the case of the foldon domain, however, rapid formation of intermolecular interactions is essential for efficient folding. Formation of the burst phase intermediate (I) leads (i) to pronounced fluorescence changes of the single Trp residue, which is part of the β-hairpin, (ii) to chain compaction as indicated by efficient energy transfer from the two tyrosine residues to the tryptophan and (iii) to significant changes in the far-UV CD signal (data not shown). This suggests that the monomeric intermediate involves structure formation of the β-hairpin, which is in agreement with the results of NMR studies on a monomeric state of the foldon domain (A-state) observed in equilibrium at low pH. In this state, the β-hairpin forms essentially the same structure as in the folded trimer (S.M. *et al.*, unpublished results). This shows that the monomeric foldon domain has a high propensity to form a β-hairpin even in the absence of intermolecular interactions. The similarity between the burst phase intermediate and the A-state is supported by the identical fluorescence emission maxima of the two states (Figure 3B). Rapid formation of the β-hairpins in the monomers probably facilitates the subsequent association reactions, since the hairpins provide highly specific surfaces that allow fast formation of intermolecular interactions. This model is in agreement with the folding mechanisms of the dimeric GCN4 fragment, for which rapid formation of α-helical structure was shown to accelerate the subsequent bimolecular association step.[26] Obviously, optimized energy landscapes for folding of monomeric and oligomeric proteins are different. Monomeric proteins fold fastest when the intermediates are marginally less stable than the unfolded state,[25] whereas populated intermediates promote the subsequent concentration-dependent association reactions in oligomeric proteins. The rapid formation of a partially folded state in the 27 residue foldon monomer shows that small, single-domain proteins should be able to rapidly form partially folded intermediates, provided this was beneficial for efficient folding.

Conclusions

Our folding studies on the foldon domain provide the first detailed information on the folding mechanism of a fast-folding trimeric protein. Previous studies on trimeric proteins focused mainly on large proteins, which fold very slowly and are accompanied by irreversible aggregation reactions[5] or on collagens, which are limited in folding by prolyl *cis–trans* isomerization reactions.[27] Our results suggest that in the small 27 amino acid residue foldon domain, all folding and association steps are optimized for rapid formation of a stable trimer. The two consecutive association reactions show bimolecular rate-constants similar to the values found in the fastest-folding small dimeric proteins. Further, the first-order folding reaction, which becomes rate limiting at high concentrations of protein, has a rate constant comparable to those of fast-folding, small, single-domain proteins. In contrast to previously studied small multimeric proteins or protein fragments, where association involved mainly helix–helix interactions, the foldon domain trimerizes *via* β-hairpins. The high bimolecular rate constants for both association steps in the foldon domain demonstrate that β-hairpins allow very rapid and specific formation of intermolecular interactions during protein assembly.

Materials and Methods

Protein synthesis and purification

The foldon domain was synthesized chemically with an ABIMED economy peptide synthesizer EPS 221 (Abimed, Germany) using standard F-moc chemistry. Pre-coupled resin was purchased from Novabiochem, Switzerland. Amino acids were from Alexis Biochemicals, USA or from Iris Biotech, Marktredwitz, Germany. Solvents and other chemicals were from Fluka (Buchs, Switzerland). The protein was purified using HPLC with a C-8 reverse-phase preparative column (Hibar, LiChrosorb®100 RP-8 from Merck, Darmstadt, Germany). ^{13}C,^{15}N-labeled foldon for NMR spectroscopy was expressed in *E. coli* BL21(DE3) cells grown in ^{13}C,^{15}N-enriched M9 minimal medium. Protein purity was confirmed by nanospray mass spectrometry and analytical HPLC. Expression and purification of the foldon domain in *E. coli* was performed as described.[4]

