

TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Biophysikalische Chemie

Intramolecular and intermolecular diffusion processes  
in protein folding and assembly

Karin N. Joder

Vollständiger Abdruck der von der Fakultät für Chemie  
der Technischen Universität München zur Erlangung des akademischen Grades eines  
Doktors der Naturwissenschaften  
genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. Dr.h.c. B. Rieger

Prüfer der Dissertation:

1. Univ.-Prof. Dr. Th. Kiefhaber
2. Univ.-Prof. Dr. M. Groll
3. Univ.-Prof. Dr. Chr. F. W. Becker

Die Dissertation wurde am 13.04.2011 bei der Technischen Universität München  
eingereicht und durch die Fakultät für Chemie  
am 27.06.2011 angenommen.

## Table of Contents

<b>1. INTRODUCTION.....</b>	<b>1</b>
1.1 Proteins in biology.....	1
1.2 Protein folding.....	2
1.2.1 Protein stability .....	2
1.2.2 Intermediates and barriers in protein folding.....	4
1.3 Early steps in protein folding.....	7
1.3.1 Polymer chain models.....	8
1.3.2 Triplet-triplet energy transfer.....	10
1.3.3 Chain dynamics in unfolded polypeptides.....	12
1.3.4 Aggregation prone polyglutamine sequences.....	15
1.4 Protein-protein interactions.....	16
1.4.1 Analytical models for bimolecular reactions.....	16
1.4.2 Site-specific protein-protein interactions.....	18
1.4.2.1 Folding and assembly of dimeric proteins.....	18
1.4.2.2 Folding and assembly of trimeric proteins.....	19
1.4.2.3 Electrostatic enhancement of site-specific protein-protein interactions.....	21
1.4.2.4 Theoretical models for site-specific protein-protein interactions...	22
1.5 Effect of cellular crowding on intra- and intermolecular diffusion processes.....	25
<b>2. AIMS OF RESEARCH.....</b>	<b>29</b>
2.1 Intramolecular diffusion processes during protein folding.....	29
2.2 Intermolecular diffusion processes during protein folding and assembly.....	30

<b>3. RESULTS.....</b>	<b>31</b>
3.1 Dynamics of intramolecular loop formation in polypeptide chains.....	31
3.1.1 <i>Effect of solvent viscosity and temperature on intrachain diffusion processes.....</i>	31
3.1.1.1 <i>Effect of viscous co-solvents on intrachain dynamics in model polypeptide chains and natural protein sequences.....</i>	32
3.1.1.2 <i>Characterization of the barriers for intrachain loop formation in model polypeptide chains and natural protein sequences.....</i>	49
3.1.2 <i>Dynamics of loop formation in a fragment from the brinker protein.....</i>	59
3.1.3 <i>Dynamics of loop formation in aggregation prone polyglutamine peptides.....</i>	72
3.1.4 <i>End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains studied by FRET.....</i>	86
3.2 Dynamics of site-specific intermolecular interactions.....	90
3.2.1 <i>Intermolecular interactions between unfolded polypeptide chains.....</i>	90
3.2.2 <i>Role of electrostatic interactions in the assembly of the foldon domain..</i>	103
3.2.3 <i>The speed-limit for site-specific formation of an encounter complex during assembly of the foldon domain.....</i>	120
3.2.4 <i>Analytical models for calculation of the reactive surface on a monomeric foldon subunit.....</i>	128
3.2.5 <i>Elimination of association reactions during folding of the trimeric foldon domain.....</i>	134
3.3 Materials and methods.....	144
<b>4. SUMMARY.....</b>	<b>151</b>
<b>5. ACKNOWLEDGEMENTS.....</b>	<b>156</b>
<b>6. REFERENCES.....</b>	<b>157</b>

**7. APPENDIX..... 177**

7.1 End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation.... 177

# 1. Introduction

## 1.1. Proteins in biology

Proteins constitute the major class of macromolecules in living organisms and play an essential role in nearly every biological process. Their function is very diverse including signal processing and transduction, mechanical support of cells, catalysis of specific chemical reactions as well as transport and storage of nutrients and metabolites. Proteins are linear polymer chains made up of twenty different  $\alpha$ -L-amino acids connected by amide linkages. The amino acid sequence of a protein is determined by the sequence of nucleic acids in the encoding gene, which is translated at the ribosome during protein biosynthesis. In higher organisms, proteins are often modified after synthesis by the attachment of sugars, lipids or other chemical groups. Under physiological conditions most naturally occurring proteins adopt a specific three-dimensional structure (native conformation), which is usually an absolute prerequisite for proper function. The transition from an unstructured polypeptide chain to a highly ordered molecule with a well-defined three-dimensional structure is referred to as protein folding. Proteins that do not fold properly and form aggregates were found to be the cause for several diseases including the Alzheimer disease, the Huntington disease and prion diseases like Creutzfeldt-Jacob disease, Scrapie and Bovine Spongiforme Encephalopathy.

The linear amino acid sequence of a polypeptide chain includes the entire information for the chain to fold to its native three-dimensional structure in solution.<sup>1</sup> This allows the study of the protein folding process *in vitro* in the absence of the cellular environment. The question of how the linear amino acid sequence is able to find its native structure is a central question in biophysics. A final goal is to predict the native structure from the linear amino acid sequence what would allow the design of proteins with new structures and properties.

## 1.2. Protein folding

Starting from a large ensemble of unstructured conformations termed the unfolded or denatured state (U), the polypeptide chain has to search for energetically favorable interactions to form secondary structured elements like loops, helices or  $\beta$ -sheets. In the native state (N) secondary elements are arranged in a clear defined three-dimensional structure (tertiary structure). The unfolded and folded states of a protein are in equilibrium, thus folding and unfolding reactions occur constantly in a protein population.

### 1.2.1 Protein stability

Folding of many small proteins can be described as a two-state process with no partially folded intermediates populated to a significant amount.



Here,  $k_f$  and  $k_u$  are the rate constants for folding and unfolding, respectively. Taking the unfolded state as a reference the stability of the native protein can be expressed by the difference in free energy  $\Delta G^0$  between the folded and the unfolded state according to

$$\Delta G^0 = -RT \ln K_{eq} \quad \text{with} \quad K_{eq} = \frac{[N]_{eq}}{[U]_{eq}} = \frac{k_f}{k_u} \quad (2)$$

$K_{eq}$  is the equilibrium constant and  $[N]_{eq}$  and  $[U]_{eq}$  are the concentrations of native and unfolded molecules in equilibrium. The stability of a protein is usually small with a free energy for folding ( $\Delta G^0$ ) in the range of -10 to -40 kJ/mol. The small stability is the result of several large, but compensating, contributions. Upon folding intramolecular van-der-Waals interactions, H-bonds and salt bridges are established, which is energetically favorable. On the other hand the protein loses interactions, mainly hydrogen bonds with the solvent, which is energetically destabilizing. Similar compensating effects are observed for the reaction entropy. The native state is destabilized by a decrease in chain entropy while it is stabilized by an increase in entropy due to the release of water molecules that were bound to the unfolded chain.

The native state was supposed to be the thermodynamic most stable state on the energy landscape. But today several cases are known where this assumption is not true. Examples are proteins that are natively unstructured and only fold in the presence of their binding partner (DNA or another protein) and misfolded proteins involved in prion diseases.

The equilibrium between the native and the unfolded state can be shifted by a change in external conditions. The Gibbs fundamental equation (Equation 3) shows that the equilibrium can be perturbed by pressure, temperature or changes in chemical potential, e.g. upon addition of chemical denaturants like urea or guanidinium hydrochloride (GdmCl).

$$d\Delta G^0 = \Delta V^0 dp - \Delta S^0 dT + \sum \Delta \mu_i^0 dn_i \quad (3)$$

$\Delta V^0$ ,  $\Delta S^0$  and  $\Delta \mu_i^0$  are the changes in volume, entropy and chemical potential upon folding. Including the effect of a denaturant on protein folding, equation 3 can be written as

$$d\Delta G^0 = \Delta V^0 dp - \Delta S^0 dT + m d[D] \quad (4)$$

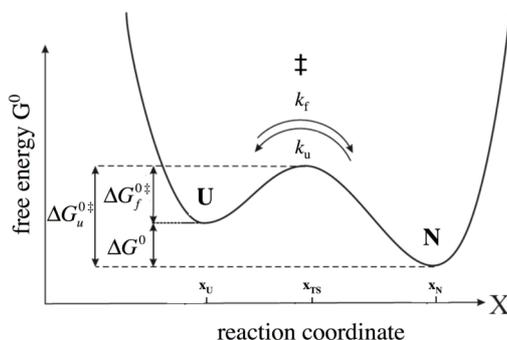
The  $m$ -value describes the experimentally observed linear change in stability upon addition of a denaturant<sup>2</sup> and was found to correlate well with the change in accessible surface area (ASA) between the native and the unfolded state.<sup>3</sup> Experimentally, protein stability is usually determined in an unfolding transition induced by increasing the temperature or by the addition of denaturants. Some proteins can also be denatured by a change in pH due to repulsion of equally charged amino acids and the loss of salt bridges in very acidic or basic conditions. A good indication for a two-state system is given if equilibrium transitions studied by experimental probes that measure different properties of the polypeptide chain superimpose each other.

## 1.2.2 Intermediates and barriers in protein folding

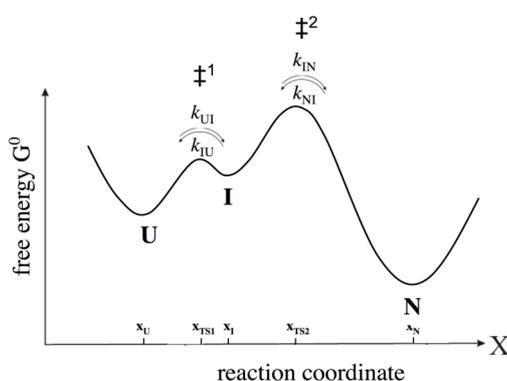
It is a major unsolved question in biophysics how a protein finds its native state. Insight into the folding process is given by kinetic studies: How fast does a protein fold? What is the folding mechanism? Does folding occur in a concerted reaction or are there intermediates? What are the properties of the free energy landscape for protein folding and dynamics? Are there side reactions like aggregation? The time a protein would need to find its native state by randomly sampling all possible conformations is astronomically large.<sup>4,5</sup> Experimental studies showed that folding times of small proteins range from microseconds to seconds,<sup>6</sup> a contradiction that is known as the Levinthal's paradox. In different theories obligatory intermediates are suggested to guide proteins along defined reaction pathways and by this speed up the folding process.<sup>7,8</sup> In hierarchical models, like the framework model or the diffusion-collision model, local interactions guide folding by formation of secondary structure and subsequent stabilization by long range interactions.<sup>7-15</sup> The nucleation-condensation model proposes the early rate-limiting formation of a folding nucleus followed by condensation of structure around this nucleus.<sup>16</sup> In contrast, the picture of fast folding due to a funnel-like energy landscape<sup>17,18</sup> with a small energy bias towards the native state<sup>19</sup> describes intermediates as non-productive kinetic traps. Experimental studies on many proteins revealed transient population of intermediates, which favored hierarchical folding. Because of their transient nature, intermediates are difficult to observe and to characterize. However, under certain conditions these partially folded states can be populated and studied experimentally. Nevertheless, several small single domain proteins were found to fold and unfold without population of kinetic or thermodynamic intermediates (Figure 1A).<sup>6,20,21</sup> Recent results on apparent two-state folders suggested the presence of high-energy intermediates, which are obligatory for fast folding (Figure 1B).<sup>22,23</sup> These intermediates are higher in energy than the unfolded and the native state and do not become populated to detectable amounts during folding/unfolding or in equilibrium.

The folding process can be described by a potential energy curve as shown in figure 1A for a two-state folding reaction. The native state (N) and the unfolded state (U) are separated by the transition state (TS or ‡). The folding kinetics is determined by the rate constants for folding and unfolding,  $k_f$  and  $k_u$ .  $\Delta G_{f,u}^{0\ddagger}$  is the activation free

energy for the folding and unfolding reaction, respectively. Figure 1B shows the potential energy curve for an apparent two-state folder going through a high-energy intermediate (I) during folding and unfolding.



**Figure 1A:** The free energy surface for a two-state folding reaction is shown as the free energy  $G^0$  of the reaction as a function of a certain reaction coordinate, which describes the progression of the folding reaction followed by a certain experimental observable.



**Figure 1B:** The free energy surface for an apparent two-state folder going through a high-energy intermediate I during folding/unfolding.

Due to the existence of free energy barriers, transition state theory (TST) of Eyring<sup>24</sup> was applied to protein folding reactions. This results in a reaction rate constant that is determined only by a pre-exponential factor  $k_0$  and the height of the barrier  $\Delta G^{0\ddagger}$

$$k = k_0 \exp\left(\frac{-\Delta G^{0\ddagger}}{RT}\right) \quad (5)$$

The pre-exponential factor  $k_0$  represents the maximum rate constant in the absence of any barrier and is specific for each reaction. In classical TST,  $k_0$  is given by

$$k_0 = \frac{k_B T}{h} \quad (6)$$

where  $h$  is the Planck constant. This reflects the maximum reaction rate constants for elementary reactions in organic chemistry where covalent bonds are formed and broken. At room temperature this corresponds to  $6 \cdot 10^{12} \text{ s}^{-1}$ , which reflects the vibration frequency of a covalent bond. This is not an appropriate description for folding reactions. Alternatively, in Kramer's theory<sup>25</sup> the solvent collision rate is faster than barrier crossing and thus the movement over the energy barrier is driven by temperature dependent Brownian motion. Coupling to solvent motion is introduced by a friction coefficient  $\gamma$  that enters the pre-exponential factor and the rate constant for folding becomes

$$k = \frac{C}{\gamma} \exp\left(\frac{-\Delta G^{0\ddagger}}{RT}\right) \quad (7)$$

The maximum rate constant depends on the friction coefficient and on a constant  $C$ , which is specific for each reaction. Under the assumption that friction is proportional to the solvent viscosity  $\eta$  this leads to an inverse dependence of the reaction on  $\eta$ . Equation 7 was shown to be valid for many reactions proceeding in aqueous solution.

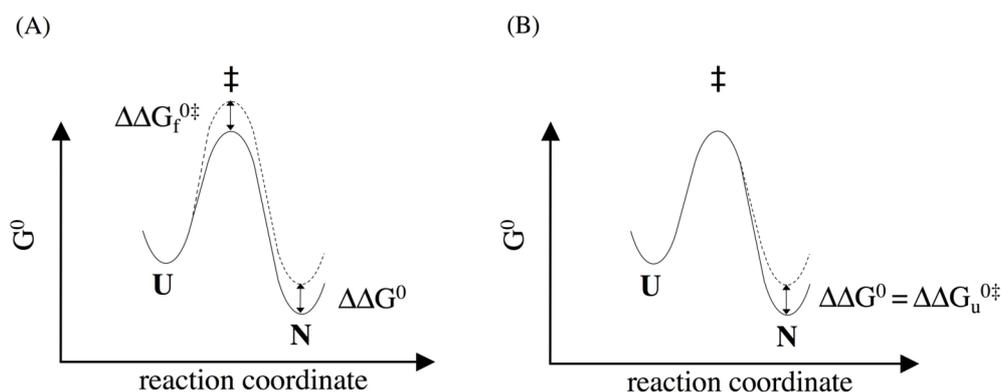
In analogy to equation 4, Gibbs fundamental equation applied to transition state properties is given by

$$d\Delta G^{0\ddagger} = \Delta V^{0\ddagger} dp - \Delta S^{0\ddagger} dT + m^\ddagger d[D] \quad (8)$$

A change in pressure gives information about the volume of the transition state,<sup>26</sup> variation of the temperature about its entropy, enthalpy and heat capacity<sup>20,27-29</sup> and variation of the denaturant concentration about its solvent accessible surface area. The change in activation free energy ( $\Delta\Delta G^{0\ddagger}$ ) relative to the change in reaction free energy ( $\Delta\Delta G^0$ ) with respect to a perturbation ( $\partial x$ ) is a measure for the position of the transition state on the reaction coordinate probed by  $\partial x$ <sup>30</sup>

$$\alpha_x = \frac{\partial\Delta G^{0\ddagger} / \partial x}{\partial\Delta G^0 / \partial x} \quad (9)$$

If  $\alpha_x = 1$  the change in  $\Delta G^0$  is already completely observed in the transition state ( $\partial\Delta G^0 / \partial x = \partial\Delta G_f^{0\ddagger} / \partial x$ ) what indicates that the transition state resembles the native state with respect to the property tested (Figure 2A). If  $\alpha_x = 0$  the change in  $\Delta G^0$  can be entirely attributed to  $\partial\Delta G_u^{0\ddagger} / \partial x$ , indicating that the transition state resembles the unfolded state with respect to the property tested (Figure 2B).



**Figure 2:** Effect of a perturbation  $\partial x$  on a protein folding reaction where A: the transition state resembles the native state ( $\alpha_x = 1$ ) or B: the transition state resembles the unfolded state ( $\alpha_x = 0$ ) with respect to the properties tested.

Extending this concept to structural perturbations, achieved by amino acid mutations, allows to determine to what extent individual side chain interactions are already formed in the transition state ( $\phi$ -value analysis).<sup>31</sup>

### 1.3 Early steps in protein folding

During protein folding the unstructured chain has to search for energetically favorable side-chain and backbone interactions to find its native structure. Thus, formation of site-specific interactions in a polypeptide chain is the first basic step during protein folding and sets an upper limit for protein folding rate constants. The maximum rate at which two points in a polymer chain can make contact is determined by intrachain diffusion. These early steps in protein folding correspond to a polymer problem, which depends on structural and dynamic properties of an unfolded polypeptide chain. To understand structural properties and chain dynamics in the unfolded state it

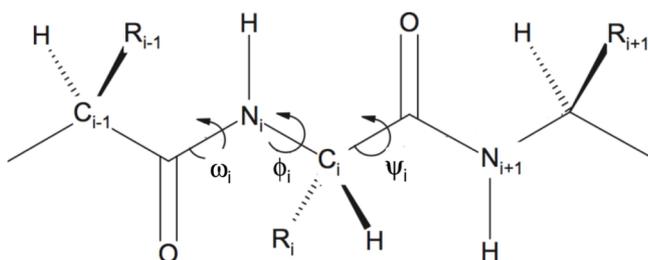
is convenient to use simplified polymer chain models for unfolded polypeptide chains.

### 1.3.1 Polymer chain models

In a simplified view, unstructured polypeptide chains can be described as random coil polymers, which can adopt a huge number of different conformations with no specific interactions between individual residues. Such an ideal polymer with  $n$  bonds of length  $l$  can be explained as freely jointed chain with an average end-to-end distance of

$$\langle r^2 \rangle = nl^2 \quad (10)$$

This chain model fails to describe local structural properties and is therefore only realistic to describe global properties of long chains. Further, it assumes equally probable angles at each joint what is not the case in real polypeptide chains. Here, neighboring segments influence each other and thus angles become restricted. The conformation of a polypeptide backbone can be described by specifying the dihedral angles  $\phi$  and  $\psi$  of all residues (Figure 3). For polypeptide chains, the allowed torsion angles are presented in conformational energy plots<sup>32-35</sup> and in the Ramachandran map.<sup>36</sup>



**Figure 3:** Chemical structure of the peptide backbone. Side chains are represented as  $R_i$ .

The phenomenon of restricted angles can be explained by chain stiffness. The end-to-end distance of such a chain is larger than expected from equation 10. Flory introduced the characteristic ratio  $C_n$  as a measure for the dimension of a stiff chain compared to a freely jointed chain.<sup>37</sup>

$$\langle r^2 \rangle = C_n nl^2 \quad (11)$$

$C_n$  increases with chain length until it reaches a limiting value  $C_\infty$  for long chains and

$$\langle r^2 \rangle = C_\infty n l^2 \quad (12)$$

$C_\infty$  equals the average number of chain segments that propagate in the same direction. For polypeptide chains it depends on the amino acid composition, with little contribution from flexible glycine residues and large contributions from the stiffer amino acids.  $C_\infty \cdot l$  is termed the Kuhn length  $b$ , which describes the hypothetical chain length of segments that can move freely without feeling chain stiffness.<sup>38,39</sup> The Kuhn length is directly connected to the persistence length  $l_p$  ( $b=2l_p$ ), which is a measure for the distance that an infinitely long chain propagates in the same direction. The excluded volume effect, introduced by Kuhn<sup>38</sup> describes that two points in a real chain cannot overlap. This results in increased chain dimensions and a larger end-to-end distance<sup>40</sup>

$$\sqrt{\langle r^2 \rangle} \propto l \cdot n^\nu \quad \text{with } \nu=0.59 \quad (13)$$

Flory introduced the isolated-pair hypotheses, which describes that each pair of  $\phi, \psi$  angles is independent of the values of its neighbors.<sup>37</sup> This hypothesis was shown to be invalid for polypeptide chains where steric effects extend beyond nearest-chain neighbors.<sup>41</sup> Compared to a dipeptide, the region of allowed  $\phi, \psi$  combinations in the Ramachandran map is decreased in longer chains. This results in more extended conformations and in a significantly restricted conformational space for the unfolded chain.

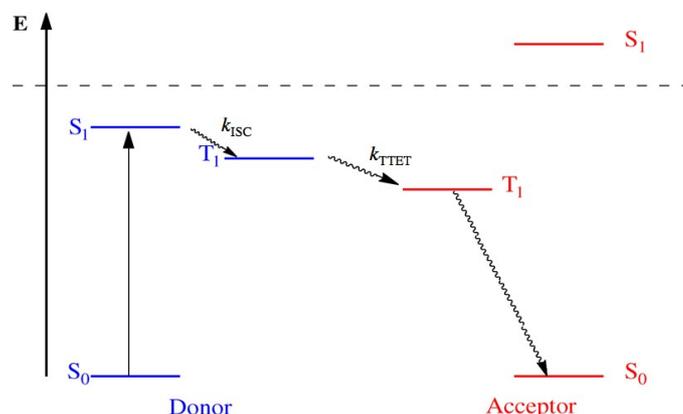
Chain structure and dimension are directly connected to macroscopic chain dynamics. Several theoretical approaches were developed to describe the dynamics in polymer chains. Based on Gaussian chains, the Rouse model<sup>42</sup> and the Zimm model<sup>43</sup> describe the dynamics of unstructured polymers in dilute solution. The optimized Rouse-Zimm approach (ORZ) further includes polymers of finite length and significant degree of chain stiffness.<sup>44</sup> Another important approach to describe unfolded polypeptide chains is the Szabo, Schulten, Schulten theory (SSS), which treats the kinetics of end-to-end contact formation.<sup>45</sup>

### 1.3.2 Triplet-triplet energy transfer

Triplet-triplet energy transfer (TTET) was shown to be a well-suited tool to directly measure the formation of site-specific interactions within a polypeptide chain in order to study the process of intrachain diffusion.<sup>46-48</sup> During TTET, excited-state energy is transferred from a triplet donor D to a triplet acceptor moiety A



via the simultaneous exchange of two electrons between the two groups.<sup>49</sup> The superscript indicates the electronic state of the molecule, with one for the singlet and three for the triplet state. Figure 4 shows a Jablonski diagram describing this process.



**Figure 4:** Jablonski diagram for triplet-triplet energy transfer from a triplet donor to a triplet acceptor. The rate  $k_{ISC}$  corresponds to formation of the donor triplet state via intersystem crossing.  $k_{TTET}$  is the rate constant for the energy transfer step.

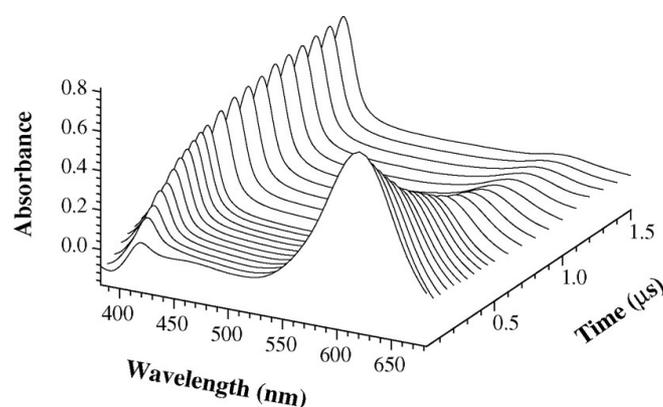
The rate constant for energy transfer was shown to decay exponentially<sup>49,50</sup> as the distance  $r$  separating donor and acceptor increases

$$k_{TTET} = A \cdot \exp\left(-\frac{2r}{L}\right) \quad (15)$$

$A$  denotes a factor that is dependent on photophysical properties and  $L$  is the average van-der-Waals radius of donor and acceptor. The strong distance dependence results in close contact (van-der-Waals contact) being a prerequisite for efficient energy

transfer. The strong distance dependence of TTET is an important difference to the well-known fluorescence resonance energy transfer (FRET), which designates the radiationless transfer of excited-state energy between a donor and an acceptor fluorophore. FRET is mediated by resonant oscillation of transient dipoles in the two chromophores. It occurs through space and the rate of energy transfer decays with  $1/r^6$  with  $r$  being the interchromophore distance.

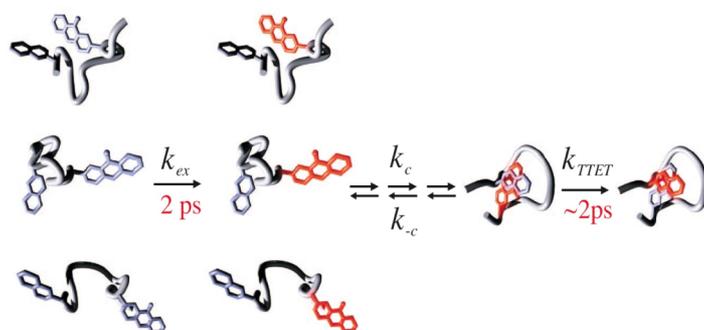
An ideal donor/acceptor pair for TTET studies should fulfill the following properties: Formation of the excited donor state must occur fast and with a high quantum yield. The intrinsic lifetime of the donor triplet state should be long to get access to a large experimental time window. This requirement is usually fulfilled since reversion to the ground state is spin-forbidden and thus slow. Further, the triplet energy of the acceptor must be several kJ below the triplet energy of the donor to obtain efficient, irreversible energy transfer upon formation of van-der-Waals contact. Xanthone and naphthalene were shown to be a suitable donor/acceptor pair. TTET in this system was shown to be diffusion-controlled.<sup>47</sup> The xanthone triplet state is formed within  $\sim 2\text{ps}$ <sup>51</sup> with a high quantum yield of  $\sim 99\%$ . Triplet transfer to naphthalene is irreversible and occurs with a time constant  $1/k_{\text{TTET}}$  of  $\sim 2\text{ps}$ . In degassed aqueous solution the intrinsic lifetime of the xanthone triplet is around  $20\mu\text{s}$ . Since the triplet states of xanthone and naphthalene show intense absorption bands, the energy transfer process can easily be detected (Figure 5).



**Figure 5:** TTET can easily be detected by monitoring the decay in xanthone triplet absorbance at 590 nm or the concomitant increase in naphthalene triplet absorbance at 420nm. Taken from reference 47.

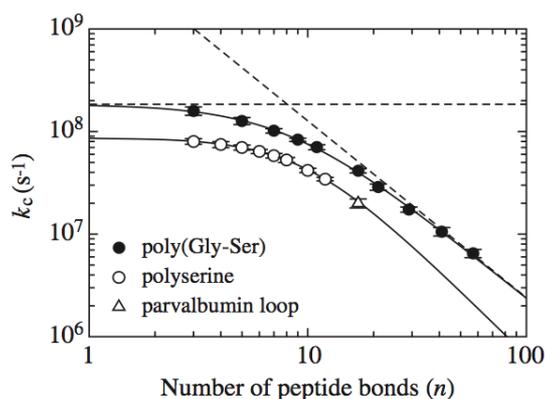
### 1.3.3 Chain dynamics in unfolded polypeptides

Introducing the TTET labels into a polypeptide chain (Figure 6) allows measurements of absolute rate constants for site-specific contact formation within the chain.



**Figure 6:** Scheme of an intramolecular TTET experiment to study site-specific contact formation in an unstructured polypeptide chain. Taken from reference 51.

Recent results on model polypeptide chains, revealed that the kinetics of end-to-end contact formation depend on chain length and on amino acid sequence.<sup>47</sup> TTET experiments performed on the nanoseconds to microseconds time scale revealed single exponential kinetics. This can be explained by very fast interconversion between the different conformations in the large unfolded ensemble. In poly(Gly-Ser) and polyserine peptides the rate limiting steps for motions over short and long distances are different in origin (Figure 7).<sup>47</sup> The time constants ( $1/k_c$ ) for end-to-end contact formation over short distances are almost independent of chain length and reach a limiting value of 5 – 10 ns for flexible glycine-serine chains. For longer poly(Gly-Ser) chains with more than 20 peptide bonds end-to-end contact formation scales with  $N^{-1.7\pm 0.1}$ , with  $N$  being the number of peptide bonds. In stiffer polyserine chains the end-to-end contact formation rate constants are significantly reduced and reach a twofold slower limiting value.

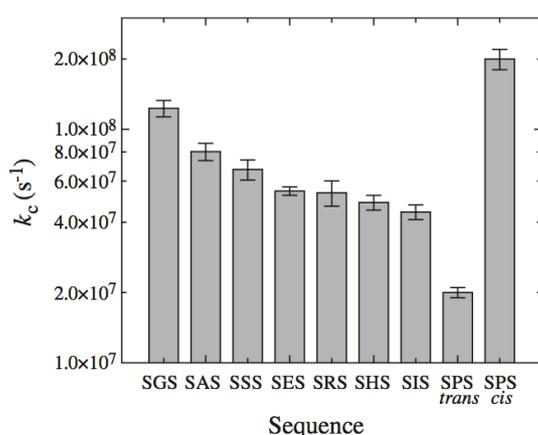


**Figure 7:** Rate constants  $k_c$  for end-to-end contact formation in model polypeptide chains and in a natural loop sequence derived from the parvalbumin protein. Adapted from references 47 and 53.

The observation that different processes limit intrachain loop formation for short and long loops is in agreement with polymer theory. Short chains cannot be treated like ideal chains due to strong influence of chain stiffness on chain properties. Long chains behave like Gaussian chains with contributions from excluded volume effects. However, the factors determining intrachain dynamics in the different regimes are still not well understood. A subpopulation of the large unfolded ensemble turned out to show much more complex loop closure kinetics on the 50 to 500 ps time.<sup>52</sup> This was explained by conformations, which can form loops by just a few bond rotations.

Intrachain contact formation kinetics in naturally occurring protein loops revealed rate constants comparable to the expected value from results on polyserine chains (Figure 7).<sup>53,54</sup> A problem of TTET in natural polypeptide sequences is that the amino acids tryptophane, tyrosine and methionine interact with the xanthone triplet state.<sup>53</sup>

The effect of amino acid sequence on the dynamics of loop formation was studied in host-guest peptides of the sequence Xan-Ser-Xaa-Ser-NAla-Ser-Gly where Xaa denotes one of the 20 naturally occurring amino acids, Xan is xanthone and NAla is the non-natural amino acid naphthylalanine. The results show that the amino acid side chain has only a small effect on the contact formation rate constant (Figure 8).<sup>47</sup> Glycine as a guest residue leads to the fastest dynamics consistent with its high flexibility. Amino acids with short side chains show slightly faster loop formation than long side chains. The proline peptide shows two rate constants corresponding to the *cis* and *trans* state of the Ser-Pro peptide bond, respectively.



**Figure 8:** Rate constants  $k_c$  for end-to-end contact formation in host-guest peptides to study the effect of individual amino acids on local chain dynamics. Taken from reference 47.

To improve the understanding of limiting factors for intrachain dynamics the effect of solvent viscosity and of temperature on intramolecular contact formation was recently addressed applying different approaches. These experiments yield information on the role of solvent friction and allow estimation of intrinsic free energy barriers. Results obtained from different methods disagree in several points. TTET experiments revealed fractional viscosity dependencies ( $k_c \sim \eta^{-\beta}$  with  $\beta < 1$ )<sup>46,55</sup> and significant free energy barriers<sup>55</sup> for end-to-end loop formation in short model polypeptide chains, indicating that loop formation in these peptides is not fully diffusion-controlled. Huang et al. studied end-to-end contact formation in unstructured fragments from the N-terminal  $\beta$ -hairpin of ubiquitin applying fluorescence quenching of 2,3-diazabicyclo[2.2.2]oct-2-ene (DBO) by tryptophane (Trp) and tyrosine (Tyr).<sup>56</sup> DBO quenching by Trp occurs via exciplex formation with a nearly diffusion-controlled bimolecular rate constant  $k_{q,H_2O} = 2.0 \cdot 10^9 \text{ M}^{-1}\text{s}^{-1}$  while quenching by Tyr occurs via hydrogen atom abstraction with slower rate constants of  $k_{q,H_2O} = 5.8 \cdot 10^8 \text{ M}^{-1}\text{s}^{-1}$ .<sup>57</sup> The results did not indicate any intrinsic barriers for loop formation in the flexible turn sequence. However, the results for the two stiffer strand sequences revealed significant intrinsic barriers for loop formation, indicating contributions from internal friction to chain dynamics in these sequences.<sup>56</sup> In agreement with results from TTET, the DBO/Trp system indicated intrinsic barriers for end-to-end loop formation in short poly(Gly-Ser) chains and further showed that the activation energy decreases with increasing chain length.<sup>58</sup> The absolute rate constants for loop formation measured by DBO/Trp quenching are slower by a factor of around two compared to the corresponding values measured by TTET from Xan to Nal. The DBO/Trp system further revealed a slightly increased activation energy for end-to-end contact formation in a (Ala-Gly-Gln)<sub>1</sub> chain compared to the value expected for a flexible poly(Gly-Ser) chain of similar length.<sup>59</sup> Another method used to study intramolecular contact formation is based on Trp triplet quenching by cysteine (Cys). The quenching mechanism for this system is not completely understood. The bimolecular quenching rate constant of  $k_q = 5.0 \cdot 10^8 \text{ M}^{-1}\text{s}^{-1}$  is significantly lower than the diffusion limit, indicating inefficient energy transfer upon contact formation.<sup>60,61</sup> Thus, this method does not allow direct measurement of contact formation rate constants but requires correction of experimentally determined quenching rate constants for the probability of quenching during the lifetime of the contact pair.

Lapidus et al. applied this system to study end-to-end loop formation in poly(Ala-Gly-Gln) chains using a correction factor of 2.<sup>62</sup> Significantly larger correction factors (17-28) were proposed by results from MD simulations.<sup>63</sup> The rate constants determined by the Trp/Cys pair are significantly slower compared to results from Xan/Nal triplet quenching and DBO/Trp fluorescence quenching. The Trp/Cys system was used to study end-to-end loop formation in poly(Ala-Gly-Gln) chains of different length.<sup>64,65</sup> The results indicated significant contributions from internal friction to dynamics in these chains.<sup>65</sup> Recently, a single molecule technique based on fluorescence quenching of the oxazine fluorophore MR121 by Trp monitored by fluorescence correlation spectroscopy (FCS) was introduced to measure specific contact formation.<sup>66</sup> Fluorescence quenching of MR121 by Trp occurs via photoinduced electron transfer upon van-der-Waals contact formation with a diffusion-limited bimolecular rate constant of  $k_q = 4.0 \cdot 10^9 \text{ M}^{-1} \text{ s}^{-1}$ .<sup>67,68</sup> Due to the requirement of defined mutual orientation between the aromatic groups, the rate constants determined in the FCS experiments need to be corrected to reveal rate constants for contact formation. Rate constants for end-to end contact formation in poly(Gly-Ser) chains and their dependence on loop length determined by the MR121/Trp system agree with results from TTET measurements.<sup>69</sup> Modulation of the solvent viscosity by adding sucrose showed that  $\beta = 1$  for  $(\text{Gly-Ser})_n$  with  $n=2,4,6,8,10$  and 14, indicating that loop formation in these chains is fully diffusion-controlled. This result is supported by both the observation that loop formation dynamics are not influenced by internal friction indicated by  $\tau = 1/k_c \rightarrow 0$  for  $\eta \rightarrow 0$  and by the absence of intrinsic energy barriers.<sup>69</sup> These results are not in agreement with results from TTET experiments, which indicate intrinsic barriers and  $\beta < 1$  for loop formation in short chains.<sup>55</sup>

#### 1.3.4 Aggregation prone polyglutamine sequences

Long polyglutamine stretches in proteins are associated with a variety of neurological human disorders, of which Huntington's disease is the most common.<sup>70</sup> Dependent on the length of the polyglutamine domain these proteins may aggregate to typically fibrillar aggregates, which are supposed to be connected to cellular toxicity.<sup>71-79</sup> Currently, these sequences are subject to extensive research to elucidate the kinetics and thermodynamics of the aggregation process<sup>80-82</sup> and to study different aggregate morphologies.<sup>83,84</sup> Aggregation of long polyglutamine sequences has been suggested

to occur via a nucleated-dependent polymerization mechanism for which an energetically unfavourable,  $\beta$ -structure containing species within the bulk monomer pool serves as nucleus.<sup>85</sup> Other observations propose that soluble unstructured monomers, which are somewhat more extended than true random coils aggregate in a more complex assembly mechanism via soluble aggregates lacking regular  $\beta$ -sheet structure.<sup>86</sup> Singh et al. used monomeric polyglutamine peptides of four to 16 amino acids length to study intrinsic chains dynamics by Trp/Cys contact quenching.<sup>87</sup> They found very slow rate constants for contact formation and proposed an extreme stiffness with a persistence length of 13Å for polyglutamine sequences. However, this method is not diffusion-controlled, which makes a quantitative evaluation impossible. It is still unclear whether structural and dynamic properties of unstructured polyglutamine chains are different compared to other unstructured polypeptide chains and whether this is connected to aggregate formation.

## 1.4 Protein-protein interactions

Protein-protein interactions play a central role in diverse biological processes like enzyme catalysis, inhibition and regulation, cellular signal transduction and folding of oligomeric proteins. Bimolecular reactions involving proteins are subject to extensive research. Nevertheless, the speed limit for site-specific interaction between two proteins or between a protein and a small ligand is still not well understood, however, it is a prerequisite for the detailed understanding of these processes.

### 1.4.1 Analytical models for bimolecular reactions

A bimolecular reaction between particles of species A and B can be written as



The first step, in which the encounter complex A\*B is formed from the free partners, depends on relative diffusion of the interacting particles. A subsequent unimolecular step forms the final complex C. The actual reversibility of the second step can be neglected under the assumption that C is very stable. If  $k_c \gg k_{-d}$  the reaction becomes

diffusion-limited and the association rate constant  $k_A \approx k_d$ . In the opposite limit where  $k_c \ll k_d$  the unimolecular step becomes rate limiting and  $k_A \approx k_c k_d / k_d$ .

According to the classical Smoluchowski theory,<sup>88</sup> which models the interacting molecules as uniformly reactive spheres, diffusion sets the upper limit of the association rate constant to  $5 \cdot 10^9 \text{ M}^{-1} \text{ s}^{-1}$  according to

$$k_S = \frac{N_L}{1000} 4\pi a (D_A + D_B) \quad (17)$$

$D_{A/B}$  are the diffusion coefficients of the reaction partners. The particles react as soon as their radial separation reaches the reaction distance  $a$  for the first time. This simple diffusion model assumes that diffusion occurs in a hydrodynamic continuum with no forces acting between the reactants what brings about several complications concerning its application to protein-protein association.<sup>89</sup> Under the assumption of a spherical shape the diffusion coefficient can be determined via the Stokes-Einstein equation

$$D = \frac{k_B T}{6\pi\eta r_H} \quad (18)$$

Here,  $r_H$  is the hydrodynamic or Stokes radius and  $\eta$  is the solvent viscosity. Experimentally  $D$  can be determined via classical diffusion measurements based on Fick's second law where the spreading of an initial sharp boundary between a buffer and a solution of known concentration of the molecule of interest is followed as a function of time. Since biological macromolecules have usually a very small diffusion coefficient the boundaries spread very slowly making a precise determination of  $D$  difficult. In practice other techniques are used to determine diffusion coefficients very accurately, among which the most important ones are analytical ultracentrifugation, dynamic light scattering (DLS) and nuclear magnetic resonance (NMR) spectroscopy. Debye found that the association rate constant between charged molecules can be affected by electrostatic interactions.<sup>90</sup> For two uniformly reacting spheres of charge  $z_A e$  and  $z_B e$  at the position  $r_A$  and  $r_B$  in space, the interaction potential energy is given by

$$U(r_A - r_B) = \frac{z_A z_B e^2}{4\pi\epsilon_0\epsilon|r_A - r_B|} \quad (19)$$

The diffusion-limited association rate constant is then given by<sup>90</sup>

$$k_{DS} = 4\pi(D_A + D_B) / \int_a^\infty e^{U(r_A - r_B)/k_B T} (r_A - r_B)^{-2} dr \quad (20)$$

When other, non-reactive ions are nearby, electrostatic interactions between charged reactants are partially screened. This results in an attenuated interaction potential energy<sup>91</sup> and thus in a decreased effect of electrostatic interactions on the association rate constant.<sup>90</sup> While bimolecular reactions between spherical and uniformly reactive molecules are quite well understood, less is known for protein-protein interactions and protein-ligand interactions.

### 1.4.2 Site-specific protein-protein interactions

Since proteins are usually not uniformly reactive over their entire surface, proper mutual orientation is a prerequisite for productive protein-protein and protein-ligand interactions. Due to the orientational constraints the maximum rate constant for intermolecular reactions involving proteins is expected to be smaller than the Smoluchowski limit of  $5 \cdot 10^9 \text{ M}^{-1} \text{ s}^{-1}$ . Several theoretical studies put the limit for site-specific protein-protein interactions in the range from  $10^5$  to  $10^6 \text{ M}^{-1} \text{ s}^{-1}$ .<sup>92-95</sup> It is difficult to experimentally test whether the association process is diffusion-controlled, reaction-controlled or a mixture of both limiting cases. A general hint for a protein-protein or protein-ligand interaction to be diffusion-controlled is a large association rate constant  $k_A$ . Further indications are a strong dependence of  $k_A$  on solvent viscosity and a low energy of activation. Coupled folding and association introduces additional complexity in folding studies of oligomeric proteins.

#### 1.4.2.1 Folding and assembly of dimeric proteins

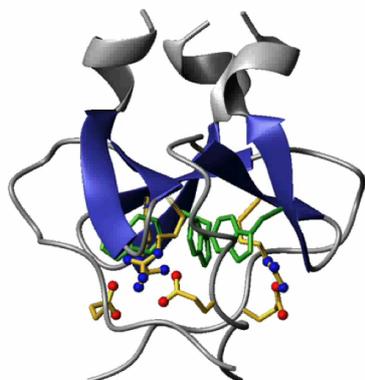
The yeast transcriptional activator GCN4 is one of the best-characterized dimeric proteins. Here, two identical alpha helices of 33 amino acids length each form a parallel two-stranded coiled coil.<sup>96</sup> Formation of the final dimer occurs with a

bimolecular rate constant  $k_A$  of  $4.2 \cdot 10^5 \text{ M}^{-1} \text{ s}^{-1}$ .<sup>97</sup> In a mutational analysis it was found that elements of secondary structure are already formed in the transition state of the rate-limiting step in the folding and association reaction and thus play a fundamental role during assembly.<sup>98</sup> Another well-studied homodimeric protein is the Arc repressor, which regulates transcription of the *ant* gene during lytic growth of bacteriophage P22. In the native Arc dimer, two identical monomers consisting of 53 amino acids each are strongly intertwined with each other forming a single globular domain with a conventional hydrophobic core.<sup>99,100</sup> Kinetic stopped-flow experiments revealed a bimolecular rate constant for formation of the Arc dimer of approximately  $2 \cdot 10^7 \text{ M}^{-1} \text{ s}^{-1}$  under physiological conditions.<sup>101</sup> Unfolding turned out to be fast with a rate constant of around  $0.1 \text{ s}^{-1}$  what results in a short average lifetime of ten seconds for the native dimer. A mutational analysis revealed transition state properties, which can be explained by simultaneous folding and association.<sup>102</sup> In wild type Arc, formation of salt bridges and hydrogen bond interactions mediated by the side chains of Arg31, Glu36 and Arg40 determine the rate-limiting step for folding. Replacing these interactions by the mutations Arg31Met, Glu36Tyr and Arg40Leu (MYL mutant) reduces the kinetic barrier for folding with almost no effect on the overall structure.<sup>103</sup> Compared to the wild type protein the rate constant for folding is increased by a factor of around ten to  $k_A = 3 \cdot 10^8 \text{ M}^{-1} \text{ s}^{-1}$ . While the assembly of wild type Arc depends only modestly on solvent viscosity,  $k_A$  for the MYL mutant is inversely proportional to solvent viscosity indicating that the reaction is diffusion-controlled.

#### 1.4.2.2 Folding and assembly of trimeric proteins

Detailed kinetic studies on the assembly of trimeric proteins are quite rare. Among the best-studied trimeric proteins is the tail spike protein of Salmonella phage P22<sup>104-106</sup> and collagen.<sup>107</sup> The large size of the proteins makes an extensive investigation of the association process very complex. Folding and assembly of the foldon domain from bacteriophage T4 fibrin is a well-studied trimerization reaction. Fibrin is a rod-like structural protein of bacteriophage T4. It is attached via its N-terminus to the neck of the bacteriophage and forms a homotrimer consisting of a N-terminal anchor domain (residues 1-46), a central coiled-coil domain (residues 47-456) and a small C-terminal globular domain (residues 457-486).<sup>108</sup> The C-terminal domain is called foldon

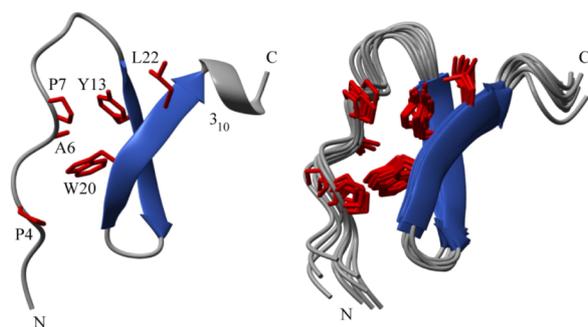
domain, which was shown to be an evolutionarily optimized trimerization motif that is essential for assembly of fibritin.<sup>108-110</sup> In isolation, the foldon domain adopts the same trimeric  $\beta$ -propeller structure as in fibritin (Figure 9).<sup>111</sup>



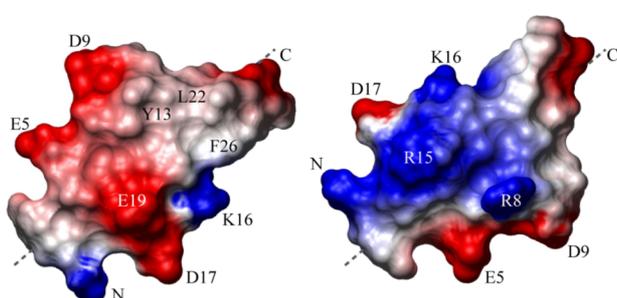
**Figure 9:** NMR structure of the trimeric foldon domain. The native trimer is stabilized by intermolecular H-bonds, salt bridges and hydrophobic interactions.<sup>111</sup>

Due to its small size the foldon domain is a good system to study folding and association reactions during the assembly process. Previous studies yielded detailed information on the mechanism of coupled folding and association during the assembly process.<sup>111</sup> During assembly, a monomeric intermediate with native-like topology is formed on the submillisecond time scale. This intermediate assembles into a partially folded trimer in two very fast consecutive bimolecular reactions with rate constants of  $1.9(\pm 0.5) \cdot 10^6 \text{ M}^{-1} \text{ s}^{-1}$  and  $5.4(\pm 0.3) \cdot 10^6 \text{ M}^{-1} \text{ s}^{-1}$  in the presence of 0.58 M guanidinium chloride (GdmCl). A final folding step yields the native trimer. Although the bimolecular reactions during foldon assembly are very fast it is unclear whether they are diffusion-limited. Mutation of an intermolecular salt bridge (E5R variant) was shown to lead to a monomeric state that resembles the monomeric folding intermediate in its spectroscopic properties and allowed the characterization of the structure and the dynamics of this kinetic intermediate using solution NMR spectroscopy.<sup>112</sup> Structural studies revealed essential native like topology of the monomeric intermediate with defined hydrogen bonds and side chain interactions, but largely increased flexibility and reduced stability compared to the native trimer (Figure 10A). The low stability of the intermediate allows a large free energy gain upon trimerization, which serves as driving force for rapid assembly. It is still unclear whether structure formation in the monomeric folding intermediate is important for fast assembly. The binding surfaces of the monomeric subunits in the native trimer

are oppositely charged (Figure 10B). Whether this asymmetric charge distribution guides the rapid assembly process electrostatically is still unknown.