NMR spectroscopy

NMR samples ([U-^{15}N]foldon 95% H_2O/5% ^2H_2O; [U-^{13}C,^{15}N]foldon 95% H_2O/5% ^2H_2O; [U-^{13}C,^{15}N]foldon 100% ^2H_2O) of 300 μl volume (Shigemi NMR microtubes) were prepared as 0.3 mM protein solutions at pH 7.1 in 5 mM sodium phosphate, 0.02% (w/v) NaN$_3$ without any further addition of salt. Residual alignment of [U-^{13}C,^{15}N]foldon for the measurement of residual dipolar couplings was introduced by lamellar ether/ *n*-hexanol phases.[28] A set of standard triple and double-resonance assignment, quantitative *J*-coupling, nuclear Overhauser effect (NOE) spectroscopy (NOESY) and ^{15}N relaxation experiments similar to those described[29] were

performed on a Bruker DRX 600 spectrometer at 25 °C. Standard data processing and analysis was carried out as described.[29]

Structure calculation

Experimental NOE distance, torsion angle and residual dipolar restraints derived from the NMR data are listed in Table 1. Structure calculations were carried out with a simulated annealing protocol using the program CNS.[30] The structural statistics for the best 40 structures are given in Table 1.

Protein Data Bank accession code

The structural statistics have been deposited in the Brookhaven Protein Data Bank with PDB accession code 1RFO.

Denaturant-induced equilibrium transitions

Denaturant-induced equilibrium transitions were recorded in an AMINCO Bowman series 2 spectrofluorimeter (SLM Aminco, USA) and with an Aviv 62ADS spectropolarimeter (Aviv, USA). All transitions were measured at 10 mM sodium phosphate (pH 7.1 at 20 °C). For fluorescence measurements at concentrations of total monomer ([M]$_0$) of 5 μM and 30 μM, the excitation wavelengths were 278 nm and 298 nm, respectively, at 2 nm bandwidth. Emission was recorded at 320 nm (2 nm bandwidth). CD measurements were performed at 228 nm with 0.5 cm (5 μM) or 0.1 cm (30 μM) path-lengths.

Equilibrium transition curves were analyzed assuming a two-state transition from native trimer (N) to unfolded monomer (U):

$$N \rightleftharpoons 3U \tag{5}$$

The resulting equilibrium constant is given by:

$$K_{eq} = \frac{[U]^3}{[N]} \tag{6}$$

and the total monomer concentration ([M]$_0$) can be expressed as:

$$[M]_0 = [U] + 3[N] \tag{7}$$

The fractions of monomers in the unfolded state (f_U) and in the native state (f_N) are given by:

$$f_U = \frac{[U]}{[U] + 3[N]} = \frac{[U]}{[M]_0} \tag{8a}$$

$$f_N = \frac{3[N]}{[U] + 3[N]} = \frac{3[N]}{[M]_0} = 1 - f_U \tag{8b}$$

Thus, equation (6) becomes:

$$K_{eq} = \frac{[U]^3}{[N]} = \frac{3f_u^3[M]_0^2}{1 - f_u} \tag{9}$$

To fit the equilibrium transitions to equation (9) we expressed the fraction of unfolded protein at a given denaturant concentration (x) as:[31]

$$f_U(x) = \frac{S_N(x) - S(x)}{(S_N(x) - S(x)) + (S(x) - S_U(x))} \tag{10}$$

where $S(x)$ corresponds to the measured signal at the given denaturant concentration, x. S_N and S_U represent the spectroscopic signals of the native and unfolded state, respectively, obtained from linear extrapolation of the baselines according to:[32]

$$S_N(x) = S_N(H_2O) + m_N x \tag{11a}$$

$$S_U(x) = S_U(H_2O) + m_U x \quad (11b)$$

Using:

$$\Delta G^0 = -RT \ln K_{eq} \quad (12)$$

and a linear denaturant-dependence of ΔG^0 :[32,33]

$$\Delta G^0(x) = \Delta G^0(H_2O) + mx \quad (13)$$

we can obtain the free energy for unfolding in water $\Delta G^0(H_2O)$ from fitting the transition curve to equations (9)–(13) in a single step, similar to the procedure described by Santoro & Bolen for a two-state transition of monomeric proteins.[34] This yields the following equation for a two-state unfolding transition from a native trimer to unfolded monomers:

$$S(x) = S_N(x) - \frac{S_N(x) - S_U(x)}{3[M]_0}$$

$$\times \left(\sqrt[3]{e^{\frac{-\Delta G^0(H_2O)+mx}{RT}} \left(\frac{9}{2}[M]_0 + \sqrt{D} \right)} \right.$$

$$\left. + \sqrt[3]{e^{\frac{-\Delta G^0(H_2O)+mx}{RT}} \left(\frac{9}{2}[M]_0 - \sqrt{D} \right)} \right) \quad (14a)$$

with:

$$D = \left(\frac{9}{2} \right)^2 [M]_0^2 + e^{\frac{-\Delta G^0(H_2O)+mx}{RT}} \quad (14b)$$

Equations (14a) and (14b) were used to fit the individual transition curves monitored by fluorescence or CD and to globally fit all data obtained by measuring various spectroscopic probes at different concentrations of total monomer ($[M]_0$).

Kinetic experiments

All stopped-flow experiments were performed with a SX18.MV stopped-flow instrument from Applied Photophysics (Leatherhead, UK) equipped with a Hamamatsu R1104 photomultiplier tube for single-wavelength kinetics and a Hamamatsu R6095 photomultiplier tube for all other measurements. For experiments using a cut-off filter (≥ 320 nm), the excitation wavelength was 278 nm (2 nm bandwidth). For single-wavelength refolding experiments, the excitation wavelength was 280 nm (4.5 nm bandwidth) for final concentrations of protein up to 10 µM and 295 nm (4.5 nm bandwidth) for higher concentrations. The emission bandwidth was 12 nm. For refolding experiments, the protein was allowed to unfold for at least ten hours in 6.38 M GdmCl, 10 mM sodium phosphate (pH 7.1) before the measurements. At final concentrations of protein above 10 µM monomer, the protein was unfolded at the same concentration of denaturant at pH 2. Unless stated otherwise, all measurements were performed in 10 mM sodium phosphate (pH 7.1 at 20 °C). Refolding was initiated by 11-fold dilution to a final concentration of GdmCl of 0.58 M. The fluorescence intensity of the unfolded state under refolding conditions (Figure 3) was determined by extrapolating the fluorescence signal of the unfolded foldon domain (measured in the stopped-flow instrument) to 0.58 M GdmCl. The effect of human cyclophilin 18 on the slow-refolding reaction was measured by manual mixing fluorescence measurements in 10 mM sodium cacodylate (pH 7.1) at $[M]_0 = 5$ µM foldon domain in the absence and in the presence of 3.3 µM

cyclophilin. The use of sodium cacodylate instead of sodium phosphate had no effect on the folding kinetics.

Stopped-flow interrupted refolding experiments were used to monitor the formation of native molecules during refolding in 0.58 M GdmCl, 20 mM sodium phosphate (pH 7.0) at 20.0 °C. Completely unfolded foldon (in 3.4 M GdmCl, 20 mM sodium phosphate (pH 1.7), $[M]_0 = 60$ µM) was diluted sixfold into final conditions of 10 µM protein, 0.58 M GdmCl, 20 mM sodium phosphate (pH 7.0) to initiate refolding. After various times (t_i), refolding was interrupted by transferring the solution into final conditions of 6.7 M GdmCl, 20 mM sodium phosphate (pH 7.0), final protein concentration 1.7 µM. Native foldon unfolds with double-exponential kinetics under these conditions (S.Gü. & T.K., unpublished results) with $\tau_1 = 7.8(\pm 0.5)$ seconds (80% amplitude) and $\tau_2 = 0.24(\pm 0.02)$ seconds (20% amplitude). The relative amplitudes of the two reactions is independent of the refolding time. The amplitude of the major, slow-unfolding reaction ($\tau_1 = 7.8(\pm 0.5)$ seconds) was used as a measure for the amount of native protein that was present after the time t_i, when refolding was interrupted. The observed unfolding amplitudes after various times of refolding were normalized against the amplitude of completely refolded foldon to yield the fraction of native molecules that were present after t_i.