**Figure 10A:** Comparison of a single subunit in the foldon wild type trimer (left) with the ten lowest energy structures of the E5R monomer (right). Taken from reference 112.



**Figure 10B:** Charge distribution on the different sides of a foldon wild type monomer with Glu at position 5 based on the E5R structure. Taken from reference 112.

### 1.4.2.3 Electrostatic enhancement of site-specific protein-protein interactions

The Debye-Smoluchowski theory predicts a modest effect of electrostatic interactions on diffusion-controlled bimolecular reactions. More realistic models for protein-protein association predict a much larger effect of electrostatics on the association kinetics.<sup>113,114</sup> In many systems, rapid association was indeed found to be supported by long-range electrostatic forces, which can enhance the bimolecular rate constant by several orders of magnitude. Optimization of the association rate constant is mainly important in situations where speed is essential.<sup>115-117</sup> Wendt et al. studied a designed heterodimeric parallel leucine zipper consisting of an acidic and a basic chain of 30 amino acids each.<sup>118</sup> They obtained a rate constant for dimer formation of  $k_A = 1 \cdot 10^9 \text{ M}^{-1} \text{ s}^{-1}$  at low ionic strength. Increasing the ionic strength significantly slowed down association resulting in  $k_A = 5 \cdot 10^5 \text{ M}^{-1} \text{ s}^{-1}$  at very high ionic strength. These results indicate a significant enhancement of the association process by favorable electrostatics. Another well-studied example is the process of complex formation of the nuclease barnase with its intracellular inhibitor barstar.<sup>119</sup> The binding sites of barnase and barstar are oppositely charged. In the absence of salt, association occurs

with a rate constant of  $k_A = 5 \cdot 10^9 \text{ M}^{-1} \text{ s}^{-1}$ . Increasing the ionic strength significantly slows down association and extrapolation to very high ionic strength where electrostatic forces are completely shielded yields a basal value for  $k_A$  of  $10^5 \text{ M}^{-1} \text{ s}^{-1}$ .<sup>119</sup> Selzer et al. introduced a strategy to predict the effect of electrostatic forces on the association rate constants of protein-protein interactions.<sup>120,121</sup> Besides generally speeding up protein-protein association, electrostatic attractive forces may also help guiding the interaction partners into proper mutual orientation and by this increasing the frequency of productive collisions.

#### **1.4.2.4 Theoretical models for site-specific protein-protein interactions**

Since proteins are usually not uniformly reactive over their entire surface, proper mutual orientation is a prerequisite for productive protein-protein and protein-ligand interactions. Due to the orientational constraints a reduced rate constant  $k_A = \Gamma \cdot k_S$  will be observed. The simple theoretical approach of multiplying the Smoluchowski rate constant  $k_S$  by the probability that in a random encounter the interacting molecules are properly oriented (“geometric rate constant”) yields rate constants that are usually several orders of magnitude smaller than experimentally observed values. This can be explained by reorientation during a single encounter event due to translational and rotational diffusion. Various analytical models were developed to quantify the retarding effect of partial reactivity and the role of rotational diffusion on the association rate constant. Solc et al. presented a formal solution for association of two spherical molecules bearing axially symmetric reactive patches, which cover a portion of the overall molecule surface.<sup>122</sup> Later they numerically studied the simplified case where one of the molecules is uniformly reactive.<sup>123</sup> This model is referred to as the model of Solc and Stockmayer and describes the binding of a small ligand to a partially reactive macromolecule. Schmitz and Schurr numerically investigated the reaction between a mobile sphere bearing a single axially symmetric reactive site and a stationary reactive site on a planar surface.<sup>124,125</sup> This model is referred to as the model of Schmitz and Schurr. Shoup et al. introduced the generally applicable constant flux approximation that allowed simplification of the complex formal solution of Solc et al. and Schmitz et al. to closed analytic expressions for the association rate constant.<sup>126</sup> They applied their approach to the model of Solc and Stockmayer for association of a uniformly reactive molecule B of radius  $r_B$  with a

molecule A of radius  $r_A$  that is reactive only over the axially symmetric region  $0 < \theta < \theta_A^0$ . In the limit  $\kappa \rightarrow \infty$  (diffusion-controlled reaction) this results in an expression for the bimolecular rate constant of

$$k_a = -k_s(1 - \cos\theta_A^0)^2 \left( \sum_{l=0}^{\infty} \frac{K_{l+1/2}(\xi^{*'})}{lK_{l+1/2}(\xi^{*'}) - \xi^{*'} K_{l+3/2}(\xi^{*'})} \cdot \frac{(P_{l-1}(\cos\theta_A^0) - P_{l+1}(\cos\theta_A^0))^2}{2l+1} \right)^{-1} \quad (21)$$

with 
$$\xi^{*'} \equiv \sqrt{\frac{D_A^{\text{rot}} l(l+1)}{D_A + D_B}} \cdot a \quad (22)$$

$K_l$  are modified Bessel functions of the second kind and  $P_l(x)$  is the  $l$ th order Legendre polynomial.  $D_A^{\text{rot}}$  is the rotational diffusion coefficient of molecule A. In the limit of small  $\theta_A^0$ ,  $r_A \gg r_B$  and  $\kappa \rightarrow \infty$  the bimolecular rate constant becomes

$$k_a = 4\pi D_B r_A (a / \pi r_A) \quad (23)$$

where  $a = r_A \theta_A^0$ . In 1993, Zhou used the constant flux approximation and presented an expression for the association rate constant of two spherical molecules bearing axially symmetric patches and considered the influence of electrostatic interactions.<sup>113</sup> In 2002, Schlosshauer et al. presented a general expression for the partially diffusion-controlled association rate constant of two molecules with an asymmetric reactive patch on each surface.<sup>127</sup> Within the constant flux approximation and in the limiting case of axially symmetric reactive patches they arrive at the solution presented by Zhou. For  $\kappa \rightarrow \infty$  the expression for the rate constant simplifies to

$$k_a = -k_s(1 - \cos\theta_A^0)^2 (1 - \cos\theta_B^0)^2 \left( \sum_{l, l_1, l_2} S(l, l_1, l_2) \right)^{-1} \quad (24)$$

The sum over the indices  $l, l_1, l_2$  runs from 0 to  $\infty$  and the summands are defined as

$$S(l, l_1, l_2) \equiv \frac{(2l+1)K_{l+1/2}(\xi^*)}{lK_{l+1/2}(\xi^*) - \xi^* K_{l+3/2}(\xi^*)} (2l_1+1)(2l_2+1) I_{l_1, A}^2 I_{l_2, B}^2 \begin{pmatrix} l & l_1 & l_2 \\ 0 & 0 & 0 \end{pmatrix} \quad (25)$$

$$\text{with } \xi^* \equiv \sqrt{\frac{D_A^{rot} l_1(l_1 + 1) + D_B^{rot} l_2(l_2 + 1)}{D_A + D_B}} \cdot a \quad (26)$$

$$I_{l,X} \equiv \int_0^{\theta_X^0} d\theta \sin\theta P_l(\cos\theta) = \frac{P_{l-1}(\cos\theta_X^0) - P_{l+1}(\cos\theta_X^0)}{2l+1}, \quad X = A, B \quad (27)$$

The matrix  $\begin{pmatrix} l & l_1 & l_2 \\ 0 & 0 & 0 \end{pmatrix}$  denotes the Wigner 3-j-symbol. Assuming one particle to be uniformly reactive the approach of Schlosshauer et al. coincides with the results introduced by Solc et al. and Shoup et al. In the fully diffusion-controlled limit and assuming both spheres to be uniformly reactive they arrive at the Smoluchowski rate constant. Further, they obtain a very good agreement with the results obtained from Brownian dynamics simulations.<sup>92</sup> The theoretical models for site-specific protein-protein interactions indicate that reorientation after initial contact formation significantly reduces the retarding effect of partial surface reactivity. They demonstrate that the deviation from the naïve geometric rate constant increases with increasing angular constraints. Another interesting finding is that for reactions between molecules with large difference in radii ( $r_A/r_B \rightarrow \infty$ ) partial surface reactivity of the small molecule can be neglected due to very fast rotational diffusion. At the same time rotational diffusion for the large molecule is slowed down and the molecule becomes apparently immobilized.

Theoretically, protein-protein and protein-ligand association reactions can be characterized by simulating translational and rotational diffusion of the interacting molecules.<sup>128</sup> Since diffusion of biological macromolecules is quite slow these Brownian dynamics simulations rely on the accessibility of a relatively large time window. This is achieved via simplification by neglecting conformational dynamics within the individual subunits. Usually many trajectories are simulated starting with the proteins at random orientations separated by a certain distance. If part of these trajectories lead to productive association, then computer algorithms can be applied to derive the association rate constant.<sup>113,129-131</sup> To describe the criteria for productive association detailed and accurate absorbing boundary conditions have to be defined. Definition of the boundary conditions is usually quite complex and often coupled to

experimental results in a way to get good agreement of predicted and experimental association kinetics.<sup>132</sup> To predict association rate constants from theory alone, the encounter complex ensemble has to be specified and located on the reaction coordinate without reference to experiment.<sup>128</sup> However, it is still unclear how the encounter complex of protein-protein and protein-ligand interactions does look like. Sometimes the encounter complex<sup>132</sup> is referred to as the transient complex.<sup>93,133</sup> Often it is assumed that the interacting molecules in the encounter complex have mutual separations and orientations that are close to the native alignment.<sup>93,133-136</sup> Generally, two extreme scenarios are proposed concerning the binding mechanism between two proteins or between a protein and a small ligand.<sup>137</sup> The induced fit model suggests binding in the unfolded state and subsequent folding. On the other hand, the conformational selection model suggests that binding cannot occur in the unfolded state. It proposes rare formation of folded/active molecules in the unbound state and subsequent binding.

## **1.5 Effect of cellular crowding on intra- and intermolecular diffusion processes**

Inside living cells macromolecules may occupy as much as 40% of the total intracellular volume and reactions proceed in a very crowded environment. This could have a significant effect on thermodynamics and kinetics of intracellular diffusion processes.<sup>138</sup> Experiments carried out in the test tube in the absence of cellular components may thus be misleading for the understanding of diffusion processes in nature. To study the effect of crowding *in vitro* several types of crowding agents that mimic the dense environment were introduced. Beside nucleic acids, mostly folded proteins account for crowding inside cells. Therefore, proteins like ovalbumin or hemoglobin were used as mimics for the cellular environment. Other widely used crowding agents are polyethylene glycols, dextrans and Ficoll. Compared to the real *in vivo* environment, all these solutions are quite well defined and easy to manipulate.

A major effect from cellular crowding is increased solvent viscosity. The concept of the Reynolds number nicely demonstrates that viscous forces become very important when small things move through fluids.<sup>139</sup> The Reynolds number  $R$  is a dimensionless

parameter in the equations of motion in a fluid that indicates the relative size of terms that describe inertial forces and viscous forces. It is given by

$$R = \frac{vL\rho}{\eta} \quad (28)$$

Here,  $v$  is the velocity of the particle moving through a fluid with specific gravity  $\rho$  and viscosity  $\eta$ .  $L$  is the dimension of the particle. For a small molecule, like a flagellated bacteria the Reynolds number is very small, which means that it moves itself by using viscous share without experiencing any inertia. This becomes clearer from calculating the distance the bacterium coasts when it stops swimming. The result is impressive revealing that it will come to a stop within about 1  $\mu$ s and will coast a distance of around  $4 \cdot 10^{-12}$  m, a distance small compared to the diameter of a hydrogen atom. However, the bacterium does not actually stop because it is still subject to Brownian motion. For intra- and intermolecular protein motions the Reynolds number is very small. This indicates that inertia is not important for protein diffusion and intrachain diffusion and results in motions, which are determined by solvent friction and other frictional forces.

Processes that are limited by solvent friction are expected to decrease with  $1/\eta$ .<sup>25</sup> Using small viscous additives like glycerol or ethylene glycol, end-to-end contact formation in model polypeptide chains was found to depend strongly on solvent viscosity.<sup>46,140</sup> The effect of increased solvent viscosity on end-to-end loop formation is significantly reduced in the presence of large crowding agents.<sup>69,140</sup> The finding that large viscous co-solutes have a small effect on intrachain dynamics is consistent with the observation that crowding in the cytoplasm usually slows down diffusion of small solutes and many macromolecules by only a few-fold compared to their diffusion in water.<sup>141</sup> However, several observations describe large reductions in solute diffusion inside cells. This might be explained by an additional steric effect of crowding on solute movement by providing fixed barriers or by interactions between solutes and cellular components.<sup>141</sup>

For protein-protein interactions the effect of increased viscosity is expected to be larger for long-range translational diffusion than for the faster and more local

rotational diffusion. Together this is expected to lead to a very complex dependence of association reactions on viscosity. Several experimental studies on diffusion inside cells<sup>142,143</sup> and in cell-simulating solutions<sup>144-146</sup> indicate that the relative size of crowding and associating molecules might be an important parameter.

A second major effect of macromolecular crowding is the restriction in available space due to mutual impenetrability of all molecules. Considering a single polypeptide chain, the excluded volume theory predicts that macromolecular crowding decreases the dimension of the unfolded state and stabilizes more compact conformations.<sup>147,148</sup> This will favor protein folding entropically. Several theoretical approaches were developed to describe the effect of excluded volume on association reactions.<sup>149-151</sup> They predict that the excluded volume effect will result in an effective entropically based distance-dependent attractive potential, which is acting when the associating molecules are very close to each other and results in an enhancement of association reactions in crowded solutions.<sup>150,152</sup> This was confirmed in experiments on excluded volume induced association.<sup>153,154</sup>

The overall effect of molecular crowding on intra- and intermolecular dynamics is complex. On the one hand dynamics can be enhanced by the excluded volume effect. On the other hand, crowding can slow down kinetics by reducing diffusion. The effect of crowding will thus depend on whether the investigated reaction is diffusion-controlled or reaction-controlled. For a folding reaction the excluded volume effect may reduce  $\Delta G^{0\ddagger}$  and by this speed up the reaction. If intrachain diffusion plays a critical role crowding may act to reduce the rate of intrachain diffusion and hence decrease the value of the pre-exponential factor in Kramer's equation. Different contributions of the opposing effects explain the different results from experiments on the effect of crowding on intrachain dynamics in unfolded polypeptide chains,<sup>69</sup> for formation of secondary structure elements<sup>155</sup> and for folding of large proteins.<sup>138,156-163</sup> For larger proteins the excluded volume effect seems to be dominant resulting in increased folding rate constants.

In analogy, fast diffusion-controlled protein-protein interactions are more affected by reduced diffusion while slow association reactions can be significantly enhanced by

the excluded volume effect. Several experimental studies revealed that the crowding environment has only a small effect on many protein-protein interactions.<sup>164-167</sup>

## 2. Aims of Research

### 2.1. Intramolecular diffusion processes during protein folding

#### *Kinetics of loop formation in unstructured polypeptide chains*

Formation of site-specific interactions in a polypeptide chain represents the elementary step in protein folding and thus sets an upper limit to protein folding rate constants. In earlier experiments triplet-triplet energy transfer (TTET) was used to study the kinetics of end-to-end loop formation in poly(glycine-serine) and polyserine model peptides of different length. Rate constants for loop formation were found to approach a limiting value in short chains while they scale with a power law in longer chains, indicating that the rate-limiting step for motion over short and long distances are different in origin. We wanted to characterize the rate-limiting step for motion in these different regimes in more detail by analyzing the effect of solvent viscosity and of temperature on the kinetics of end-to-end loop formation. This allows to assess contributions from solvent interactions and to characterize the free energy barriers for loop formation. To obtain more information on dimension and internal dynamics of unfolded polypeptide chains we wanted to apply fluorescence resonance energy transfer (FRET). Usually, rather harsh conditions are required to populate the unfolded state of a protein. Natively unfolded proteins provide good model systems to study the unfolded state under physiological conditions. To study loop formation in a natural unfolded polypeptide chain we wanted to apply TTET to a fragment from the natively unfolded brinker protein.

#### *Kinetics of loop formation in aggregation prone polyglutamine peptides*

Several neurodegenerative diseases associated with proteins are based on the presence of abnormally long polyglutamine stretches. The mutant protein aggregates into intraneuronal, nuclear inclusions, which are assumed to play a central role in the disease process. The reason why polyglutamine stretches beyond a critical length lead to aggregate formation is not known. A possible explanation is special properties in the unfolded state of polyglutamine chains. We wanted to apply TTET to study end-to-end loop formation in short unstructured polyglutamine peptides to test whether polyglutamine chains have different dynamical and structural properties compared to other polypeptide chains.

## **2.2 Intermolecular diffusion processes during protein folding and assembly**

Site-specific protein-protein and protein-ligand interactions play an important role in diverse biological processes. These processes are limited by diffusion of the interacting molecules. The speed-limit for site-specific interactions is still not known.

### *Site-specific interactions between unfolded polypeptide chains*

To improve the understanding of site-specific protein-protein and protein-ligand interactions we applied intermolecular TTET to different model systems. We aimed to answer the following questions: 1) What is the diffusion-controlled rate constant for site-specific association between two unfolded polypeptide chains and between an unfolded polypeptide chain and a small molecule? 2) Is the association process slowed down by orientational constraints? 3) Is the bimolecular rate constant influenced by intermolecular electrostatic interactions?

### *Assembly of the foldon domain from bacteriophage T4 fibrin*

Due to its small size the trimeric foldon domain from bacteriophage T4 fibrin is a good model to study formation of site-specific protein-protein interactions during protein assembly. Assembly of the native trimer is limited by two consecutive bimolecular association reactions. The foldon domain can be synthesized chemically what allows introduction of the labels for TTET measurements. We wanted to determine the maximum possible rate constant for site-specific formation of an encounter complex between two monomeric foldon subunits by applying TTET to the monomeric foldon E5R variant. This allows to test whether dimer formation during assembly of the wild type protein is diffusion-controlled. Results from NMR solution spectroscopy indicate that the binding surfaces of the monomeric subunits in the native trimer are oppositely charged. By varying the salt concentration, we wanted to investigate the role of long-range electrostatic interactions on the assembly of the wild type trimer and on the maximum rate constant for encounter complex formation between two monomeric subunits. To study how fast folding can occur in the absence of the association reactions, we wanted to eliminate relative translational diffusion of the monomeric subunits during the assembly process. We wanted to apply two strategies in which three monomeric foldon subunits are either placed within a single continuous amino acid chain or are connected using a cyclic template.

## **3. Results**

### **3.1 Dynamics of intramolecular loop formation in polypeptide chains**

Specific contact formation between two sites in an unfolded polypeptide chain represents the elementary step in protein folding and determines how fast the vast conformational space can be explored in search for energetically favourable conformations. Characterization of the dynamics in unfolded polypeptide chains is therefore important to understand the earliest steps in protein folding and thus to characterize the protein folding process in detail. Recently, several experimental methods were introduced to study the dynamics of specific contact formation and were applied to unfolded model polypeptide chains and natural protein sequences (see chapter 1.3.3). The different approaches revealed different rate constants for a certain contact formation reaction and results further lead to different conclusions on limiting processes for chain dynamics in the unfolded state. Thus, the factors determining intrachain dynamics in unfolded polypeptide chains are still not well understood.

#### **3.1.1 Effect of solvent viscosity and temperature on intrachain diffusion processes**

The goal of this part of the work was to gain a detailed understanding of the dynamics of intramolecular loop formation in polypeptide chains. TTET measurements on the dynamics of end-to-end loop formation in different polypeptide chains had shown that the kinetics of formation of long loops are in agreement with the behaviour expected for a pure diffusive search of a Gaussian chain.<sup>47</sup> In contrast, formation of short loops is slower than expected for a pure diffusive search<sup>47</sup>, indicating the presence of barriers and a change in the rate-limiting step that should affect the viscosity-dependence of the reaction. To characterize the dynamics and barriers of loop formation in the different regimes of short and long loops in more detail, we determined the effect of solvent viscosity and temperature on the formation of intramolecular van-der-Waals contacts.

To obtain absolute free energy barriers and viscosity-dependencies the TTET process between the two chromophores has to be fully diffusion-controlled. The effect of

solvent viscosity on the rate constant of a chemical reaction can be described by the empirical equation

$$k_c = k_{c,0} \left( \frac{\eta}{\eta_0} \right)^{-\beta} \quad (29)$$

where  $\eta_0$  is the reference solvent viscosity of water at  $T = 22.5^\circ\text{C}$  and  $k_{c,0}$  is the rate constant at  $\eta_0$ . The exponent  $\beta$  reflects the sensitivity of the reaction to solvent viscosity. For purely diffusive reactions, which are fully viscosity-dependent a  $\beta$ -value of 1 is expected. Reactions with high barriers should be completely independent of solvent viscosity resulting in  $\beta = 0$ . Previous results from our group showed that bimolecular TTET from xanthone to naphthalene has a  $\beta$ -value of 1 in accordance with a diffusion controlled-reaction.<sup>140</sup> This result was confirmed by studies on the effect of temperature on the bimolecular rate constant for TTET from xanthone to naphthalene. The viscosity-corrected bimolecular rate constants determined at different temperatures revealed a low activation energy of  $2.6(\pm 0.4)$  kJ/mol,<sup>140</sup> which is on the order of  $RT$  as expected for a diffusion-controlled reaction. These results show that TTET from xanthone to naphthylalanine is diffusion-controlled and not limited by a free energy barrier. Thus, TTET allows determination of absolute rate constants, viscosity-dependencies and free energy barriers for van-der-Waals contact formation processes.

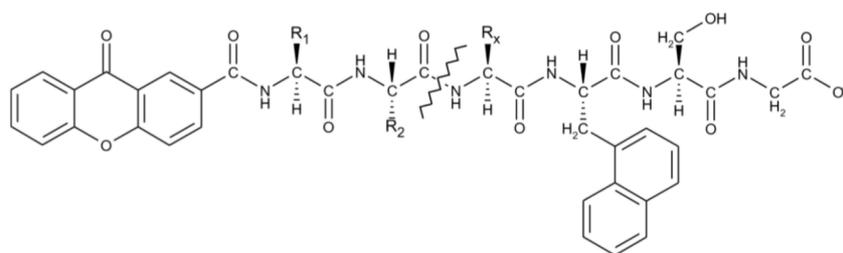
### 3.1.1.1 Effect of viscous co-solvents on intrachain dynamics in model polypeptide chains and natural protein sequences

According to Kramer's theory the solvent collision rate for a chemical reaction is faster than barrier crossing and thus movement over a free energy barrier ( $\Delta G^{0\dagger}$ ) is driven by temperature dependent Brownian motion.<sup>25</sup> Coupling of a chemical reaction to solvent motion is introduced into the rate equation by a frictional coefficient,  $\gamma$ , which enters the pre-exponential factor

$$k = \frac{C}{\gamma} \exp\left(\frac{-\Delta G^{0\dagger}}{RT}\right) \quad (7)$$

where  $C$  is a reaction-specific constant. Assuming that solvent friction is proportional to solvent viscosity,  $\eta$ , equation (7) leads to  $k \propto 1/\eta$ . Kramer's relation was confirmed in early experimental studies on the effect of solvent viscosity on the rate constants of conformational transitions of polymers in solution<sup>168-170</sup> and was further applied to describe dynamic states of proteins involved in enzymatic reactions.<sup>171,172</sup> We used TTET to study the effect of solvent viscosity on end-to-end loop formation in different polypeptide chains. Earlier studies had shown that the kinetics of formation of long loops is in agreement with the behaviour expected for a pure diffusive search of a Gaussian chain and thus a  $1/\eta$  viscosity dependence of the rate constant for loop formation ( $k_c$ ) is expected. In contrast, formation of short loops is slower than expected for a pure diffusive search indicating the presence of barriers or a change in the rate-limiting step that should affect the viscosity-dependence of the reaction.