Data fitting and simulations

Data evaluation was carried out using the programs ProFit (Quantumsoft, Zurich, Switzerland) and Matlab (The MathWorks, Natick, MA, USA). Interrupted refolding experiments at $[M]_0 = 10$ µM and ten direct fluorescence-detected refolding traces with $[M]_0$ ranging from 0.5 µM to 102 µM were analyzed globally by nonlinear, least-squares curve fitting. The experimental data were fit to the numerical solution of the kinetic scheme depicted in equation (4). Rate constants and relative signal amplitudes of the different kinetic species were fitted as global parameters. The equilibrium constant between native and unfolded protein determined by a global fit of the equilibrium unfolding transitions measured by fluorescence and CD at various concentrations of protein (Figure 2) was used as an additional constraint for the fit. To ensure that the fit converged to the global minimum, it was repeated 60 times with randomly chosen starting values for the fitting parameters.

References

1. Tao, Y., Strelkov, S. V., Mesyanzhinov, V. V. & Rossmann, M. G. (1997). Structure of bacteriophage T4 fibritin: a segmented coiled coil and the role of the C-terminal domain. *Structure*, **5**, 789–798.
2. Letarov, A. V., Londer, Y. Y., Boudko, S. P. & Mesyanzhinov, V. V. (1999). The carboxy-terminal domain initiates trimerization of bacteriophage T4 fibritin. *Biochemistry (Moscow)*, **64**, 817–823.
3. Boudko, S. P., Londer, Y. Y., Letarov, A. V., Sernova, N. V., Engel, J. & Mesyanzhinov, V. V. (2002). Domain organization, folding and stability of bacteriophage T4 fibritin, a segmented coiled-coil protein. *Eur. J. Biochem.* **269**, 833–841.
4. Frank, S., Kammerer, R. A., Mechling, D., Schulthess, T., Landwehr, R. J. B., Guo, Y. *et al.* (2001). Stabilization of short collagen-like triple helices by protein engineering. *J. Mol. Biol.* **308**, 1081–1089.
5. Seckler, R. (2000). Assembly of multi-subunit

structures. In *Mechanisms of Protein Folding* (Pain, R., ed.), pp. 279–308, Oxford University Press, Oxford.

6. Stetefeld, J., Frank, S., Jenny, M., Schulthess, T., Kammerer, R. A., Boudko, S. P. *et al.* (2003). Collagen stabilization at atomic level: crystal structure of designed (GlyProPro)$_{10}$-foldon. *Structure*, **11**, 339–346.

7. Dosset, P., Hus, J. C., Blackledge, M. & Marion, D. (2000). Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J. Biomol. NMR*, **16**, 23–28.

8. 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.

9. Milla, M. E. & Sauer, R. T. (1994). P22 arc repressor: folding kinetics of a single-domain, dimeric protein. *Biochemistry*, **33**, 1125–1133.

10. Fischer, G., Wittmann, L. B., Lang, K., Kiefhaber, T. & Schmid, F. X. (1989). Cyclophilin and peptidyl-prolyl *cis–trans* isomerase are probably identical proteins. *Nature*, **337**, 476–478.

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

12. Kiefhaber, T. (1995). Protein folding kinetics. In *Methods in Molecular Biology* (Shirley, B. A., ed.), vol. 40, pp. 313–341, Humana Press, Totowa, NJ.

13. Bieri, O. & Kiefhaber, T. (2000). Kinetic models in protein folding. In *Protein Folding: Frontiers in Molecular Biology* (Pain, R., ed.), 2nd edit., pp. 34–64, Oxford University Press, Oxford.

14. Moore, J. W. & Pearson, R. G. (1981). *Kinetics and Mechanisms*, Wiley, New York.

15. Gutfreund, H. (1995). *Kinetics for the Life Sciences*, Cambridge University Press, Cambridge.

16. Jackson, S. E. (1998). How do small single-domain proteins fold? *Fold. Des.* **3**, R81–R91.

17. Zitzewitz, J. A., Bilsel, O., Luo, J., Jones, B. E. & Matthews, C. R. (1995). Probing the folding mechanism of a leucine zipper peptide by stopped-flow circular dichroism spectroscopy. *Biochemistry*, **34**, 12812–12819.