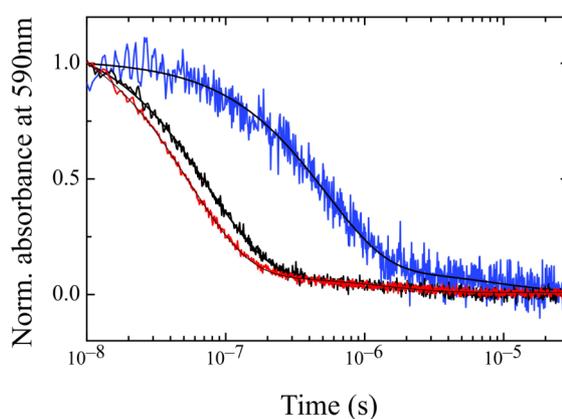
We investigated loop formation in peptides of the general form shown in figure 11 with the triplet donor xanthone (Xan) attached at the N-terminus and the triplet acceptor naphthylalanine (Nal) located near the C-terminus. Since TTET between Xan and Nal is a diffusion-controlled and fast process, the measurement of the transfer kinetics yields the rate constant for intramolecular loop formation ( $k_c$ ) that leads to van-der-Waals contact between the labels.



**Figure 11:** General form of the used model peptides to study end-to-end loop formation by TTET. The triplet donor xanthone was attached to the N-terminus and the acceptor group was introduced via the non-natural amino acid naphthylalanine near the C-terminus.

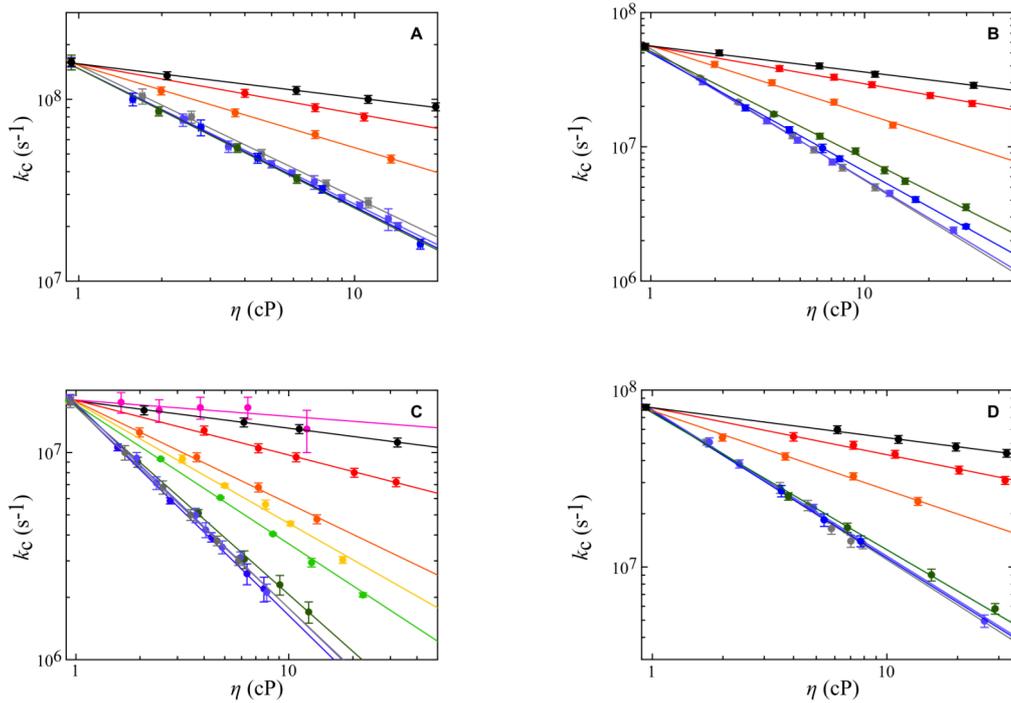
Figure 12 compares TTET kinetics measured by the transient triplet absorbance decay of the Xan donor at 590 nm in a (Gly-Ser)<sub>14</sub> chain in water with  $\eta=1$  cP and at  $\eta=10.3$  cP in the presence of either glycerol or PEG 20000. A double exponential function is

necessary to describe the observed triplet absorbance decay under all conditions. The slower kinetic phase has only about 5 % amplitude and its time constant of  $1.4(\pm 0.1)$   $\mu\text{s}$  in water corresponds to the intrinsic xanthone lifetime. This reaction is caused by molecules that do not undergo TTET during the lifetime of the donor triplet state, probably due to formation of small aggregates. The faster kinetic phase with a time constant of  $54.7(\pm 4.0)$  ns in water is due to TTET from Xan to Nal and reflects the kinetics of intramolecular loop formation. Figure 12 shows that loop formation in the  $(\text{Gly-Ser})_{14}$ -peptide is slowed down at increased solvent viscosity, but the magnitude of the effect is different for glycerol and PEG 20000.



**Figure 12:** Transient triplet absorbance decay observed for a  $(\text{Gly-Ser})_{14}$  chain in water (—) and at  $10.3(\pm 0.2)$  cP in the presence of glycerol (—) and PEG 20000 (—), respectively. Measurements were carried out in 10 mM KP pH 7.0 at  $T=22.5$  °C.

Figure 13 shows the effect of solvent viscosity in the presence of different co-solutes on the rate constant  $k_c$  of end-to-end loop formation in  $(\text{Gly-Ser})_n$  chains with  $n = 1, 6$  and 14 and in a  $\text{Ser}_2$  chain. The results reveal that increasing the solvent viscosity slows down end-to-end loop formation in all peptides, but the magnitude of the effect depends on the co-solute size. The effect of Ficoll-70 on  $k_c$  was investigated since this co-solute has been widely applied in experimental crowding studies.<sup>69</sup>



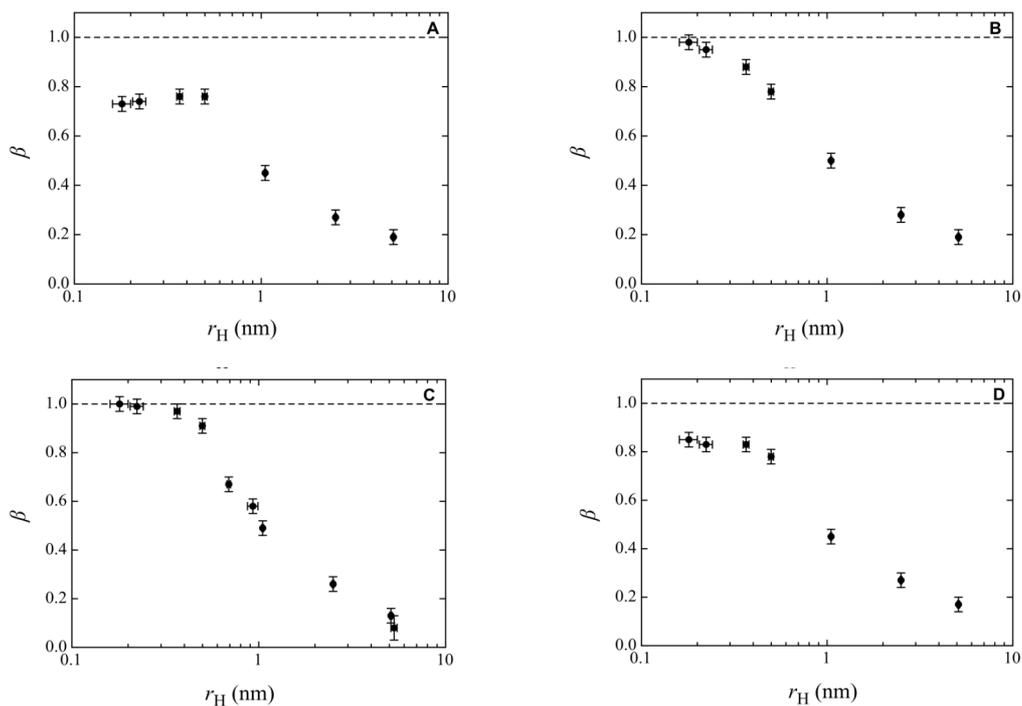
**Figure 13:** Effect of solvent viscosity on the rate constant  $k_c$  for end-to-end loop formation in  $(\text{Gly-Ser})_n$  peptides with  $n=1$  (A),  $n=6$  (B),  $n=14$  (C) and in a  $(\text{Ser})_2$  peptide (D). The lines represent a fit of equation 29 to the data. The viscosity was adjusted by adding Ficoll-70 (●), PEG 20000 (●), PEG 6000 (●), PEG 1500 (●), PEG 1000 (●), PEG 600 (●), sucrose (●), glucose (●), glycerol (●) and ethylene glycol (●). Part of the results was taken from reference 140. Measurements were carried out in 10 mM KP pH 7.0 at  $T=22.5^\circ\text{C}$ .

The viscosity dependence of  $k_c$  can be described by the empirical equation

$$k_c = k_{c,0} \left( \frac{\eta}{\eta_0} \right)^{-\beta} \quad (29)$$

A fit of equation 29 to the data shown in figure 13 reveals that the  $\beta$ -value for intrachain loop formation depends on the nature of the co-solute used to modulate solvent viscosity. This result indicates that chain motions do not see the macroscopic viscosity of the solvent. The microviscosity experienced by the polypeptide chain should depend on the size of the co-solute. Figure 14 shows  $\beta$  as function of the co-solute's hydrodynamic radius  $r_H$ , measured by dynamic light scattering (for ethylene

glycol, glycerol, glucose, sucrose, PEG 400, PEG 600 and PEG 1000) or taken from references 173 (for PEG 1500, PEG 6000 and PEG 20000) and 69 (Ficoll-70). The  $\beta$ -value is small for large co-solutes and increases with decreasing size of the co-solute until it reaches a limiting value below a critical co-solute size (Figure 14). This result shows that the macroscopic solvent viscosity and the microscopic viscosity experienced by a polypeptide chain are identical only for small co-solutes, which can thus be used to determine the actual viscosity-dependence of loop formation. This result is interesting in context of the effect of cellular crowding on diffusion processes and suggests that increased viscosity inside cells, which is due to the presence of high molecular weight molecules, has only a small effect on intrachain dynamics. The reduced effect of large co-solutes on  $k_c$  is in agreement with the observation that increased solvent viscosity in the presence of large crowding agents also has only a small effect on bimolecular diffusion processes.<sup>141</sup>



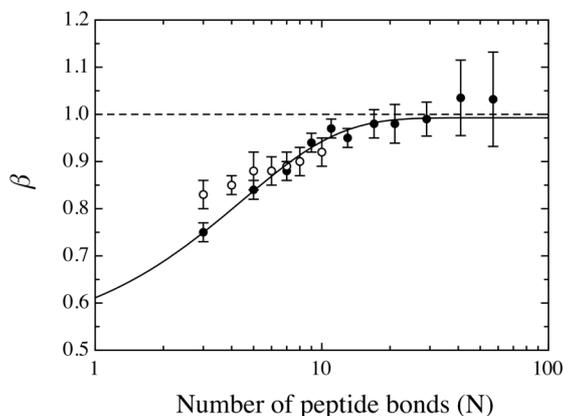
**Figure 14:** Sensitivity ( $\beta$ ) of loop formation rate constants to solvent viscosity determined for  $(\text{Gly-Ser})_n$  chains with  $n=1$  (A),  $n=6$  (B),  $n=14$  (C) and in  $(\text{Ser})_2$  (D) vs. the hydrodynamic radius  $r_H$  of the co-solutes used to modulate solvent viscosity.

The origin of the fractional viscosity-dependence ( $\beta < 1$ ) observed for large co-solvents can be explained by a microviscosity experienced in an inhomogeneous

medium.<sup>141,174</sup> According to the Stokes-Einstein relation and the Stokes-Einstein-Debye relation diffusional processes in solution are inversely proportional to solvent viscosity. This theory assumes a continuous and homogeneous solvent to which the concept of Brownian diffusive motion is applicable. This assumption does neither agree with the situation *in vivo*, where the local viscosity is governed by the presence of a relatively small number of large macromolecules nor with adding viscous co-solvents in order to adjust solvent viscosity *in vitro*. The co-solutes used to adjust solvent viscosity are much larger and less mobile than water. Thus, the solute will not see a continuous solvent but an inhomogeneous medium. Studies on translational Brownian motion of small molecules and on rotational Brownian motion of proteins in viscous aqueous solution pointed out deviations from the Stokes-Einstein relation<sup>174</sup> and from the Stokes-Einstein-Debye relation,<sup>175</sup> respectively. A detailed analysis of the viscosity dependent rate constant of oxygen escape from photodissociated oxyhemerythrin revealed that the size of the viscous co-solute is a critical parameter.<sup>176</sup>

We measured the effect of solvent viscosity modulated by the addition of glycerol on the rate constant of end-to-end loop formation in poly(Gly-Ser) chains and in polyserine chains to determine the viscosity dependence of  $k_c$ . Glycerol is a small molecule and thus exerts the full viscous effect on the dynamics of loop formation. To test whether viscous additives affect the dimension of the unfolded polypeptide chains we performed DLS measurements on the following amino acid sequence: *H-Gly-Arg-Gly-Ser-Gly-Ser-Gly-Ser-Arg-Gly-Ser-Gly-Ser-Arg-Gly-Ser-Nal-Ser-Gly-OH* with Nal = naphthylalanine. The arginine residues were introduced to increase solubility of the polypeptide chain in order to reach peptide concentrations high enough for DLS measurements. Results revealed a hydrodynamic radius of  $r_H = 1.00(\pm 0.15)$  nm for this polypeptide chain in the absence and in the presence of 30 % (w/v) glycerol, indicating that glycerol has no effect on the dimension of the studied model polypeptide chain. Figure 15 shows that the sensitivity of  $k_c$  on solvent viscosity is a function of loop length. In (Gly-Ser) $_n$  chains with  $n > 8$  the rate constant of end-to-end loop formation decreases with  $1/\eta$  ( $\beta = 1$ ). This indicates that loop formation in long chains is diffusion-controlled. In shorter chains with  $n < 8$ ,  $k_c$  exhibit a fractional viscosity dependence ( $\beta < 1$ ) even for small co-solutes, indicating that the reaction is

not fully diffusion-controlled. Similar results were found for stiffer polyserine chains. However, due to limitations in polypeptide synthesis, we were not able to study long polyserine chains. It is thus still unclear whether  $\beta$  also reaches a limiting-value of 1 in long polyserine chains.

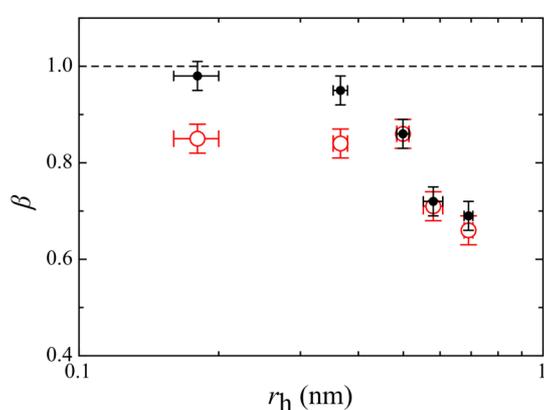


**Figure 15:** Sensitivity of end-to-end loop formation at  $T=22.5\text{ }^{\circ}\text{C}$  in poly(Gly-Ser) chains (●) and polyserine chains (○) to solvent viscosity in the presence of glycerol. Adapted from reference 140.

The observation of two different limits for formation of short loops and long loops is in agreement with previous TTET measurement on the effect of loop-length on  $k_c$  in unfolded poly(Gly-Ser) and polyserine chains.<sup>47</sup> These experiments show that loop formation scales with  $N^{-1.7\pm 0.1}$  for long poly(glycine-serine)<sub>n</sub> chains with  $n > 10$  in agreement with the value expected from polymer theory for a Gaussian chain with excluded volume effects. The rate constant for end-to-end loop formation over short distances becomes almost independent of chain length and reaches a limiting value of  $1.8(\pm 0.2)\cdot 10^8\text{ s}^{-1}$  for flexible poly(Gly-Ser) chains. A similar behaviour was observed for stiffer polyserine chains with a limiting value of  $8.7(\pm 0.8)\cdot 10^7\text{ s}^{-1}$  and a scaling law of  $k_c \sim N^{-2.1\pm 0.3}$  for longer chains. Our results on the effect of solvent viscosity on intrachain loop formation in polypeptide chains of different length support the model of different rate-limiting processes for formation of long and short loops.

As shown in figure 14, the effect of solvent viscosity on end-to-end loop formation in poly(Gly-Ser) and polyserine chains becomes stronger with decreasing co-solute size and reaches a limit for small co-solvents. For a (Gly-Ser)<sub>14</sub> chain this limit ( $\beta = 1$ ) is reached between glycerol and glucose (Figure 14C), which have hydrodynamic radii of 0.22 nm and 0.37 nm, respectively. Comparing the corresponding hydrodynamic diameters ( $d_H = 2\cdot r_H$ ) to the length of a peptide bond in a flexible chain (0.35nm)

reveals that the dimensions of glycerol and glucose correspond to the length of 1.3 and 2.1 peptide bonds, respectively. This is in the range of the persistence length of  $l_p = 2$  peptide bonds (0.7 nm) expected for flexible poly(Gly-Ser) chains.<sup>177</sup> This result suggests that the upper limit for the critical co-solute size that still exerts the full viscous effect corresponds to the persistence length of the chain. To test this model we studied the effect of solvent viscosity on end-to-end loop formation in a stiffer (Thr-Ser)<sub>7</sub> chain and compared the results to the flexible (Gly-Ser)<sub>14</sub> peptide. Due to low solubility of the (Thr-Ser)<sub>7</sub> peptide, two additional arginine residues were introduced at the C-terminal end. A corresponding (Gly-Ser)<sub>14</sub> peptide with two additional arginine residues at the C-terminus was synthesized for comparison. According to Flory, a (Thr-Ser)<sub>7</sub> chain and a (Gly-Ser)<sub>14</sub> chain have similar end-to-end distances.<sup>37</sup> TTET measurements in water reveal similar rate constants for end-to-end contact formation in water for the (Thr-Ser)<sub>7</sub> and (Gly-Ser)<sub>14</sub> peptides. The two additional N-terminal arginine residues in the (Gly-Ser)<sub>14</sub> peptide slightly decrease the rate constant for loop formation but they do not affect the critical co-solute size (Figure 16). Measurements in the presence of different viscous additives revealed a limit of  $\beta = 0.85$  for (Thr-Ser)<sub>7</sub> in the presence of small co-solutes. This limit is reached between sucrose and PEG 400 with hydrodynamic radii of 0.50 nm and 0.58 nm, respectively, corresponding to the length between 2.9 and 3.4 peptide bonds. This value is higher compared to the limiting size for (Gly-Ser)<sub>14</sub>, it is, however, shorter than the persistence length in the range of 5 to 7 peptide bonds predicted for an ideal polypeptide chain. However, the effect of peptide solvent interactions and intramolecular hydrogen bond formation may reduce the persistence length in real solutions.



**Figure 16:** Effect of co-solute dimension on the sensitivity ( $\beta$ ) of end-to-end loop formation at  $T=22.5$  °C to solvent viscosity for a (Thr-Ser)<sub>7</sub> chain (●) and a (Gly-Ser)<sub>14</sub> chain (●), respectively.

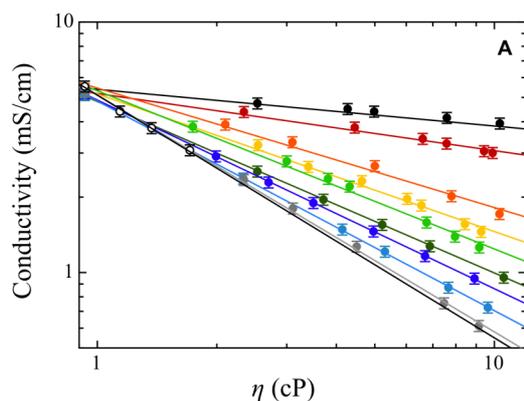
If the critical maximum size that a co-solute may have to exert the full effect of solvent viscosity on loop formation is dependent on the persistence length of the polypeptide chain, a critical co-solute size should not be observed for free diffusion of small molecules, which occurs on a very small length scale. To test for this assumption we studied the effect of solvent viscosity on ion motion by measuring electrical conductivity in a 50 mM sodium chloride solution. The measured conductivity  $\kappa$  can be described by the Nernst-Einstein equation

$$\kappa = \frac{z^2 F^2}{RT} [D_+ + D_-] \quad (30)$$

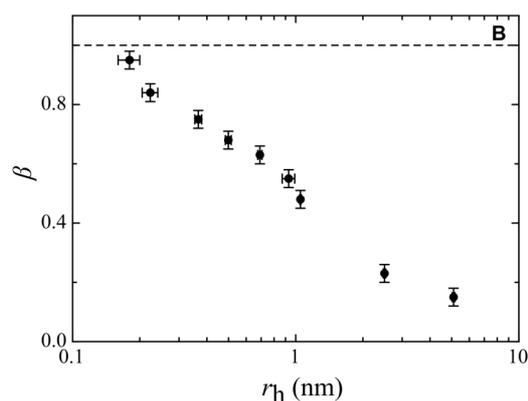
where  $z$  is the charge of the ion,  $F$  is Faraday's constant,  $R$  is the gas constant,  $T$  is the absolute temperature in K.  $D_{+/-}$  are the diffusion coefficients of the cation and the anion, respectively. According to the Stokes-Einstein equation

$$D = \frac{k_B T}{6\pi\eta r_H} \quad (18)$$

the diffusion coefficient is inversely proportional to solvent viscosity and thus, the effect of solvent viscosity on conductivity should also be described by equation (29). Figures 17A and 17B show that  $\beta$  does not reach a limiting value for the motion of small ions in solution. This result is consistent with the idea that the critical co-solute size observed for loop formation in unfolded polypeptide chains depends on the dimensions of the underlying motions and is thus a function of the persistence length in polypeptide chains.



**Figure 17A:** Electrical conductivity  $\kappa$  in a 50 mM NaCl solution at  $T = 22.5^\circ\text{C}$  as a function of solvent viscosity. The viscosity was adjusted by adding PEG 20000 ( $\bullet$ ), PEG 6000 ( $\bullet$ ), PEG 1500 ( $\bullet$ ), PEG 1000 ( $\bullet$ ), PEG 600 ( $\bullet$ ), sucrose ( $\bullet$ ), glucose ( $\bullet$ ), glycerol ( $\bullet$ ), ethylene glycol ( $\bullet$ ) or urea ( $\bullet$ ).



**Figure 17B:** Sensitivity ( $\beta$ ) of ion movement to solvent viscosity vs. co-solute dimension.  $T = 22.5^\circ\text{C}$

To compare the effect of solvent viscosity on the dynamics of loop formation in the homopolypeptide chains to the effect on naturally occurring amino acid sequences, we investigated peptides derived from the natural proteins carp muscle  $\beta$ -parvalbumin, the DNA-binding domain of the brinker protein and the B1 domain of protein G. Figure 18 shows the sequence of the studied protein fragments and their structure within the context of the full-length protein. In general, aromatic amino acids were replaced by the triplet labels, since they have similar size and chemical properties. All fragments are unfolded as judged by CD spectroscopy (see chapter 3.1.2 and reference 53).



Fragment	Sequence
GB1 <sub>45-52</sub>	Xan-DDATKT-NAla-SG-NH <sub>2</sub>
GB1 <sub>45-52</sub> T49G	Xan-DDAGKT-NAla-SG-NH <sub>2</sub>
GB1 <sub>45-52</sub> T49P	Xan-DDAPKT-NAla-SG-NH <sub>2</sub>
PA DE-loop	Xan-KADARALTAGETKT-NAla-SG-NH <sub>2</sub>
PA EF-loop	Xan-LKAGSDGDGKIGVDE-Nala-SG-NH <sub>2</sub>
Brk <sub>66-89</sub>	Xan-GKGNQRATARKFNIFRRQIQK-Nala-LQ-NH <sub>2</sub>

**Table 1:** Amino acid sequence of natural protein fragments modified to apply TTET. Xanthone (Xan) is attached to the N-terminus while naphthylalanine (Nala) is introduced near the C-terminus. The C-terminus of all peptides was amidated. In all fragments Ser-Gly was introduced at the C-terminus to increase flexibility near the resin during SPPS.

The GB1<sub>45-52</sub>-peptide comprises residues 45 to 52 from the B1 domain of protein G (GB1) which forms a  $\beta$ -hairpin loop structure in the native protein (Figure 18A). In this eight amino acid fragment tyrosine at position 45 was replaced by the triplet donor xanthone and phenylalanine at position 52 by the triplet acceptor naphthylalanine (Table 1). These positions are well-suited for labelling since the side chains of residues 45 and 52 are close in the native structure of GB1. To test for fast reactions occurring in the dead-time of the photoflash experiments ( $\sim 10$  ns) we compared the TTET kinetics in the donor/acceptor peptides with triplet decay of donor-only peptides, in which the triplet acceptor naphthylalanine was replaced by alanine, which cannot undergo TTET with xanthone. The results show a significant loss of the xanthone triplet absorbance in the experimental dead-time (10 ns) indicating subnanosecond reactions of loop formation in the GB1<sub>45-52</sub> peptide (Figure 19A and 19B) compatible with fast loop formation reactions on the 200 ps time scale observed in short polyserine peptides by a combination of femtosecond- and nanosecond-laser flash experiments.<sup>52</sup> In addition to the dead-time reactions TTET in the GB1<sub>45-52</sub> fragment occurs in two resolved kinetic phases with time constants of  $1.0(\pm 0.1)$   $\mu$ s and  $30.3(\pm 1.9)$  ns. The slow kinetic phase corresponds to the intrinsic xanthone lifetime and the fast kinetic phase corresponds to end-to-end loop formation. Comparison to the dynamics in the homopolypeptide models shows that loop formation in GB1<sub>45-52</sub> in water is slower than the value observed for a polyserine

chain of the same length (Figure 19C and table 2). Additionally, we studied two variants of the GB1<sub>45-52</sub> peptide to test for the effect of proline and glycine residues on the dynamics of loop formation, since these residues frequently occur in turn regions.<sup>181,182</sup> Replacing Thr49 by the more flexible glycine residue accelerates end-to-end loop formation in water. Introducing the stiffer proline residue, in contrast, leads to slower kinetics (Figure 19 and table 2). At the same time the dead-time amplitude decreases when a proline is introduced (Figure 19A and 19B). TTET measurements in the presence of glycerol reveal a  $\beta$ -value of 0.9 for the wild-type GB1<sub>45-52</sub> fragment and for both variants (Figure 19D and table 2) indicating that proline and glycine residues have no effect on the contributions of solvent interactions to the dynamics of loop formation in these fragments.

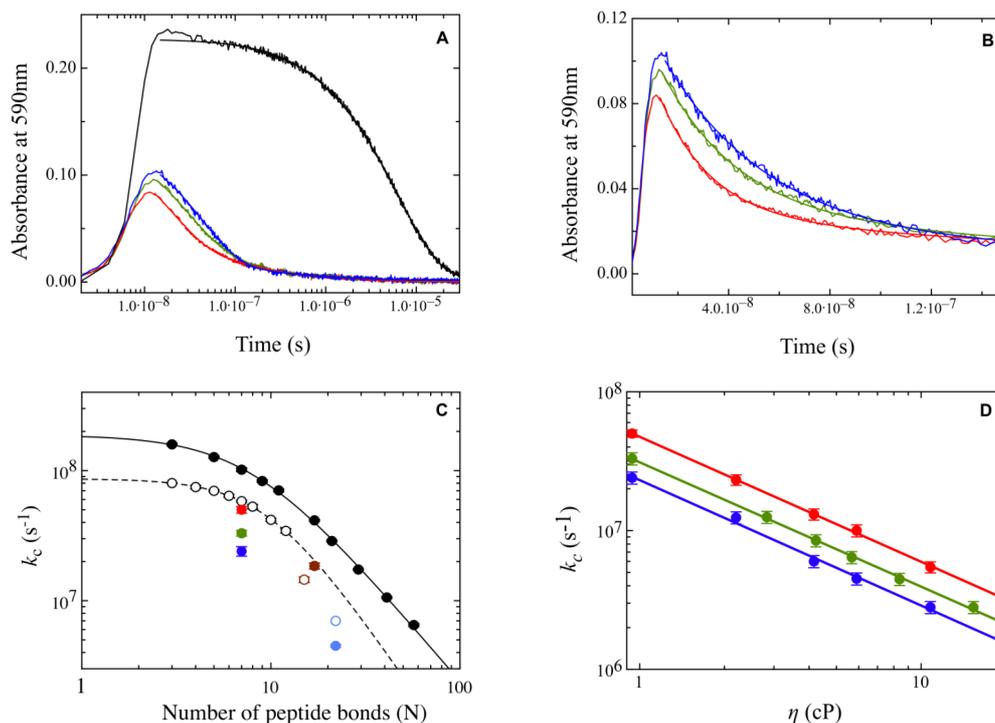
We further studied end-to-end loop formation in two loops derived from  $\beta$ -carp parvalbumin (Figures 18B and 18C). The DE-loop comprises residues 70 to 85 whereas the EF-loop fragment comprises residues 85 to 102 and binds Ca<sup>2+</sup> in the native state. Both loops connect two  $\alpha$ -helices in the native structure of parvalbumin. Xanthone and naphthylalanine were introduced at positions of phenylalanine residues 70 and 85 in the DE-loop and at positions of phenylalanine residues 85 and 102 in the EF-loop. These positions are within van-der-Waals contact in the native structure (Figures 18B and 18C). The time constant for loop formation in the EF-loop is 52.6( $\pm$ 2.8) ns, which is comparable to the value expected for a polyserine chain of similar length (Fig. 19).<sup>53</sup> The rate constant for loop formation in the DE-loop is significantly slower than in the polyserine chains (Figure 19C). This difference can be explained by the higher content of glycine residues in the EF-loop compared to the DE-loop. TTET measurements in the presence of glycerol reveal  $\beta$ -values of 0.86 for the DE-loop and  $\beta = 0.85$  for the EF-loop (Table 2) indicating that loop formation in both fragments is sensitive to solvent viscosity but not fully diffusion-controlled.

We further studied end-to-end loop formation in a fragment consisting of amino acids 66 to 89 from the DNA-binding domain of the brinker protein (Brk<sub>66-89</sub>, see chapter 3.1.2). This domain is natively unfolded and only forms structure upon binding to DNA. The 24 amino acid Brk<sub>66-89</sub>-fragment includes helices II and III from the DNA-binding motif (Figure 18D). In the absence of DNA  $\alpha$ -helix formation in this segment

is prevented by repulsive  $i, i+4$  interactions between lysine and arginine side chains (Figure 18D). These interactions are screened in the presence of negatively charged phosphate groups from the DNA backbone, which induces  $\alpha$ -helix formation. Xanthone was attached to the N-terminus of the Brk<sub>66-89</sub>-fragment and naphthylalanine was introduced at the position of tryptophane 87 (Table 1). Since Tyr77 and His80 in Brk<sub>66-89</sub> interfere with TTET experiments these residues were replaced by phenylalanines (Table 1). Loop formation in the unfolded Brk<sub>66-89</sub>-fragment is significantly slowed down compared to model polypeptide chains of similar length (Figure 19A and table 2). An acceleration of loop formation in the presence of sodium chloride indicates that the effect can partly be attributed to repulsive interactions between positively charged side chains (Figure 19C). TTET measurements in the presence of glycerol reveal a  $\beta$ -value of 0.87 for the Brk<sub>66-89</sub>-fragment, which is similar to the  $\beta$ -values found for the parvalbumin fragments (Table 2). In summary, all studied natural protein sequences showed fractional viscosity dependences in the range of  $\beta = 0.86$  to  $0.90$  indicating that end-to-end loop formation in all fragments is not fully diffusion-controlled.

Fragment	$k_0$ (H <sub>2</sub> O) (s <sup>-1</sup> )	$\beta$
GB1 <sub>45-52</sub>	$3.3(\pm 0.2) \cdot 10^7$	$0.90(\pm 0.05)^A$
GB1 <sub>45-52</sub> T49G	$5.0(\pm 0.3) \cdot 10^7^A$	$0.90(\pm 0.04)^A$
GB1 <sub>45-52</sub> T49P	$2.4(\pm 0.2) \cdot 10^7^A$	$0.90(\pm 0.05)^A$
PA DE-loop	$1.5(\pm 0.1) \cdot 10^7^A$	$0.86(\pm 0.03)^A$
PA EF-loop	$1.9(\pm 0.1) \cdot 10^7^B$	$0.85(\pm 0.03)^A$
Brk <sub>66-89</sub>	$4.5(\pm 0.1) \cdot 10^6$	$0.87(\pm 0.02)^A$
Brk <sub>66-89</sub> +salt	$7.0(\pm 0.1) \cdot 10^6$	-

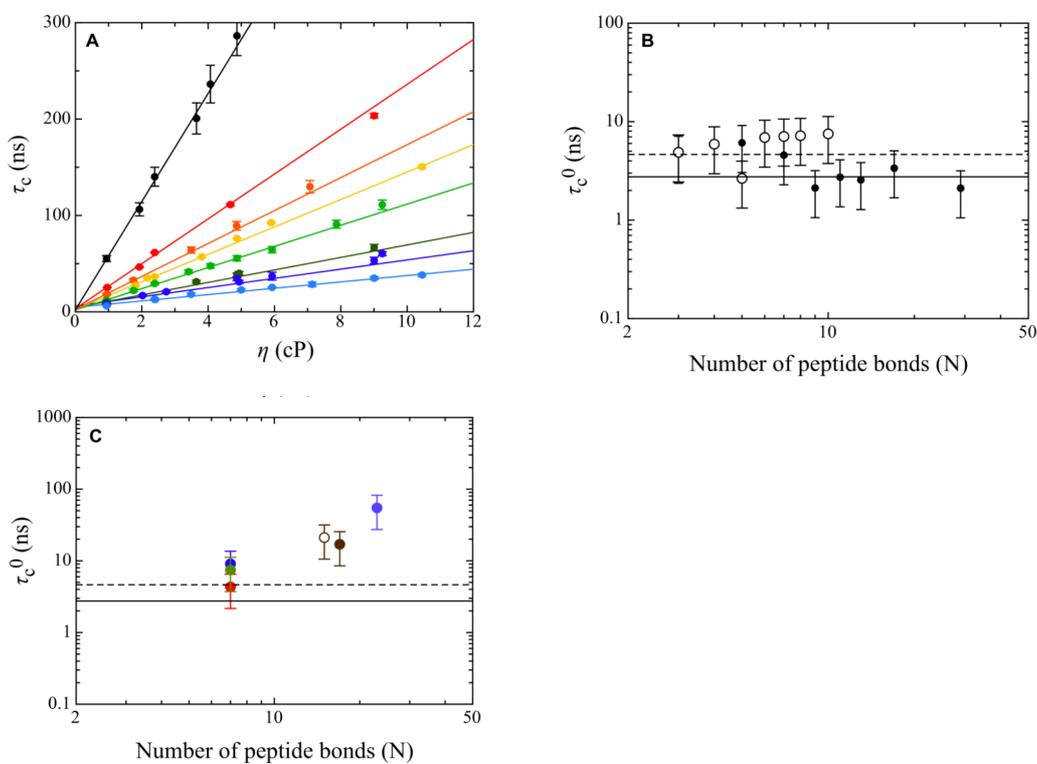
**Table 2:** Rate constant  $k_c$  (H<sub>2</sub>O) for end-to-end loop formation in natural protein sequences in water at  $T=22.5$  °C and the corresponding  $\beta$ -values determined from TTET measurements in the presence of glycerol. A: Data taken from reference 140, B: Data taken from reference 53.



**Figure 19:** A and B: Transient triplet absorbance decay observed in the donor/acceptor peptides  $GBI_{45-52}$  (green),  $GBI_{45-52}$  T49G (red) and  $GBI_{45-52}$  T49P (blue) and in a  $GBI_{45-52}$  donor-only peptide (black). C: Comparison of  $k_c$  for end-to-end loop formation in natural loop sequences and model polypeptide chains in water: poly(Gly-Ser) ( $\bullet$ ), polyserine ( $\circ$ ),  $GBI_{45-52}$  ( $\bullet$ ), in  $GBI_{45-52}$  T49G ( $\bullet$ ) and  $GBI_{45-52}$  T49P ( $\bullet$ ), DE-loop ( $\circ$ ), EF-loop ( $\bullet$ ),  $Brk_{66-89}$  ( $\bullet$ ) and  $Brk_{66-89}$  in the presence of 300 mM NaCl ( $\circ$ ). D: Viscosity dependence of  $k_c$  for loop formation in  $GBI_{45-52}$  ( $\bullet$ ),  $GBI_{45-52}$  T49G ( $\bullet$ ) and  $GBI_{45-52}$  T49P ( $\bullet$ ). Part of the data was taken from references 53 and 140 (see table 2). Measurements were carried out in 10 mM KP pH 7.0 ( $GBI$ -peptides and homopolypeptides), in 10 mM NaCac pH 7.0 (parvalbumin fragments) or in 20 mM NaAc 20 mM NaCl, pH 5.7 ( $Brk_{66-89}$ ) at  $T = 22.5^\circ\text{C}$ .

Recent studies on the effect of solvent viscosity on folding of cytochrome *c* and TrpCage showed that folding time constants ( $1/k_f$ ) in both proteins vary linearly with solvent viscosity and extrapolate to a positive intercept  $\tau_c^0$  as  $\eta \rightarrow 0$ .<sup>183,184</sup>  $\tau_c^0$  was suggested to represent the maximum time constant for folding in the absence of solvent friction and was interpreted as the contribution of intrinsic friction to the folding reaction. Analysis of the viscosity-dependence at 20° C revealed  $\tau_c^0=0.7 \mu\text{s}$  for TrpCage<sup>183</sup> and  $\tau_c^0=6 \mu\text{s}$  for cytochrome *c* folding from the compact M-state to

the native state<sup>184</sup>. Figure 20A shows the same analysis for the viscosity-dependence of loop formation in poly(Gly-Ser) and polyserine chains of different length.



**Figure 20:** A: Time constant for loop formation in  $(\text{Gly-Ser})_n$  chains as a function of solvent viscosity:  $n = 1$  (●), 2 (●), 3 (●), 4 (●), 5 (●), 6 (●), 8 (●), 14 (●). B and C: Limiting time constant  $\tau_c^0$  for  $\eta \rightarrow 0$  vs. loop length for poly(Gly-Ser) (●), polyserine (○), GBI<sub>45-52</sub> (●), GBI<sub>45-52</sub> T49G (●) and GBI<sub>45-52</sub> T49P (●), DE-loop (○), EF-loop (●), Brk<sub>66-89</sub> (●). The lines represent the average  $\tau_c^0$  for polyserine chains (---) and poly(Gly-Ser) chains (—). Data on poly(Gly-Ser), polyserine, GBI-peptides and parvalbumin fragments was adapted from reference 140.

A linear relationship between the time constant for loop formation ( $\tau_c = 1/k_c$ ) and solvent viscosity is observed for all polypeptides. Extrapolation to  $\eta \rightarrow 0$  reveals a positive intercept ( $\tau_c^0$ ) as observed for cytochrome *c* and TrpCage. A linear fit gives values of  $\tau_c^0 = 2.8(\pm 2.0)$  ns for poly(Gly-Ser) chains and of  $\tau_c^0 = 4.6(\pm 2.0)$  ns for poly(serine) chains. For both series of peptides  $\tau_c^0$  does not depend on chain length (Figure 20B). The values for  $\tau_c^0$  are only about 2-fold smaller than the limiting values of  $5.5 \pm 1.0$  ns and  $11.5 \pm 1.0$  ns observed for short chains in the length-dependence of loop formation in poly(Gly-Ser) and polyserine chains, respectively (Figure 7). This

comparison shows that internal friction contributes significantly to the dynamics of loop formation in short chains.

For the natural protein fragment GB1<sub>45-52</sub> and the T49G and T49P variants  $\tau_c^0$  is 7.5( $\pm$ 3.7) ns, 4.3( $\pm$ 2.2) ns and 9.1( $\pm$ 4.5) ns, respectively, which is comparable to the values observed in unfolded model polypeptide chains (Figure 20C). However, Figure 20C shows that this limit is significantly increased in the DE-loop ( $\tau_c^0=21.1(\pm 10.6)$  ns) and the EF-loop fragments ( $\tau_c^0=17.0(\pm 8.5)$  ns) from  $\beta$ -carp parvalbumin and in the Brk<sub>66-89</sub>-fragment ( $\tau_c^0=54.7(\pm 27.4)$  ns). The  $\tau_c^0$ -values in the range of 2-55 ns for loop formation in all studied peptides indicate that in the absence of solvent friction other processes become rate limiting for intrachain dynamics. From our measurements we cannot determine whether these processes are solvent dependent or not. Possible origins of these processes are internal friction and/or bond rotations.

### 3.1.1.2 Characterization of the barriers for intrachain loop formation in model polypeptide chains and natural protein sequences

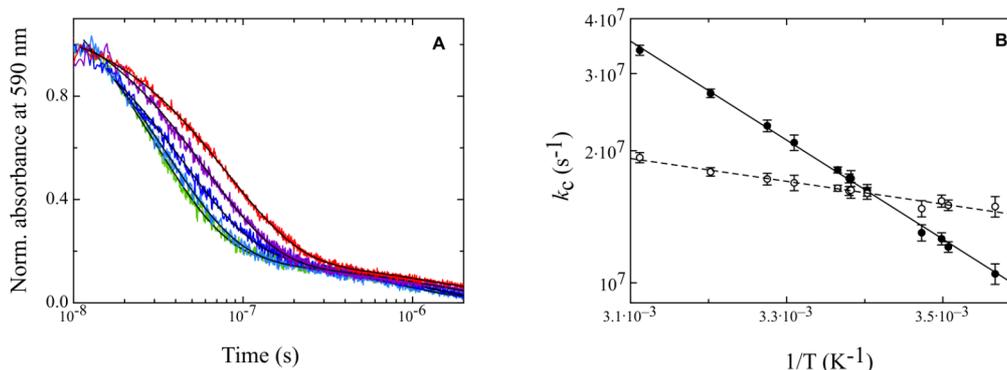
To characterize the barriers for intrachain loop formation in the different limits of formation of short and long loops in more detail we applied TTET from xanthone to naphthylalanine attached to different polypeptide chains and measured the effect of temperature on end-to-end loop formation. To determine the effect of chain length on the barriers for loop formation we studied homopolypeptides consisting of poly(Gly-Ser) or polyserine groups. The general form of the used model polypeptides is shown in figure 11. Figure 21A shows the observed transient triplet absorbance decay in a (Gly-Ser)<sub>14</sub> chain, which reveals a significant acceleration of loop formation with increasing temperature. The absorbance decay can be described by a double exponential function with the slower kinetic phase corresponding to the intrinsic xanthone lifetime  $\tau = 1.5(\pm 0.1)$   $\mu\text{s}$ . The faster kinetic phase can be attributed to TTET and corresponds to the rate constant  $k_c$  for end-to-end loop formation. Varying the temperature changes the solvent viscosity with an apparent activation energy of  $E_{a,\text{app}} = 17$  kJ/mol. Therefore, the observed rate constant  $k_c$  measured at different temperatures, has to be corrected against viscosity according to

$$k'_c = k_c \left( \frac{\eta}{\eta_0} \right)^{-\beta} \quad (31)$$

using the  $\beta$ -values determined from intramolecular TTET measurements in the presence of glycerol (Figure 21B; see chapter 3.1.1.1). Increasing the temperature speeds up both the observed rate constants  $k_c$  and the viscosity-corrected rate constants  $k'_c$  in all studied peptides. The data can be described by the Arrhenius equation:

$$k'_c = A \cdot e^{\frac{-E_a}{RT}} \quad (32)$$

where  $A$  is the Arrhenius pre-exponential factor and  $E_a$  is the Arrhenius activation energy. The observation that  $k_c$  shows linear Arrhenius behaviour indicates a change in heat capacity of  $\Delta C_p^\ddagger \approx 0$  for loop formation.



**Figure 21:** A: Transient triplet absorbance decay monitored at 590 nm in a (Gly-Ser)<sub>14</sub> chain in 10 mM KP pH 7.0 at T=10 °C (—), 20 °C (—), 30 °C (—), 40 °C (—) and 50 °C (—). The continuous lines represent fits of double exponential functions to the data. B: Observed (●) and viscosity-corrected (○) rate constants for end-to-end loop formation in (Gly-Ser)<sub>14</sub>. The lines represent a fit of equation 32 to the data. Measurements were carried out by F. Krieger.<sup>140</sup>

Comparing the results from different (Gly-Ser)<sub>n</sub> chains with *n* ranging from 1 to 14 shows that the activation energy reaches a limiting-value around 5 kJ/mol for long loops with *n* > 8 and increases for shorter loops with a value of 13.9 kJ/mol for (Gly-Ser)<sub>1</sub> (Figure 22A). Similarly, the activation energy decreases from 16.1(±0.3) kJ/mol for a (serine)<sub>2</sub> peptide to 9.2(±1.0) kJ/mol for a (serine)<sub>9</sub> peptide (Figure 22A). Due to limitation in peptide synthesis the activation energies for longer polyserine peptide chains could not be determined. For peptides with the same number of amino acids between the triplet labels, the activation energy is increased in polyserine chains compared to the poly(Gly-Ser) chains (Figure 22A). An explanation for lower activation energies in poly(Gly-Ser) peptides compared to the poly(serine) peptides is decreased chain stiffness, since a glycy residue increases the flexibility of a polypeptide chain compared to other amino acids.<sup>37</sup>

Comparing the general rate equation

$$k = k_0 \cdot e^{-\frac{\Delta G^{0\ddagger}}{RT}} = k_0 \cdot e^{-\frac{\Delta H^{0\ddagger}}{RT}} \cdot e^{-\frac{\Delta S^{0\ddagger}}{R}} \quad (5)$$

to the Arrhenius equation (Equation 32) shows that the enthalpic contribution (activation enthalpy  $\Delta H^{0\ddagger}$ ) to the free energy barrier can be determined from the Arrhenius activation energy according to

$$\Delta H^{0\ddagger} = E_a - RT \quad (33)$$

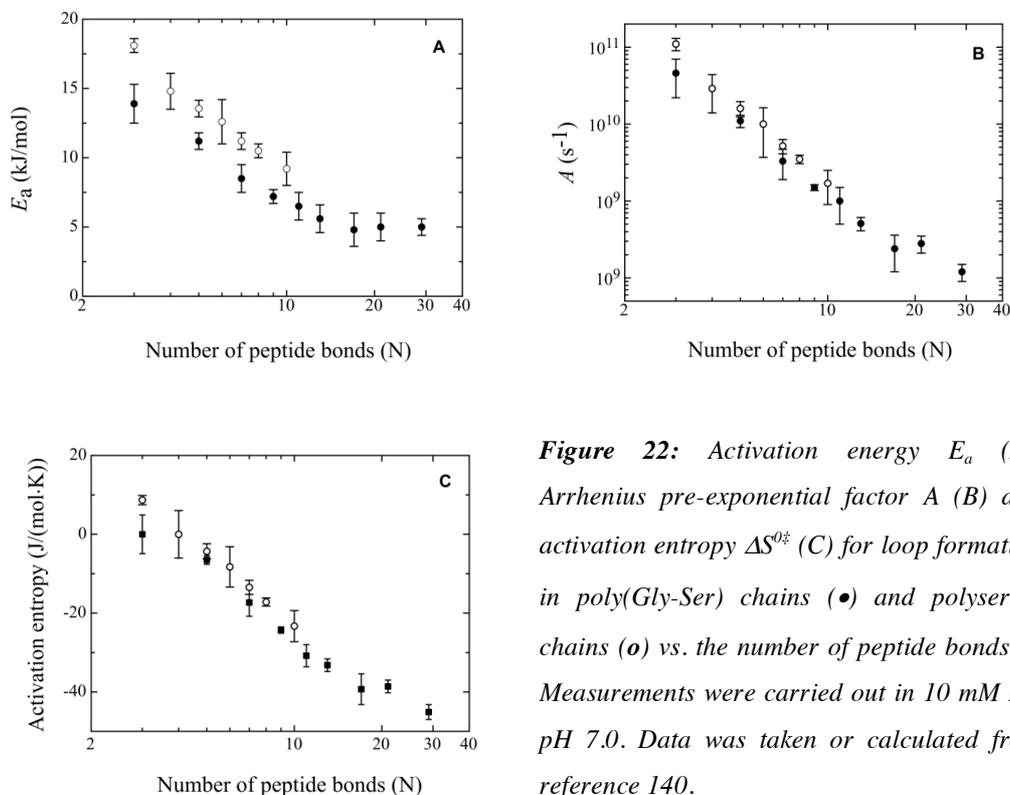
For a fully diffusion-controlled reaction  $\Delta H^{0\ddagger}$  should be close to 0. The results on the long poly(Gly-Ser) chains indicate a  $\Delta H^{0\ddagger}$ -value around 2.5 kJ/mol, which is close to zero, as expected for a purely diffusion-limited reaction. For shorter poly(Gly-Ser) chains and polyserine chains the enthalpic contributions to the barriers for intrachain loop-formation are larger with values of  $\Delta H^{0\ddagger} = 11.4$  kJ/mol for the (Gly-Ser)<sub>1</sub> peptide and  $\Delta H^{0\ddagger} = 15.6$  kJ/mol for the (serine)<sub>2</sub> peptide. These values are comparable to results from computational analysis of single bond rotation in  $\beta$ -turns.<sup>185</sup>