18. Jaenicke, R. (1987). Folding and association of proteins. *Prog. Biophys. Mol. Biol.* **49**, 117–237.

19. Wendt, H., Berger, C., Baici, A., Thomas, R. M. & Bosshard, H. R. (1995). Kinetics of folding of leucine zipper domains. *Biochemistry*, **34**, 4097–4107.

20. Waldburger, C. D., Jonsson, T. & Sauer, R. T. (1996). Barriers to protein folding: formation of buried polar interactions is a slow step in acquisition of structure. *Proc. Natl Acad. Sci. USA*, **93**, 2629–2634.

21. Gloss, L. M. & Matthews, C. R. (1998). Mechanism of folding of the dimeric core domain of *Escherichia coli* Trp repressor: a nearly diffusion-limited reaction leads to the formation of an on-pathway intermediate. *Biochemistry*, **37**, 15990–15999.

22. Reimer, U., Scherer, G., Drewello, M., Kruber, S., Schutkowski, M. & Fischer, G. (1998). Side-chain effects on peptidyl-prolyl *cis/trans* isomerization. *J. Mol. Biol.* **279**, 449–460.

23. Bachmann, A. & Kiefhaber, T. (2001). Apparent two-state tendamistat folding is a sequential process along a defined route. *J. Mol. Biol.* **306**, 375–386.

24. Sánchez, I. E. & Kiefhaber, T. (2003). Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J. Mol. Biol.* **325**, 367–376.

25. Wagner, C. & Kiefhaber, T. (1999). Intermediates can accelerate protein folding. *Proc. Natl Acad. Sci. USA*, **96**, 6716–6721.

26. Zitzewitz, J. A., Ibarra-Molero, B., Fishel, D. R., Terry, K. L. & Matthews, C. R. (2000). Preformed secondary structure drives the association reaction of GCN4-p1, a model coiled-coil system. *J. Mol. Biol.* **296**, 1105–1116.

27. Boudko, S. P., Frank, S., Kammerer, R. A., Stetefeld, J., Schulthess, T., Landwehr, R. *et al.* (2002). Nucleation and propagation of the collagen triple helix in single-chain and trimerized peptides: transition from third to first order kinetics. *J. Mol. Biol.* **317**, 459–470.

28. Ruckert, M. & Otting, G. (2000). Alignment of biological macromolecules in novel non-ionic liquid crystalline media for NMR experiments. *J. Am. Chem. Soc.* **122**, 7793–7797.

29. Kahmann, J. D., Sass, H. J., Allan, M. G., Seto, H., Thompson, C. J. & Grzesiek, S. (2003). Structural basis for antibiotic recognition by the TipA class of multidrug-resistance transcriptional regulators. *EMBO J.* **22**, 1824–1834.

30. Brünger, 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 Crystallog. sect. D*, **54**, 905–921.

31. Tanford, C. (1968). Protein denaturation. Part B. The transition from native to denatured state. *Advan. Protein Chem.* **23**, 218–282.

32. Pace, C. N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* **131**, 266–280.

33. Greene, R. F. J. & Pace, C. N. (1974). Urea and guanidine–hydrochloride denaturation of ribonuclease, lysozyme, alpha-chyomtrypsinn and beta-lactoglobulin. *J. Biol. Chem.* **249**, 5388–5393.

34. 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**, 8063–8068.

35. Cornilescu, G., Marquardt, J. L., Ottiger, M. & Bax, A. (1998). Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J. Am. Chem. Soc.*, 6836–6837.

36. 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.

37. Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55.