In addition to the activation energies and enthalpies the results from fitting the Arrhenius equation to the temperature-dependence of  $k'_c$  yields the Arrhenius pre-exponential factor of the reaction, which includes contributions from the activation entropy,  $\Delta S^{0\ddagger}$ , and from the pre-exponential factor,  $k_0$ , of the general rate-equation according to

$$A = k_0 \cdot e^{\frac{\Delta S^{0\ddagger} + R}{R}} \quad (34)$$

Figure 22B shows that the Arrhenius pre-exponential factor for loop formation decreases with increasing chain length for poly(Gly-Ser) and polyserine chains (Figure 22B). This behaviour is in accordance with a larger number of conformations adopted by longer chains, which leads to an increased loss in entropy upon loop formation. Comparing chains of the same number of peptide bonds between the triplet labels shows that poly(Gly-Ser) chains have larger pre-exponential factors compared to polyserine chains. This difference is expected since poly(Gly-Ser) is more flexible than polyserine and thus has a larger accessible conformational space and loses more entropy upon loop formation. For polyserine chains increased stiffness leads to a

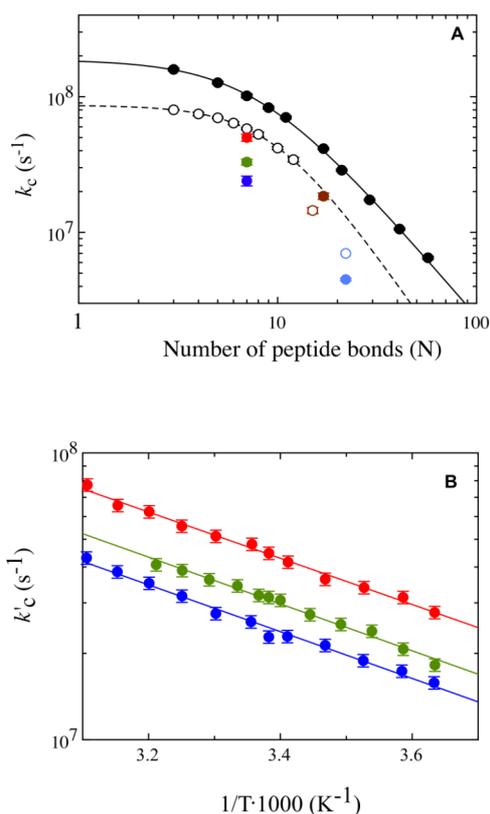
restricted conformational space, which results in less loss of entropy upon loop formation and results in a higher Arrhenius pre-exponential factor. This behaviour has previously been observed for formation of very short loops.<sup>55</sup> To estimate entropic contributions to the free energy barrier we attempted to dissect the Arrhenius pre-exponential factor into  $\Delta S^{0\dagger}$  and  $k_0$ , according to equation 34. Since the correct  $k_0$  is not known for loop formation, the determination of absolute values of  $\Delta S^{0\dagger}$  is not possible. However, it is reasonable to assume that  $k_0$  is on the order of the rate constant for a single bond rotation in a polypeptide chain, which is around  $10^{10} \text{ s}^{-1}$ . Figure 22C shows  $\Delta S^{0\dagger}$  as a function of loop length for all studied model polypeptide chains assuming  $k_0 = 10^{10} \text{ s}^{-1}$ . The results show that  $\Delta S^{0\dagger}$  is negative. This is expected since end-to-end loop formation should be entropically unfavourable.  $\Delta S^{0\dagger}$  is around 0 kJ/(mol·K) for the shortest peptides (Ser)<sub>2</sub> and (Gly-Ser)<sub>1</sub> indicating that entropy does not contribute to the free energy barrier for loop formation in these chains. This result also indicates that the value of  $10^{10} \text{ s}^{-1}$  assumed for  $k_0$  may be too low.



**Figure 22:** Activation energy  $E_a$  (A), Arrhenius pre-exponential factor  $A$  (B) and activation entropy  $\Delta S^{0\dagger}$  (C) for loop formation in poly(Gly-Ser) chains ( $\bullet$ ) and polyserine chains ( $\circ$ ) vs. the number of peptide bonds  $N$ . Measurements were carried out in 10 mM KP pH 7.0. Data was taken or calculated from reference 140.

To characterize the free energy barriers for loop formation in unstructured natural protein sequences we determined the temperature-dependence of loop formation in the same peptides used in the study of the viscosity-dependence of loop formation (see figure 18). The results from the viscosity-dependence of loop formation in these peptides described in chapter 3.1.1.1 were used to correct the rate constants for the effect of temperature on solvent viscosity according to equation 31. The results from the temperature dependence of loop formation in the natural protein sequences are summarized in table 3. For all natural protein sequences we found significant activation energies in the range of 12 - 16 kJ/mol. Loop formation in the wild type GB1<sub>45-52</sub>-peptide has an activation energy of  $E_a = 15.7(\pm 0.9)$  kJ/mol, which is around 4 kJ/mol higher than for a polyserine chain of the same length. The Arrhenius pre-exponential factor for GB1<sub>45-52</sub> is  $A = 1.8(\pm 0.7) \cdot 10^{10} \text{ s}^{-1}$ , which is significantly larger compared to a polyserine chain of the same length, indicating a decreased entropic cost for loop formation in GB1<sub>45-52</sub>. For loop formation in the T49G GB1 variant  $E_a = 15.5(\pm 0.7)$  kJ/mol and  $A = 2.4(\pm 0.7) \cdot 10^{10} \text{ s}^{-1}$  were determined. Loop formation in the T49P variant yielded  $E_a = 15.7(\pm 0.8)$  kJ/mol and  $A = 1.4(\pm 0.4) \cdot 10^{10} \text{ s}^{-1}$  (Figure 23B). These results indicate that mutations do not significantly affect the enthalpic contribution and the entropic contribution to the free energy barrier for loop formation. In the native protein G this amino acid sequence forms the loop region of a tight  $\beta$ -hairpin. The reduced entropic cost for loop formation in these peptides indicates the presence of energetically favourable loop conformations in these sequences. Population of loop conformations in equilibrium is in accordance with the large observed loss of xanthone triplet absorbance in the experimental dead time (Figure 19A). For the sequences derived from  $\beta$ -carp parvalbumin we found  $E_a = 14.3(\pm 0.7)$  kJ/mol and  $A = 4.6(\pm 1.9) \cdot 10^9 \text{ s}^{-1}$  for the DE-loop and  $E_a = 12.5(\pm 0.7)$  kJ/mol and  $A = 2.7(\pm 0.3) \cdot 10^9 \text{ s}^{-1}$  for the EF-loop, respectively. This indicates increased enthalpic barriers and decreased entropic cost for loop formation in both sequences compared to unfolded model polyserine chains of the same length. The EF-loop shows a lower activation energy and a lower Arrhenius pre-exponential factor compared to the DE loop, which can be explained by higher chain flexibility in the EF-loop due to a larger number of glycine residues. For the sequence derived from the naturally unfolded brinker protein we found  $E_a = 13.3(\pm 1.0)$  kJ/mol and  $A =$

$9.8(\pm 3.7) \cdot 10^8 \text{ s}^{-1}$ . This indicates increased enthalpic barriers and decreased entropic cost for loop formation in  $\text{Brk}_{66-89}$  compared to an unfolded model polyserine chain of the same length. This can be explained by repulsive interactions between positively charged residues in  $\text{Brk}_{66-89}$ , which might lead to increased internal friction and/or decreased chain flexibility in this sequence (see chapter 3.1.2).



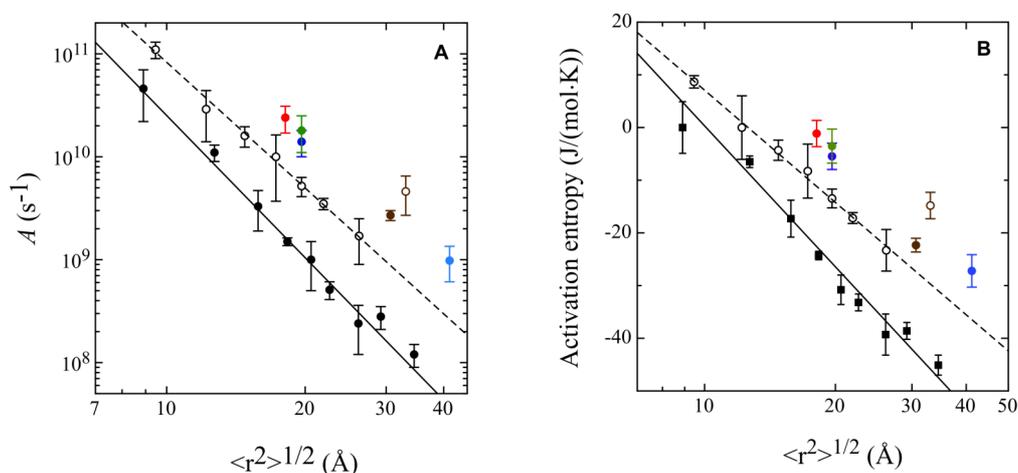
**Figure 23:** A: Rate constants ( $k_c$ ) for end-to-end loop formation in natural protein sequences and model polypeptides: poly(Gly-Ser) (●), polyserine (○),  $\text{GBI}_{45-52}$  (●),  $\text{GBI}_{45-52}$  T49G (●),  $\text{GBI}_{45-52}$  T49P (●), DE-loop (○), EF-loop (●),  $\text{Brk}_{66-89}$  (●) and  $\text{Brk}_{66-89}$  in the presence of 300 mM NaCl (○). B: Effect of temperature on end-to-end loop formation in  $\text{GBI}_{45-52}$  (●),  $\text{GBI}_{45-52}$  T49G (●) and  $\text{GBI}_{45-52}$  T49P (●). Results for poly(Gly-Ser) and polyserine were taken from reference 47. Data on natural protein sequences was partly taken from references 53 and 140 (see table 3). Measurements were carried out in 10 mM KP pH 7.0 ( $\text{GBI}$ -peptides and homopolypeptides), in 10 mM NaCac pH 7.0 (parvalbumin fragments) or in 20 mM NaAc 20 mM NaCl, pH 5.7 ( $\text{Brk}_{66-89}$ ) at  $T = 22.5 \text{ }^\circ\text{C}$ .

Fragment	$k_0$ (H <sub>2</sub> O) (s <sup>-1</sup> )	$E_a$ (kJ)	$\Delta H^{0\ddagger}$ (kJ/mol)	$A$ (s <sup>-1</sup> ) <sup>C</sup>	$\Delta S^{0\ddagger}$ (J/(mol·K)) <sup>C</sup>
GB1 <sub>45-52</sub>	$3.3(\pm 0.2) \cdot 10^7$	$15.7(\pm 0.9)^A$	$13.2(\pm 0.9)^A$	$1.8(\pm 0.7) \cdot 10^{10}^A$	$-3.5(\pm 3.2)$
GB1 <sub>45-52</sub> T49G	$5.0(\pm 0.3) \cdot 10^7^A$	$15.5(\pm 0.7)^A$	$13.0(\pm 0.7)^A$	$2.4(\pm 0.7) \cdot 10^{10}^A$	$-1.1(\pm 2.5)$
GB1 <sub>45-52</sub> T49P	$2.4(\pm 0.2) \cdot 10^7^A$	$15.7(\pm 0.8)^A$	$13.2(\pm 0.8)^A$	$1.4(\pm 0.4) \cdot 10^{10}^A$	$-5.5(\pm 2.5)$
PA DE-loop	$1.5(\pm 0.1) \cdot 10^7^A$	$14.3(\pm 0.7)^A$	$11.8(\pm 0.7)^A$	$4.6(\pm 1.9) \cdot 10^9^A$	$-14.8(\pm 2.5)$
PA EF-loop	$1.9(\pm 0.1) \cdot 10^7^B$	$12.5(\pm 0.7)^A$	$10.0(\pm 0.7)^A$	$2.7(\pm 0.3) \cdot 10^9^A$	$-22.3(\pm 1.3)$
Brk <sub>66-89</sub>	$4.5(\pm 0.1) \cdot 10^6$	$13.3(\pm 1.0)$	$10.8(\pm 1.0)$	$9.8(\pm 3.7) \cdot 10^8$	$-27.2(\pm 3.1)$
Brk <sub>66-89</sub> +salt	$7.0(\pm 0.1) \cdot 10^6$	-	-	-	-

**Table 3:** Results for loop formation in unstructured natural protein sequences.  $k_0$  (H<sub>2</sub>O) is the rate constant determined in water at  $T=22.5^\circ\text{C}$ . The activation entropy  $\Delta S^{0\ddagger}$  was determined assuming  $k_0 = 100$  ps. A: Data taken from reference 140, B: Data taken from reference 53. C: obtained from a fit of equation 5 to the temperature dependence of  $k_c'$ .

In summary, the results on the effect of temperature on end-to-end loop formation in natural protein sequences indicate significantly increased activation energies compared to polyserine chains. A possible explanation for increased activation energies in natural sequences compared to polyserine chains are longer side chains, which might lead to steric contributions to the enthalpic barrier for loop formation. The activation enthalpies  $\Delta H^{0\ddagger}$  for the natural protein sequences are in the range of 10.0 - 13.3 kJ/mol (Table 3), which is comparable to  $\Delta H^{0\ddagger}$  observed for the collapse in cytochrome c,<sup>186</sup> which indicates that the dynamics of loop formation determine the rate constant for fast collapse early in folding. The Arrhenius pre-exponential factors determined for loop formation in natural protein sequences are significantly increased compared to results on polyserine chains (Figure 24A). This indicates that the entropic contribution ( $\Delta S^{0\ddagger}$ ) to the free energy barrier for loop formation is lower in the natural protein sequences than in polyserine chains of the same length. For loop formation in natural protein sequences,  $\Delta S^{0\ddagger}$ -values using  $k_0 = 10^{10} \text{ s}^{-1}$  are summarized in table 3. Entropic contributions ( $\Delta S^{0\ddagger}$ ) to the free energy barrier for end-to-end loop formation correlate with the Flory mean square end-to-end distance  $\langle r^2 \rangle^{1/2}$  of a polypeptide chain and thus more compact conformations lead to a higher pre-exponential factor (Figure 24B). Poly(Gly-Ser) chains show the most

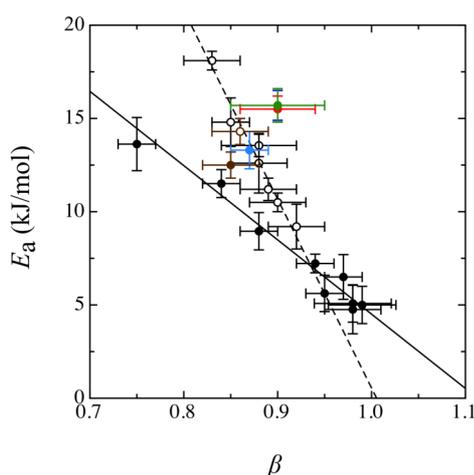
unfavourable activation entropies. This is expected since these chains can adopt a relatively large number of conformations due to high chain flexibility. Possible explanations for higher pre-exponential factors in natural protein sequences compared to polyserine peptides are side chain interactions, which might lead to more compact conformations and/or excluded volume effects, which decrease the accessible conformational space. The results show that free energy barriers for formation of short loops and for loop formation in natural protein sequences contain significant enthalpic contributions whereas free energy barriers for formation of long loops in flexible poly(Gly-Ser) model polypeptide chains contain mainly entropic contributions.



**Figure 24:** Pre-exponential factor (A) and activation entropy (B) for end-to-end contact formation vs. the mean square end-to-end distance  $\langle r^2 \rangle^{1/2}$ . The colours are as indicated in figure 23. Experimental conditions are as described in figure 23.

The observed barriers for end-to-end loop formation kinetics in short poly(Gly-Ser) and polyserine peptides are in accordance with the fractional viscosity-dependences ( $\beta < 1$ ) for loop formation in these chains. This comparison shows that formation of short loops is not fully diffusion-controlled. For long poly(Gly-Ser) chains the activation energy is in the order of the thermal energy, which is in agreement with the inverse viscosity-dependence of the rate constant for loop formation ( $\beta = 1$ ) and indicates that formation of long loops in flexible poly(Gly-Ser) chains is fully diffusion-controlled. Comparison of the results on the effect of temperature and of viscosity on loop closure kinetics reveals a negative correlation between the activation energy and the sensitivity of the reaction to solvent viscosity ( $\beta$ ) for both poly(Gly-

Ser) and polyserine chains (Figure 25). Larger activation energies for loop formation lead to smaller  $\beta$ -values. However, the increase in activation energy with decreasing  $\beta$  is weaker in flexible poly(Gly-Ser) chains compared to stiffer polyserine chains. The results on natural protein fragments from carp muscle  $\beta$ -parvalbumin (DE-loop and EF-loop) and from the natively unfolded brinker protein (Brk<sub>66-89</sub>) fit well on the correlation observed for polyserine chains. All studied GB1 fragments (GB1<sub>45-52</sub> and the two variants T49G and T49P) do not fit to the correlation observed for polyserine chains. The GB1 fragments show similar  $\beta$ -values as a polyserine chain of the same length, however the activation energy in the hairpin fragment is significantly increased. This could be due to more compact chain conformation in the hairpin fragment, which leads to steric barriers, in accordance with a larger pre-exponential factor.



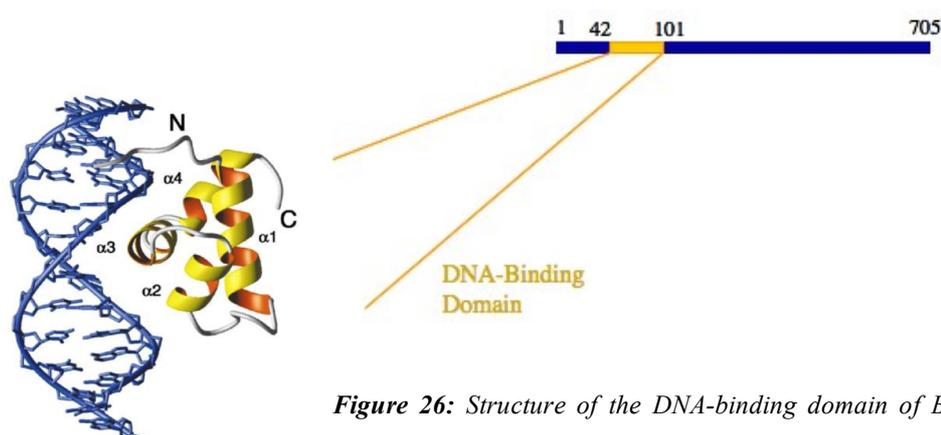
**Figure 25:** Correlation between the activation energy  $E_a$  for loop formation and the sensitivity of the reaction to solvent viscosity ( $\beta$ ). The colours are as indicated in figure 23. Data for homopolypeptides and partly for natural sequences were adapted from reference 140 (see table 3).

The observed negative correlation between  $E_a$  and  $\beta$  indicates that crossing of free energy barriers for loop formation, which contain a large enthalpic contribution ( $\Delta H^{0\ddagger}$ ) is not fully driven by temperature dependent Brownian motion ( $\beta < 1$ ) and thus the reaction rate constant is influenced by the barrier height ( $\Delta G^{0\ddagger}$ ) and the pre-exponential factor  $k_0$ . Free energy barriers for loop formation that contain mainly entropic contributions, as it is observed for long flexible poly(Gly-Ser) chains, can be overcome by Brownian motion, indicated by the inverse dependence of the rate constant on solvent viscosity ( $\beta = 1$ ). These results indicate that mainly the enthalpic contribution to the free energy barrier of loop formation determines whether the reaction is diffusion-controlled or not.

Our system used to study the effect of solvent viscosity and temperature on loop formation dynamics is diffusion-controlled (see chapter 3.1.1), indicated by an activation energy of  $E_a \approx RT$  for TTET from xanthone to naphthalene. This model is supported by an inverse dependence of the bimolecular rate constant for TTET from xanthone to naphthylalanine on solvent viscosity ( $\beta=1$ ). Thus, TTET allows determination of absolute rate constants, viscosity-dependencies and free energy barriers for end-to-end loop formation. This is in contrast to previous systems used to characterize the rate-limiting steps for end-to-end loop formation in polypeptide chains (see chapter 1.3.3). These systems are either not diffusion-controlled, contain contributions from intrinsic barriers for photophysics or are not well characterized. In a detailed study on the effect of solvent viscosity and temperature on end-to-end loop formation in flexible poly(Gly-Ser) chains we observed different regimes for intrachain motions over short and long distances. This indicates different rate-limiting steps for formation of short loops and long loops. Formation of short loops is not-diffusion controlled but is limited by intrinsic barriers, which contain mainly enthalpic contributions. Intrinsic barriers for formation of long loops contain mainly entropic contributions and can be overcome by temperature dependent Brownian motion, indicating that loop formation in these chains is fully diffusion-controlled. A similar behaviour was observed for stiffer polyserine chains. However, activation energies for loop formation in these chains are slightly increased compared to poly(Gly-Ser) chains, which can be explained by increased chain stiffness. The entropic contributions to the free energy barriers are decreased in polyserine chains compared to flexible poly(Gly-Ser) chains, which can be explained by a restricted conformational space for polyserine chains resulting in less loss of entropy upon loop formation. Free energy barriers for loop formation in natural protein sequences contain increased enthalpic contributions but decreased entropic contributions, compared to the studied homopolypeptide chains. Longer side chains in the natural protein sequences might lead to steric contributions to the enthalpic barrier for loop formation. Possible explanations for higher pre-exponential factors in natural protein sequences compared to polyserine peptides are side chain interactions, which might lead to more compact conformations and/or excluded volume effects, which decrease the accessible conformational space.

### 3.1.2 Dynamics of loop formation in a fragment from the brinker protein

Characterization of dynamics of unfolded polyserine and poly(glycine-serine) model polypeptide chains by TTET allowed insight into physical basis of chain dynamics. Compared to these model peptides biological sequences are more complex as they consist of 20 different amino acids. Usually, very harsh conditions are used to unfold proteins *in vitro* what makes characterization of the unfolded state under physiological conditions difficult. Natively unfolded proteins provide good model systems to study the unfolded state under physiological conditions. They exist as an ensemble of unstructured conformations and only adopt a well-defined three-dimensional structure in the presence of their binding partner, like DNA or other proteins. To measure end-to-end loop formation in a naturally occurring unstructured amino acid sequence we applied TTET to a highly charged fragment from the natively unfolded brinker protein. The brinker protein is a nuclear repressor playing an important role in the *Drosophila* Decapentaplegic morphogen pathway. The full-length protein consists of 704 amino acids and contains an N-Terminal sequence-specific DNA-binding domain (Figure 26). In solution the DNA-binding domain is globally unstructured and highly flexible. Binding to cognate DNA induces folding to an ordered structure consisting of four  $\alpha$ -helices forming a helix-turn-helix DNA recognition motif.<sup>187</sup>



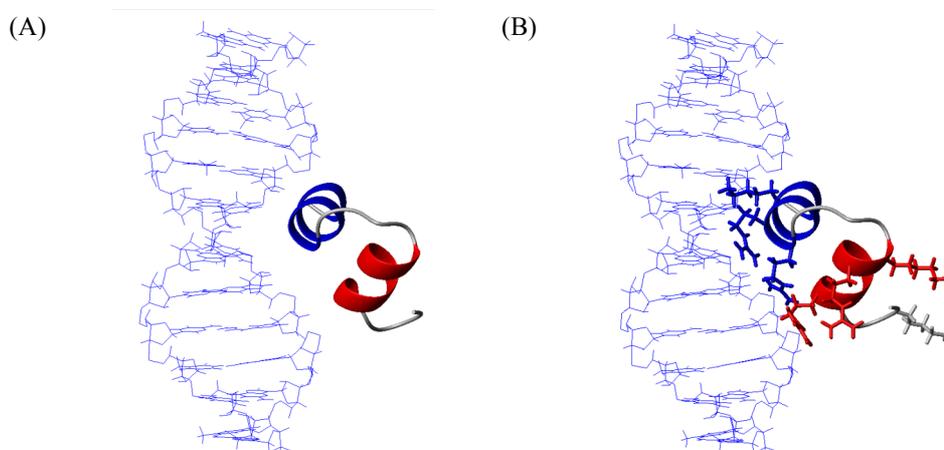
**Figure 26:** Structure of the DNA-binding domain of Brinker in complex with DNA. Ribbon diagram of the lowest energy structure taken from reference 187.

We used a fragment consisting of amino acids 66-89 of the brinker protein (Brk<sub>66-89</sub>) to investigate chain dynamics by TTET. This 24 amino acid fragment includes the sequence of helixII and helixIII from the DNA-binding motif (Figure 27A). In the

absence of DNA  $\alpha$ -helix formation in this segment is prevented by repulsive  $i, i+4$  interactions between lysine and arginine side chains. These interactions are screened in the presence of negatively charged phosphate groups from the DNA backbone, which allows  $\alpha$ -helix formation (Figure 27B).

**C<sub>66</sub>KGNQRATARKYNIHRRQIQKWLQ<sub>89</sub>**

**Figure 27A:** Wild-type amino acid sequence of the fragment Brk<sub>66-89</sub>. Residues included in helixII and helixIII in the folded conformation are shown in red and blue, respectively.

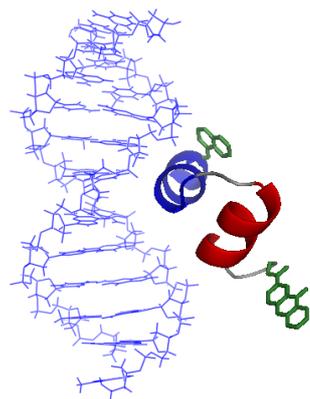


**Figure 27B:** In the presence of DNA repulsion of positively charged amino acids of Brk<sub>66-89</sub> (shown in stick representation in panel B) is screened, which allows  $\alpha$ -helix formation in this sequence. The figures were prepared using MacPyMOL and the pdb file 2GLO<sup>180</sup>.

In order to study chain dynamics in Brk<sub>66-89</sub> residues Tyr<sub>77</sub> and His<sub>80</sub> were mutated to phenylalanine since these amino acids interact with the xanthone triplet state. To study end-to-end loop formation the triplet donor xanthone (Xan) was attached to the N-terminus and the triplet acceptor naphthyl was introduced replacing Trp at position 87 by the non-natural amino acid naphthylalanine (Nal) (Figure 28). In the following Brk<sub>66-89</sub> denotes the modified sequence suitable for TTET measurements.

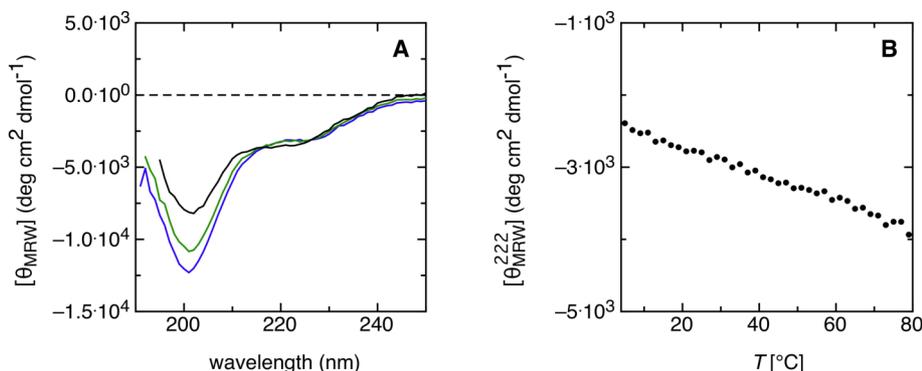
**Xan-G<sub>66</sub>KGNQRATARKFNIFRRQIQKNalLQ<sub>89</sub>**

**Figure 28A:** Amino acid sequence of Brk<sub>66-89</sub> used for TTET measurements. TTET labels and introduced mutations are in green. Residues included in helixII and helixIII in the folded conformation are shown in red and blue, respectively.



**Figure 28B:** *Brk<sub>66-89</sub>* labelled for TTET: Tryptophane at position 87 is replaced for naphthylalanine and xanthone is attached to the N-terminus of glycine at position 66. TTET labels are shown in green and in stick representation. The figure was prepared using MacPyMOL and the pdb file 2GLO<sup>180</sup>.

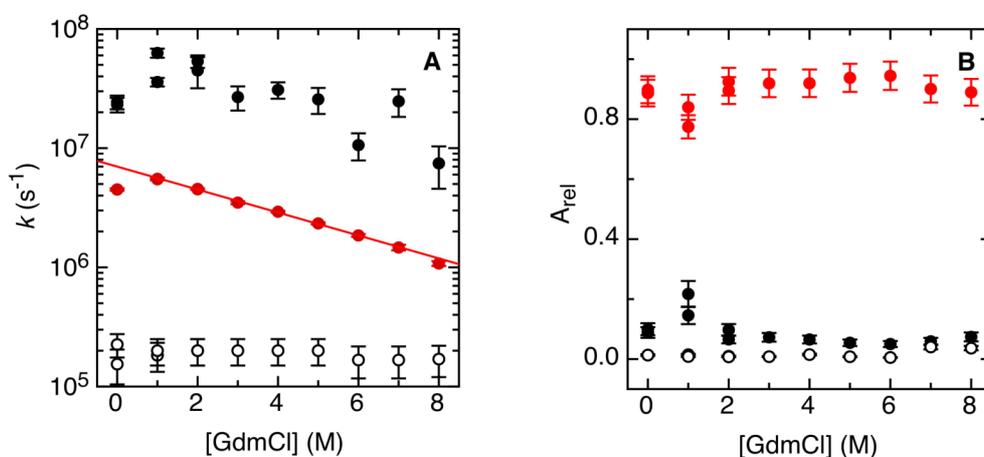
In the presence of 20mM NaAc, 20mM NaCl, pH 5.7 *Brk<sub>66-89</sub>* is unfolded, indicated by a minimum of the CD signal around 200 nm (Figure 29A). This is further supported by a thermal transition monitored at 222 nm, which shows only a small linear decrease in the CD signal with increasing temperature (Figure 29B), as observed for unfolded polypeptide chains.<sup>188</sup>



**Figure 29:** *A:* CD spectra of *Brk<sub>66-98</sub>* in 20 mM NaAc 20 mM NaCl, pH 5.7 at 5.0 °C (blue), 22.5 °C (green) and 80.0 °C (black) and *B:* Thermal transition of *Brk<sub>66-98</sub>* in 20 mM NaAc 20 mM NaCl, pH 5.7 monitored by the change in CD signal at 222 nm.

The triplet absorbance decay observed for *Brk<sub>66-89</sub>* in the presence of 20mM NaAc, 20mM NaCl, pH 5.7 at  $T = 22.5$  °C can be described by a triple exponential function (Figure 30). The slowest kinetic phase corresponds to the intrinsic xanthone lifetime with  $\tau_{\text{int}} = 1/k_{\text{int}} = 1/(1.8(\pm 0.2) \cdot 10^5 \text{ s}^{-1}) = 5.5 \pm (0.5) \mu\text{s}$  and contributes with  $2(\pm 2)\%$  to the total amplitude of the triplet absorbance decay. The second kinetic phase can be attributed to end-to-end loop formation in *Brk<sub>66-89</sub>*. It accounts for  $86(\pm 9)\%$  of the

overall amplitude and reveals a rate constant for end-to-end loop formation  $k_c$  of  $4.5(\pm 0.1) \cdot 10^6 \text{ s}^{-1}$ . The third kinetic phase has a rate constant of  $2.3(\pm 0.3) \cdot 10^7 \text{ s}^{-1}$  and accounts for  $9(\pm 5)\%$  of the overall amplitude. The origin of this phase is still unclear. The determined rate constant for end-to-end loop formation is significantly lower than the expected value from results on polyserine chains.<sup>47</sup> This is in contrast to recent results on intrachain dynamics in other unstructured biological protein fragments, which revealed rate constants comparable to the expected value from results on polyserine chains.<sup>53,54</sup> Possible explanations for the slow dynamics in Brk<sub>66-89</sub> are increased chain dimension and/or increased internal friction due to repulsive interactions between positively charged residues. This model is supported by measurements in the presence of GdmCl, which screens electrostatic forces (Figure 30). Loop formation kinetics in Brk<sub>66-89</sub> linearly depend on the GdmCl concentration extrapolating to  $7.0(\pm 0.1) \cdot 10^6 \text{ s}^{-1}$  in the absence of denaturant, which is close to the expected value from results on polyserine chains.



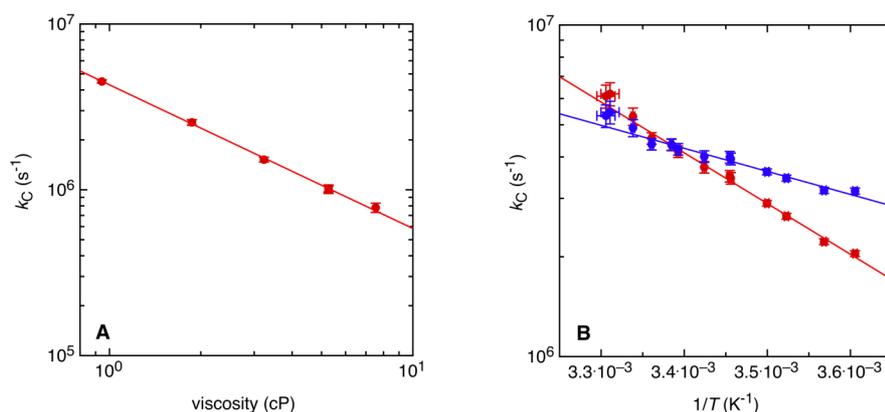
**Figure 30:** Rate constants (A) and corresponding relative amplitudes (B) as a function of the GdmCl concentration obtained by a triple exponential fit to the observed triplet absorbance decay in Brk<sub>66-89</sub>. TET measurements were carried out in 20mM NaAc, 20mM NaCl, pH 5.7 at  $T = 22.5^\circ\text{C}$ . The solid line represents a fit of equation 35 to the rate constant for end-to-end loop formation in the range of 1.0 to 8.0 M GdmCl.

The GdmCl-dependence of  $k_c$  can be described by

$$\ln k_c = \ln k_c^0 - \frac{m_c[D]}{RT} \quad (35)$$

Here  $k_c^0$  is the rate constant for end-to-end loop formation at zero denaturant concentration,  $k_c$  is the rate constant at the denaturant concentration  $[D]$  and  $m_c$  describes the sensitivity of  $k_c$  to the presence of denaturant. The results reveal a  $m_c$ -value of  $541(\pm 4)$  (J/mol)/M, which is comparable to values from results on poly(glycine-serine) chains of similar length.<sup>189</sup> The observation that the third kinetic phase does not disappear in the presence of high denaturant concentrations ( $k_3 = 4(\pm 3) \cdot 10^7 \text{ s}^{-1}$ ) indicates that it is not related to protein aggregation.

To study the origin of the slow kinetics of loop formation in Brk<sub>66-89</sub> in more detail we investigated the effect of viscosity and of temperature on  $k_c$  in the presence of 20 mM NaAc 20 mM NaCl pH 5.7 (Figure 31). A fit of equation 29 to the rate constants determined at different solvent viscosities yields  $\beta = 0.87(\pm 0.02)$ , which is low compared to a poly(glycine-serine) chain of similar length and compared to the value expected for a polyserine chain of similar length (Table 4 and chapter 3.1.1.1). The fractional dependence of  $k_c$  on solvent viscosity indicates that loop formation in Brk<sub>66-89</sub> is not fully diffusion-controlled. The temperature dependence of  $k_c$  can be described by the Arrhenius equation (Equation 32), which reveals a pre-exponential factor of  $A = 9.8(\pm 3.7) \cdot 10^8 \text{ s}^{-1}$  and an activation energy of  $E_a = 13.3(\pm 1.0) \text{ kJ/mol}$  for loop formation in Brk<sub>66-89</sub>. Both values are significantly increased compared to the results for a poly(glycine-serine) chain of similar length and compared to the values expected for a polyserine chain of similar length (Table 4 and chapter 3.1.1.2). This result indicates significant contributions from intrinsic barriers to slow loop formation in Brk<sub>66-89</sub>. On the other hand, the large Arrhenius pre-exponential factor indicates a decreased entropic cost for loop formation in Brk<sub>66-89</sub> compared to poly(glycine-serine) and polyserine chains. For poly(glycine-serine) chains and polyserine chains we found a negative correlation between the activation energy and the sensitivity of loop formation to solvent viscosity (see chapter 3.1.1.2). The results for Brk<sub>66-89</sub> fit to the correlation observed for polyserine chains (see chapter 3.1.1.2).

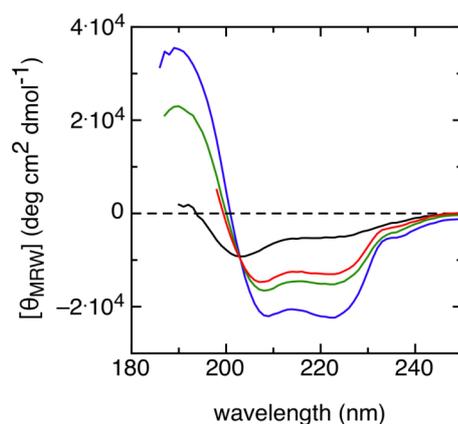


**Figure 31:** Effect of viscosity (A) and temperature (B) on the dynamics of loop formation in  $Brk_{66-89}$  in 20 mM NaAc 20 mM NaCl pH 5.7. The viscosity was adjusted by adding glycerol. A fit of equation 29 to the viscosity dependence of  $k_c$  yields  $\beta = 0.87 \pm 0.02$ , which was used to correct  $k_c$  measured at different temperatures (red) against changes in viscosity (blue). A fit of equation 32 to the viscosity-corrected data yields  $A = 9.8(\pm 3.7) \cdot 10^8 \text{ s}^{-1}$  and  $E_a = 13.3(\pm 1.0) \text{ kJ/mol}$ .

peptide	$A \text{ (s}^{-1}\text{)}$	$E_A \text{ (kJ/mol)}$	$\beta$
$Brk_{66-89}$	$9.8(\pm 3.7) \cdot 10^8$	13.3(1.0)	$0.87(\pm 0.02)$
$GS_{10}$	$2.8(\pm 0.7) \cdot 10^8$	5.0( $\pm 1.0$ )	$0.98(\pm 0.04)$
$GS_{14}$	$1.2(\pm 0.3) \cdot 10^8$	5.0( $\pm 0.6$ )	$0.99(\pm 0.04)$

**Table 4:** Effect of temperature and solvent viscosity on end-to-end loop formation dynamics in  $Brk_{66-89}$  in 20mM NaAc 20 mM NaCl pH 5.7 and comparison to results on  $(\text{glycine-serine})_n$  chains of similar length.

The presence of inorganic phosphate ( $P_i$ ) was shown to induce structure formation in the DNA-binding domain of the brinker protein by mimicking the negative charges of the DNA backbone. Similarly, structure formation in  $Brk_{66-89}$  is induced, indicated by a maximum of ellipticity at 190 nm and two minima at 208 nm and 222 nm (Figure 32). A similar effect is observed in the presence of NaCl (Figure 32).



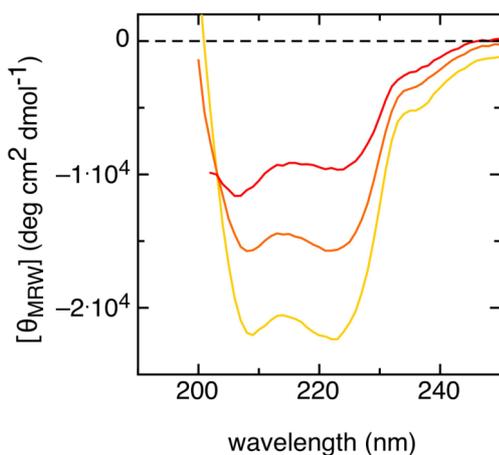
**Figure 32:** CD spectra of *Brk*<sub>66-98</sub> in 100 mM potassium phosphate (KP), pH 7.0 at 5.0 °C (blue), 22.5 °C (green) and 80.0 °C (blue) and in the presence of 20 mM NaAc, 300 mM NaCl pH 5.7 at 22.5 °C (red).

In order to quantify the formation of secondary structure in the presence of inorganic phosphate and in the presence of NaCl we calculated the helicity  $f_h$  of *Brk*<sub>66-89</sub> based on the mean residual ellipticity at 222 nm ( $[\theta]_{222}$ ) (Table 5 and material and methods). In the presence of 20mM NaAc, 20mM NaCl, pH 5.7 the CD spectrum of *Brk*<sub>66-89</sub> indicates an unstructured polypeptide chain. However, the calculation yields a helix propensity of 7-13 % under these conditions. The deviation can be explained by contributions of the TTET labels to  $[\theta]_{222}$ , which indicates that all calculated values for the helix propensity in *Brk*<sub>66-89</sub> are slightly too large.

Buffer	$f_{h,5.0^\circ\text{C}}$ (%)	$f_{h,22.5^\circ\text{C}}$ (%)	$f_{h,80.0^\circ\text{C}}$ (%)
20 mM NaAc, 20 mM NaCl pH5.7	12.9	12.4	8.0
100 mM KP pH 7.0	61.8	47.0	16.7
20 mM NaAc, 300 mM NaCl pH5.7	-	40.7	-

**Table 5:** Helix content obtained from CD spectroscopy on *Brk*<sub>66-89</sub>.

The helix content of *Brk*<sub>66-89</sub> in the presence of inorganic phosphate is reduced when urea is added (Figure 33 and table 6).

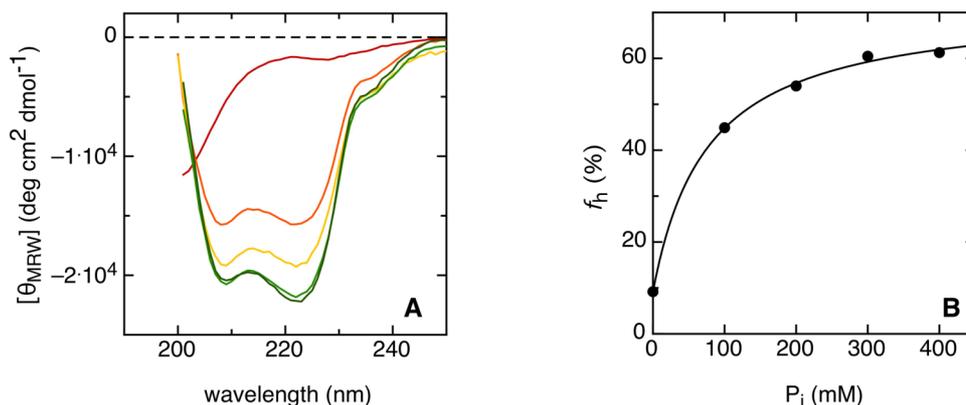


**Figure 33:** CD spectra of *Brk*<sub>66-98</sub> at  $T = 5^\circ\text{C}$  in the presence of 100 mM phosphate (yellow), 100 mM KP + 0.5 M (orange) and 100 mM KP + 1 M urea (red).

Buffer	$f_{h,5.0^\circ\text{C}}$ (%)
20 mM NaAc, 20 mM NaCl pH5.7	12.9
20 mM NaAc, 20 mM NaCl pH5.7 +0.5 M urea	9.2
100 mM KP pH 7.0	61.8
100 mM KP pH 7.0 + 0.5M urea	44.9
100 mM KP pH 7.0 + 1.0M urea	29.2

**Table 6:** Reduction of helix content of *Brk*<sub>66-89</sub> in the presence of urea at  $T = 5^\circ\text{C}$ .

In the presence of 0.5 M urea at  $5^\circ\text{C}$ , the effect of secondary structure formation in the presence of inorganic phosphate is saturated at 300 mM  $\text{P}_i$  (Figure 34 and table 7). Again one has to take into account that calculated helix propensities are somewhat too large due to contribution of the TTET labels to  $[\theta]_{222}$ . In the presence of 0.5 M urea, the spectrum of *Brk*<sub>66-89</sub> clearly indicates a random coil conformation in the absence of inorganic phosphate (Figure 34) while calculation yields a helix content of 9.2 %.



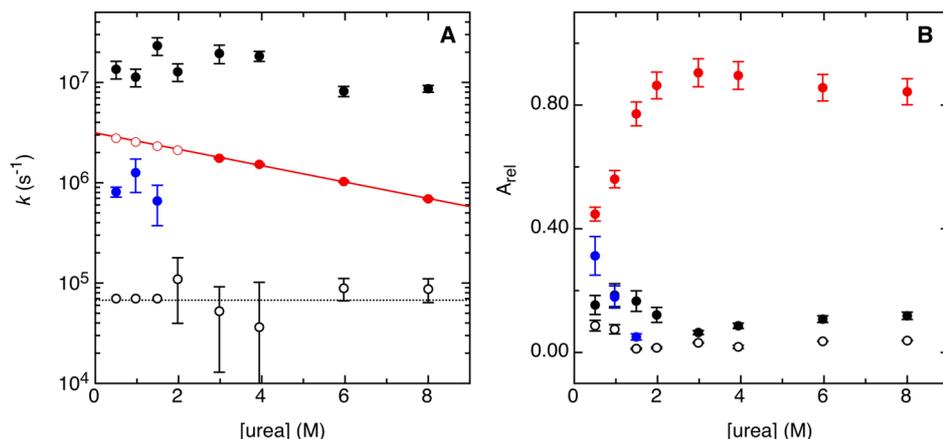
**Figure 34:** A: CD spectra of *Brk*<sub>66-98</sub> in 20 mM NaAc, 20 mM NaCl, pH 5.7 in the presence of 0.5 M urea at  $T = 5^\circ\text{C}$  (red). Helix formation is induced in the presence of inorganic phosphate:  $[P_i] = 100$  mM (orange), 200 mM (yellow), 300 mM (light green) and 400 mM (dark green). B: Helix propensity of *Brk*<sub>66-98</sub> in the presence of 0.5 M urea at  $T = 5^\circ\text{C}$  as a function of the inorganic phosphate concentration.  $f_h$  was determined from  $[\theta]_{222}$  as described in materials and methods. The line represents a fit of the Schellmann weak binding model to the data.

Buffer	$f_{h,5.0^\circ\text{C}}$ (%)
20 mM NaAc, 20 mM NaCl pH 5.7 + 0.5 M urea	9.2
100mM KP + 0.5 M urea	44.9
200mM KP + 0.5 M urea	54.0
300mM KP + 0.5 M urea	60.5
400mM KP + 0.5 M urea	61.3

**Table 7:** Helix content obtained from CD spectroscopy on *Brk*<sub>66-89</sub> in the presence of various concentrations of inorganic phosphate at  $T = 5^\circ\text{C}$ .