9. Appendix

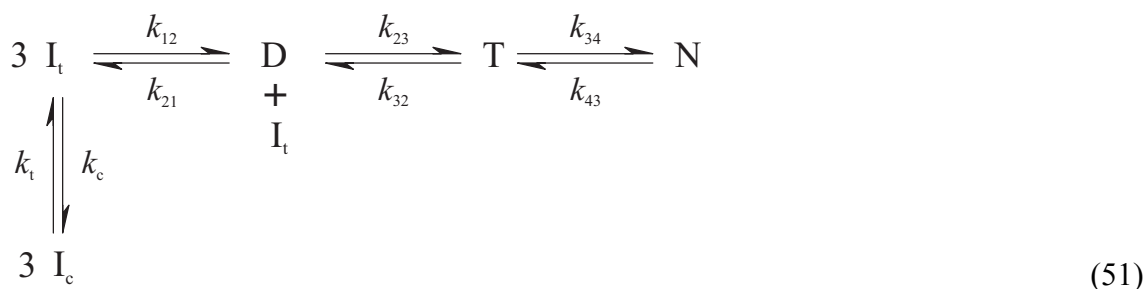## 9.1 – Analysis of the Refolding Kinetics of the Foldon Domain

Refolding kinetics of the trimerization domain of phage T4 fibritin, shortly foldon, were analyzed with the computer program MATLAB[209] (cp. chapter 3.5). In the following the most salient features of the data analysis procedure will be discussed.

Even comparatively simple kinetic reaction schemes involving bimolecular (or higher order) reaction steps possess no analytical solution. However, a reaction scheme such as the one applicable to foldon refolding, depicted in eq. (51), defines a set of (ordinary) differential equations which describe the time behaviour of the system.

$$3\ I_t \underset{k_{21}}{\overset{k_{12}}{\rightleftharpoons}} \underset{\substack{+\\ I_t}}{D} \underset{k_{32}}{\overset{k_{23}}{\rightleftharpoons}} T \underset{k_{43}}{\overset{k_{34}}{\rightleftharpoons}} N$$

$$k_t \Big\Updownarrow k_c$$

$$3\ I_c \tag{51}$$

Starting from a suitable set of initial conditions, the scheme can be solved numerically and the concentrations of the individual reactants are obtained as functions of time. MATLAB supplies several functions to accomplish this. Different ordinary differential equation (ODE) solvers are available but best results in terms of accuracy and performance were obtained with the ODE15S function. The code fragment below illustrates the use of this function. Parameters M0, D0, T0, N0 and CM0 denote the concentrations of the species I_t, D, T, N and I_c at time zero. The differential equations are defined in the function 'foldfun'.

```matlab
options=odeset('RelTol',1e-6,'AbsTol',[1e-10 1e-10 1e-10 1e-10 1e-10]);
sol=ode15s(@foldfun,t,[M0 D0 T0 N0 CM0],options,k12,k21,k23,k32,k34,
    kc,kt,k43);

% ------------------------------------

function dy=foldfun(t,y,k12,k21,k23,k32,k34,kc,kt,k43)

% define differential equations

dy=ones(5,1);
dy(1)=-2*k12*y(1)^2+2*k21*y(2)-k23*y(1)*y(2)+k32*y(3)+kt*y(5)-kc*y(1);
```

```
dy(2)=k12*y(1)^2+k32*y(3)-k21*y(2)-k23*y(1)*y(2);
dy(3)=k23*y(1)*y(2)-k34*y(3)-k32*y(3)+k43*y(4);
dy(4)=k34*y(3)-k43*y(4);
dy(5)=-kt*y(5)+kc*y(1);

% 1 -> Itrans; 2 -> D; 3 -> T; 4 -> N; 5 -> Icis
```

By fitting the experimental data to the numerical solution of the ODE system, the microscopic rate constants of the reaction system can be determined (cp. eq. (51)). A single refolding trace at one given protein concentration is not sufficient to resolve the rate constants for the forward (k_{12}, k_{23}, k_{34}) and backward (k_{21}, k_{32}, k_{43}) reactions. If, however, refolding data acquired for different total protein concentrations are analyzed globally, the rate constants for the first two association reactions can be reliably determined, since only the forward reactions are concentration-dependent. The rate constants of the last unimolecular reaction, k_{34} and k_{43}, can be resolved via the global stability of the protein. Rate constants for the *cis/trans* isomerization reaction, k_c and k_t, are defined by the initial concentrations of *cis* and *trans* isomer, $[I_c]_0$ and $[I_t]_0$, according to eq. (52).

$$[I_t]_0 / [I_c]_0 = k_t / k_c \tag{52}$$

Data fitting was carried out in MATLAB using the non-linear least-squares optimization algorithm LSQNONLIN as shown below.

```
options=optimset('LargeScale','on','LevenbergMarquardt','on','MaxIter',
    200,'Diagnostics','on','DiffMinChange',1e-7,'DiffMaxChange',1e0,
    'Display','iter','MaxFunEvals',40000,'TolFun',1e-9,'TolX',1e-9);
[fitpa,resnorm,residuals,flag,output,lambda,jacobian]=lsqnonlin(@fit_Ki
    netix_DG,initpa,lb,ub,options,tquery,sig,weight,Ctot,noel,nont,fitmo
    de,DG);
```

The function 'fit_Kinetix_DG' calculates the solution to the kinetic scheme using ODE15S as explained above. Solving the ODE system is computationally quite demanding and, accordingly, the fitting process can take a long time. A significant gain in execution speed is achieved by using static (persistent) variables to save intermediate results as illustrated below. The only fit parameters that affect the solution of the ODE system are the microscopic rate constants. If between function calls their values do not change, the ODE solution does not need to be calculated anew. Instead one can save the solution from the previous function call.

```matlab
function target=fit_Kinetix_DG(param,tquery,sig,weight,Ctot,noel,nont,
    fitmode,DG)

persistent save_param;      % to speed things up a little - save params
persistent save_result;     % save previous function values
persistent not_first_run;   % flag: first execution of this function?

% (...)

if (not_first_run)
    % check whether any rate constant changed cp.'d to last func. call

    for j=(2*nof+1-2*nont):(2*nof+11-2*nont)
        % loop over all parameters that define rate constants
        if (param(j)~=save_param(j))
            recal=1;
        end
    end
else
    not_first_run=1;
    recal=1;
end

if (recal)
    % have to recalculate solution of ODEs

    % (...)

else
    % can use previous solution
    sol=save_result;
end

% (...)

save_result=sol;
save_param=param;

% (...)
```

The array 'save_param' stores the parameter values from the previous function call. When one of the rate constants was changed, the flag variable 'recal' is set to 'true' and the ODE system has to be solved anew. If none of the rate constants changed, the solution of the ODE system can be just copied out of the array 'save_result' from the last function call.

10. Curriculum Vitae

Personal details

Name and address: Andreas Möglich

                           Biozentrum der Universität Basel

                           Abteilung Biophysikalische Chemie

                           Klingelbergstrasse 70

                           CH-4056 Basel

                           Tel.: ++ 41 (0)61 267 21 84

                           Fax: ++ 41 (0)61 267 21 89

                           Email: andreas.moeglich@unibas.ch

Education

03/2002 – 11/2005   Biozentrum der Universität Basel, Switzerland

                           Graduate (Ph. D.) studies in Biophysics

                           Thesis supervisor: Prof. Dr. Thomas Kiefhaber

                           Thesis title: "Structure and Dynamics of Unfolded Polypeptide

                           Chains"

04/2001 – 12/2001   Universität Regensburg, Germany

                           Diploma thesis in December 2001

                           Diploma supervisor: Prof. Dr. Eike Brunner

                           Thesis title: "Strukturbestimmung der Ile14→Ala-Austausch-

                           mutante des Histidine-Containing Phosphocarrier Protein aus

                           Staphylococcus carnosus unter Verwendung von

                           Restdipolkopplungen"

10/1996 – 01/2002	Universität Regensburg, Germany
	Diploma studies in biochemistry
08/1999 – 06/2000	Washington University St. Louis, MO USA
	Exchange studies
	Research supervisor: Prof. Dr. Timothy M. Lohman
09/1986 – 06/1995	Kaiserin-Friedrich-Gymnasium, Bad Homburg v. d. H., Germany
	Abitur in 1995