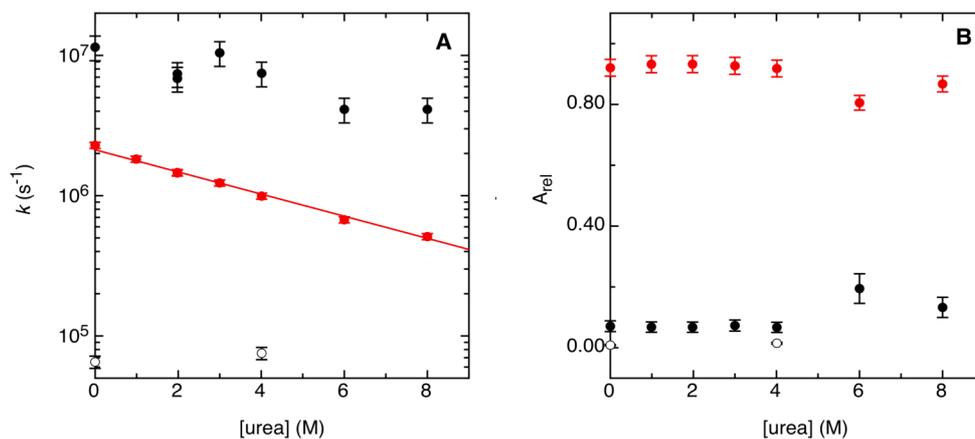
To study the effect of secondary structure formation on chain dynamics we performed TTET experiments in the presence of 300 mM potassium phosphate at  $5^\circ\text{C}$ . Above 2.0 M urea, the measured transient triplet absorbance decays can be described by a triple exponential function with the same kinetic phases that have been observed in the absence of inorganic phosphate (Figure 35). The first kinetic phase corresponds to the intrinsic xanthone lifetime with  $\tau_{\text{int}} = 1/k_{\text{int}} = 1/(7.0(\pm 3.0) \cdot 10^4 \text{ s}^{-1}) = 14 \pm (6) \mu\text{s}$  and

contributes with of 2(±2)% to the overall amplitude of the kinetics. The second kinetic phase corresponds to end-to-end loop formation in unfolded Brk<sub>66-89</sub> and accounts for 87(±3)% of the overall amplitude. A fit of equation 35 to the denaturant dependence of  $k_c$  in the range of 2.0 M to 8.0 M urea reveals a  $m_c$ -value of 427(±8) (J/mol)/M and a rate constant  $k_{c,H_2O}$  of  $3.1(±0.1) \cdot 10^6 \text{ s}^{-1}$  in the absence of denaturant. As urea is a weaker denaturant compared to GdmCl, the decreased  $m_c$ -value in urea is expected ( $m_c = 541(±4)$  (J/mol)/M for GdmCl, Figure 30). The third kinetic phase does not significantly depend on the urea concentration with a rate constant ranging from  $8 \cdot 10^6 \text{ s}^{-1}$  to  $2 \cdot 10^7 \text{ s}^{-1}$  with relative amplitudes of 9(±3)%. Below 2.0 M urea an additional kinetic phase is necessary to describe the kinetic data. The rate constant of this additional phase is not well separated from the intrinsic xathone triplet lifetime and from the rate constant for loop formation in unfolded Brk<sub>66-89</sub>. To describe the additional kinetic phase, the triplet absorbance traces monitored in the range of 0.5 M to 2.0 M urea were fitted with a quadruple exponential function in which  $k_c$  for loop formation in the unfolded conformation and  $k_{int}$  were fixed on the values expected from their urea dependence above 2.0 M urea (Figure 35). Results reveal a rate constant of  $9.0(±2.0) \cdot 10^5 \text{ s}^{-1}$  for the fourth kinetic phase and relative amplitudes, which increase with decreasing urea concentration. Since the additional kinetic phase appears at low urea concentrations and its contribution to the observed triplet absorbance decay increases with decreasing denaturant concentrations it is likely that this phase is related to secondary structure formation. If this model is correct, contribution of the fourth kinetic phase to the observed triplet absorbance decay should be highest in the absence of urea. However, measurements in the absence of denaturant could not be performed due to aggregation of Brk<sub>66-89</sub> during the degassing procedure in the presence of high concentrations of inorganic phosphate. Measurements in this sample showed that the aggregates do not contribute to observable TTET.



**Figure 35:** Effect of urea on the dynamics of loop formation in  $Brk_{66-89}$  in the presence of 300 mM KP at  $T = 5^\circ\text{C}$ . In the range of 2.0 M to 8.0 M urea results from a triple exponential fit are shown. The solid line represents a fit of equation 35 to  $k_c$  above 2.0 M urea. Below 2.0 M urea results from a quadruple exponential fit are shown in which  $k_c$  in the unfolded conformation (open red circles) and the intrinsic xanthone lifetime (open black circles) were hold fix.

In the absence of inorganic phosphate, where  $Brk_{66-89}$  is unfolded (Figures 29),  $k_c$  is linearly dependent on the urea concentration over the whole range of denaturant concentration (Figure 36).



**Figure 36:** Effect of urea on the dynamics of loop formation in  $Brk_{66-89}$  in 20 mM NaAc 20 mM NaCl pH 5.7 at  $T = 5^\circ\text{C}$ . The solid line represents a fit of equation 35 to the rate constant of the main kinetic phase.

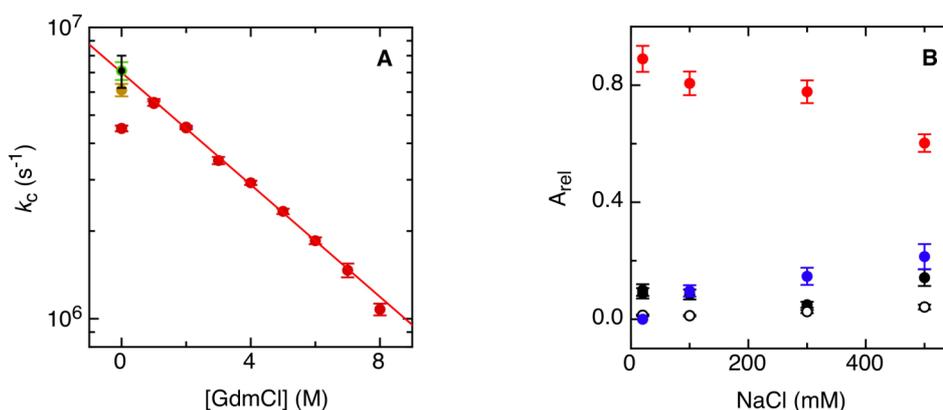
A fit of equation 35 to the data monitored at 5°C reveals an  $m_c$ -value of 420(±3) (J/mol)/M and  $k_{c,H_2O}$  of  $2.1(\pm 0.1) \cdot 10^6$  s<sup>-1</sup> in the absence of denaturant. This  $m$ -value

supports the result from measurements in the presence of inorganic phosphate at high urea concentrations ( $m_c = 427(\pm 8)$  (J/mol)/M). The observation that the addition of inorganic phosphate increases  $k_{c,H_2O}$  ( $k_{c,H_2O} = 3.1(\pm 0.1) \cdot 10^6$  s<sup>-1</sup> in 300 mM Pi at 5°C) and thus loop formation in unfolded Brk<sub>66-89</sub> at a given urea concentration supports the model of decelerated chain dynamics due to repulsion of positively charged residues at low ionic strength.

To study the contribution of repulsive interactions between positively charged side chains to the slow kinetics of loop formation in Brk<sub>66-89</sub> in more detail we performed TTET experiments in the presence of high NaCl concentrations, where electrostatic interactions are screened. A quadruple exponential function is necessary to fit transient triplet absorbance decays measured in the presence of high NaCl concentrations (Figure 37B). The slowest kinetic phase corresponds to the intrinsic xanthone triplet lifetime with  $\tau_{int} = 1/k_{int} = 1/(1.5(\pm 0.3) \cdot 10^5$  s<sup>-1</sup>) = 6.7(±1.3) μs and accounts for 2(±2)% of the total amplitude of the triplet absorbance decay. The rate constant of the second kinetic phase is independent of the NaCl concentration with  $k = 1.1(\pm 0.2) \cdot 10^6$  s<sup>-1</sup>. The relative amplitude of this phase increases with increasing salt concentration (Figure 37B), indicating that this phase might be related to end-to-end loop formation in the helical conformation. At the same time contributions from the third kinetic phase decrease with increasing salt concentration, indicating that this phase can be attributed to loop formation in unfolded Brk<sub>66-89</sub> (Figure 37B). The rate constant of the third kinetic phase is increased with increasing NaCl concentration reaching a limiting value of  $7.0(\pm 0.1) \cdot 10^6$  s<sup>-1</sup> (Figure 37A). This value is in good agreement with the rate constant for loop formation in unfolded Brk<sub>66-89</sub> extrapolated from measurements in the presence of GdmCl to zero denaturant (Figure 37A). This result supports the model of slow chain dynamics in the unfolded conformation due to repulsion of positively charged residues at low ionic strength. The fourth kinetic phase with  $k = 3.0(\pm 1.0) \cdot 10^7$  s<sup>-1</sup> and a contribution of 9(±5)% to the total amplitude of observed TTET corresponds to the fast phase of unknown origin.

Since Brk<sub>66-89</sub> tends to precipitate during the degassing procedure at high ionic strength conditions another explanation for the appearance of the second kinetic phase might be related to xanthone triplet quenching in the aggregates. This explanation is

contradicted by measurements in the presence of 300 mM Pi in the absence of denaturant, which showed that the aggregates do not contribute to observable TTET. As the lifetime of the second kinetic phase is rather well separated from the intrinsic xanthon lifetime  $\tau_{\text{int}}$  ( $1.5(\pm 0.2) \cdot 10^5 \text{ s}^{-1}$  in 300 mM NaCl at 22.5°C) it is not probable that it corresponds to intrinsic xanthone quenching.



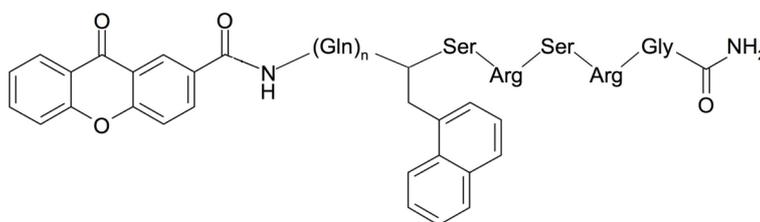
**Figure 37:** A: Rate constant for loop formation in  $Brk_{66-89}$  in 20 mM NaAc, 20 mM NaCl pH 5.7 as a function of GdmCl (red, see figure 40). In the absence of denaturant the rate constant is enhanced by the addition of NaCl: 100 mM (yellow), 300 mM (green) and 500 mM (black).  $T = 22.5^\circ\text{C}$  pH 5.7. B: Relative amplitudes of the kinetic phases obtained from a quadruple exponential fit to the observed triplet absorbance decays measured in the presence of different NaCl concentrations:  $A_{1,rel}$  (intrinsic xanthon triplet lifetime, open circles),  $A_{2,rel}$  (loop formation in helical conformation, blue),  $A_{3,rel}$  (loop formation in unfolded conformation, red) and  $A_{4,rel}$  (kinetic phase of unknown origin, black).

To study in more detail whether increased chain dimension and increased internal friction in  $Brk_{66-89}$  contribute to the slow chain dynamics in  $Brk_{66-89}$  we plan to determine end-to-end distances and intrachain diffusion constants by applying FRET.

### 3.1.3 Dynamics of loop formation in aggregation prone polyglutamine peptides

Several neurodegenerative diseases associated with proteins are based on the presence of abnormally long polyglutamine stretches. Among these diseases, the most common and best known is Huntington's disease.<sup>70</sup> The unusually long polyglutamine stretch facilitates protein aggregation, which is assumed to be related to the disease process.<sup>71-79</sup> In a recent study, Singh et al. proposed extreme chain stiffness and very extended conformations for soluble polyglutamine peptides from results of Trp/Cys contact quenching experiments.<sup>87</sup> However, this method is not diffusion-controlled, which makes a quantitative evaluation impossible. It is thus still unclear whether unfolded polyglutamine chains have different structural and dynamical properties compared to other polypeptide chains.

To test for unusual chain properties in polyglutamine peptides we applied triplet-triplet energy transfer (TTET) from xanthone (Xan) to naphthylalanine (Nal) to measure end-to-end loop formation kinetics. We investigated peptides of the type shown in figure 38 with  $n = 2, 4$  and  $9$  ((Gln)<sub>2</sub>, (Gln)<sub>4</sub> and (Gln)<sub>9</sub>) and compared the dynamics of loop formation to results from poly(glycine-serine) and polyserine model peptides.

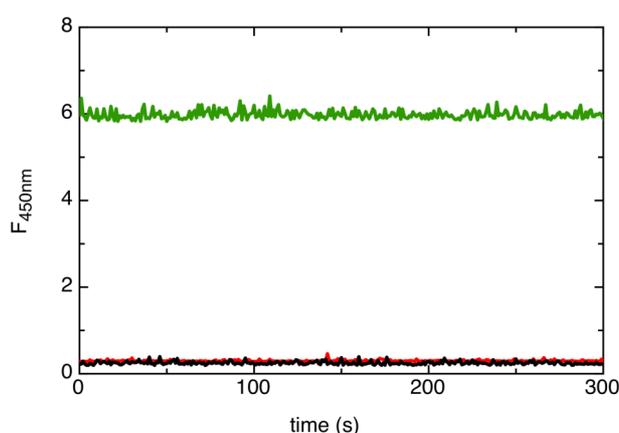


**Figure 38:** Polyglutamine peptides used in TTET experiments. Xanthone was attached to the N-terminus and naphthyl was introduced via the non-natural amino acid naphthylalanine. The short C-terminal sequence including two arginines was introduced to increase solubility. The C-terminus of the peptides was amidated.

Solutions of polyglutamine peptides are often heterogeneous due to peptide aggregation and therefore different protocols were developed to obtain homogenous samples. We used a protocol introduced by O'Nullain et al. (see materials and methods)<sup>190</sup> to remove all aggregates and aggregation nuclei in polyglutamine samples. Although spontaneous aggregation in short polyglutamine peptides is very

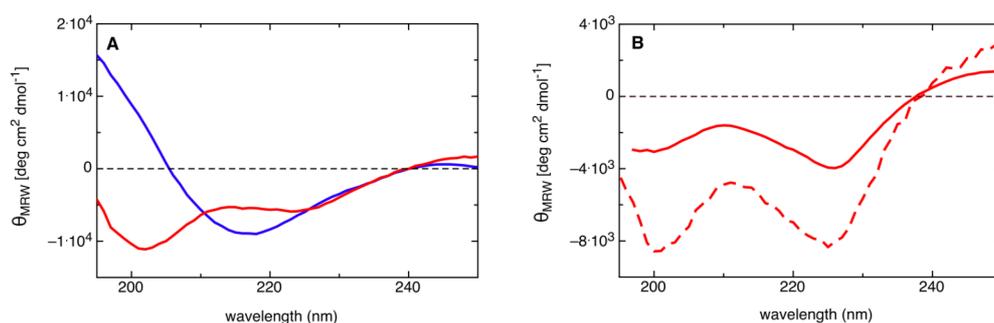
slow we tested whether applying this protocol to samples of (Gln)<sub>9</sub> prior to TTET measurements has an effect on the observed kinetics of end-to-end contact formation.

To test for successful disaggregation, the final solution was analysed by monitoring light scattering in a fluorescence spectrometer in which emission and excitation wavelength were set to 450nm.<sup>190</sup> The obtained Rayleigh light-scattering signal was comparable to scattering from buffer alone and significantly lower than the signal of a partly aggregated sample indicating complete removal of residual aggregates (Figure 39).



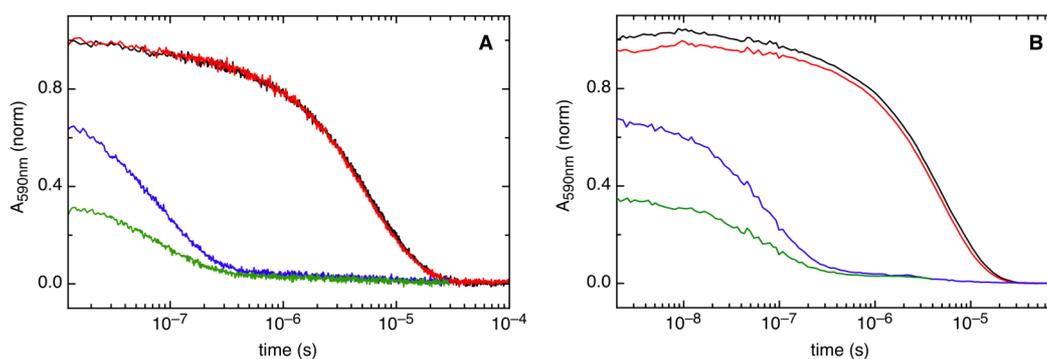
**Figure 39:** Test for residual aggregates in (Gln)<sub>9</sub>. The signal of the disaggregated solution (red) is comparable to scattering from buffer alone (black). The signal of a partly aggregated sample (green) is significantly higher.  $T = 22.5$  °C.

Studying secondary structure formation by circular dichroism spectroscopy revealed unstructured conformations only for disaggregated (Gln)<sub>9</sub> samples as indicated by a minimum of the ellipticity around 200 nm (Figure 40). Untreated (Gln)<sub>9</sub> samples show significant amounts of  $\beta$ -sheet conformations as indicated by a minimum of the ellipticity at 218 nm and a positive ellipticity below 200 nm (Figure 40). The typical  $\beta$ -sheet CD spectrum indicates the presence of residual aggregates in the untreated (Gln)<sub>9</sub> sample.<sup>85</sup> For both shorter peptides, (Gln)<sub>4</sub> and (Gln)<sub>2</sub>, we found unstructured conformations even without application of the disaggregation protocol.



**Figure 40:** A: CD spectra of  $(\text{Gln})_9$  after disaggregation (red) and without prior treatment (blue) at 22.5 °C. Using polycarbonate tubes instead of polyallomer tubes for the centrifugation step induced  $\beta$ -sheet conformations even in disaggregated samples. B: CD spectra of  $(\text{Gln})_4$  (solid line) and  $(\text{Gln})_2$  (dashed line) at 22.5 °C without prior treatment. All spectra were measured in 10 mM NaCac pH 7.0.

Figure 41A shows the kinetics of end-to-end loop formation in water for a disaggregated  $(\text{Gln})_9$  sample and for an untreated  $(\text{Gln})_9$  sample measured by TTET from xanthone to naphthylalanine induced by a 4-ns laser flash. The amplitudes of the observed triplet absorbance decays were compared to the respective amplitude measured in a  $(\text{Gln})_3$  donor-only peptide. In this reference peptide the naphthylalanine residue was replaced by a phenylalanine, which cannot undergo TTET with Xan.<sup>53</sup>



**Figure 41:** Kinetics of end-to-end contact formation for disaggregated  $(\text{Gln})_9$  (blue) and untreated  $(\text{Gln})_9$  (green) measured in nanosecond-laser flash TTET experiments (A) and picosecond-laser flash TTET experiments (B). A  $(\text{Gln})_3$  donor-only peptide was used as reference: Disaggregated (black) and without treatment (red). Measurements were carried out in 10mM NaCac pH 7.0 at  $T = 22.5^\circ\text{C}$ .

Both the disaggregated sample and the untreated sample exhibit smaller amplitudes for the observed triplet absorbance decay compared to the donor-only reference peptide. The amplitude of the observed triplet absorbance decay reaches 60% for the disaggregated sample and 30% for the untreated sample, compared to the amplitude measured in the reference peptide. The observable rate constants,  $k_c$ , for end-to-end contact formation in the disaggregated and the untreated (Gln)<sub>9</sub> sample are similar (Table 8). Together with the increased loss in signal amplitude for the untreated sample compared to the disaggregated sample this indicates that species adopting  $\beta$ -sheet conformations do not contribute to the observed TTET. Possible explanations for the loss in signal amplitude observed in the disaggregated (Gln)<sub>9</sub> sample compared to the donor-only reference are subnanosecond chain dynamics. To test for this we performed picosecond-laser flash experiments in which TTET is induced by a 120 ps laser flash and by this allow measurement of loop formation kinetics from 2 ns on. These experiments cannot resolve the missing amplitude, indicating the absence of chain dynamics on the 2 to 10 ns time scale (Figure 41B). Thus the observed loss in signal amplitude must be due to even faster processes, possibly due to  $\beta$ -hairpin formation or fast intermolecular TTET between subunits of residual aggregates.

A triple exponential function is necessary to describe the observed triplet absorbance decay in the (Gln)<sub>2</sub> peptide (Table 8). The slowest kinetic phase corresponds to the intrinsic lifetime of the xanthone triplet and accounts for 4( $\pm$ 2)% of the overall amplitude of the observed kinetics. The second kinetic phase has a rate constant of  $9.1(\pm 1.0) \cdot 10^6 \text{ s}^{-1}$  in water and accounts for 10( $\pm$ 3)% of the overall amplitude of the observed triplet absorbance decay. The origin of this phase is still unclear, however, as it is still present in the presence of high denaturant concentrations (Figures 42A and 42B) it seems to be unrelated to structure formation or aggregation. The third kinetic phase corresponds to end-to-end loop formation and accounts for 86( $\pm$ 4)% of the overall amplitude of the observed triplet absorbance decay. For (Gln)<sub>4</sub> only two exponentials are required to describe the observed triplet absorbance decay (Table 8). The slowest kinetic phase corresponds to the intrinsic lifetime of the xanthone triplet and contributes 5( $\pm$ 2)% to the overall amplitude. The second kinetic phase accounts for 95( $\pm$ 4)% of the overall amplitude and corresponds to end-to-end loop formation. For (Gln)<sub>9</sub> a triple exponential function is necessary to describe the observed triplet

absorbance decay (Table 8). The slowest kinetic phase corresponds again to the intrinsic lifetime of the xanthone triplet and accounts for 4( $\pm$ 3)% of the overall amplitude. The second kinetic phase contributes with 84( $\pm$ 3)% to the overall amplitude of the observed triplet absorbance decay and corresponds to end-to-end loop formation. The third kinetic phase has a rate constant of  $4.5(\pm 1.2) \cdot 10^7 \text{ s}^{-1}$  in water and accounts for 13( $\pm$ 4)% of the overall amplitude of the observed triplet absorbance decay. The origin of this phase is still unclear. It does neither depend on oxygen content nor on peptide concentration, arguing against a bimolecular reaction. Additionally, this phase is still present in the presence of high denaturant concentrations (Figures 42C and 42D) suggesting that it does not appear due to structure formation or aggregation. In water at 22.5 °C, TTET experiments revealed rate constants  $k_c$  for loop closure of  $4.9(\pm 0.3) \cdot 10^7 \text{ s}^{-1}$  in (Gln)<sub>2</sub>,  $2.1(\pm 0.1) \cdot 10^7 \text{ s}^{-1}$  in (Gln)<sub>4</sub> and  $1.0(\pm 0.1) \cdot 10^7 \text{ s}^{-1}$  in (Gln)<sub>9</sub> (Table 8). Degassing the samples before TTET measurements to prevent quenching by oxygen does not change these values significantly.

sample	$A_{1,\text{rel}}$	$k_1 \text{ (s}^{-1}\text{)}$	$A_{2,\text{rel}}$	$k_2 \text{ (s}^{-1}\text{)}$	$A_{3,\text{rel}}$	$k_3 \text{ (s}^{-1}\text{)}$
(Gln)2 untreated	0.04 ( $\pm 0.02$ )	$5(\pm 1) \cdot 10^5$	0.10 ( $\pm 0.03$ )	$9.1(\pm 1.0) \cdot 10^6$	0.86 ( $\pm 0.04$ )	$4.9(\pm 0.3) \cdot 10^7$
(Gln)4 untreated	0.05 ( $\pm 0.02$ )	$5(\pm 1) \cdot 10^5$	0.95 ( $\pm 0.04$ )	$2.1(\pm 0.1) \cdot 10^7$	-	-
(Gln)9 untreated	0.04 ( $\pm 0.02$ )	$4(\pm 2) \cdot 10^5$	0.84 ( $\pm 0.03$ )	$1.1(\pm 0.1) \cdot 10^7$	0.12 ( $\pm 0.04$ )	$4.5(\pm 1.2) \cdot 10^7$
(Gln)9 disaggregated	0.03 ( $\pm 0.02$ )	$6(\pm 2) \cdot 10^5$	0.84 ( $\pm 0.03$ )	$1.1(\pm 0.1) \cdot 10^7$	0.13 ( $\pm 0.04$ )	$4.4(\pm 1.1) \cdot 10^7$

**Table 8:** Kinetic phases for the studied polyglutamine peptides obtained by fitting the observed triplet absorbance decays with the sum of two or three exponentials, respectively.

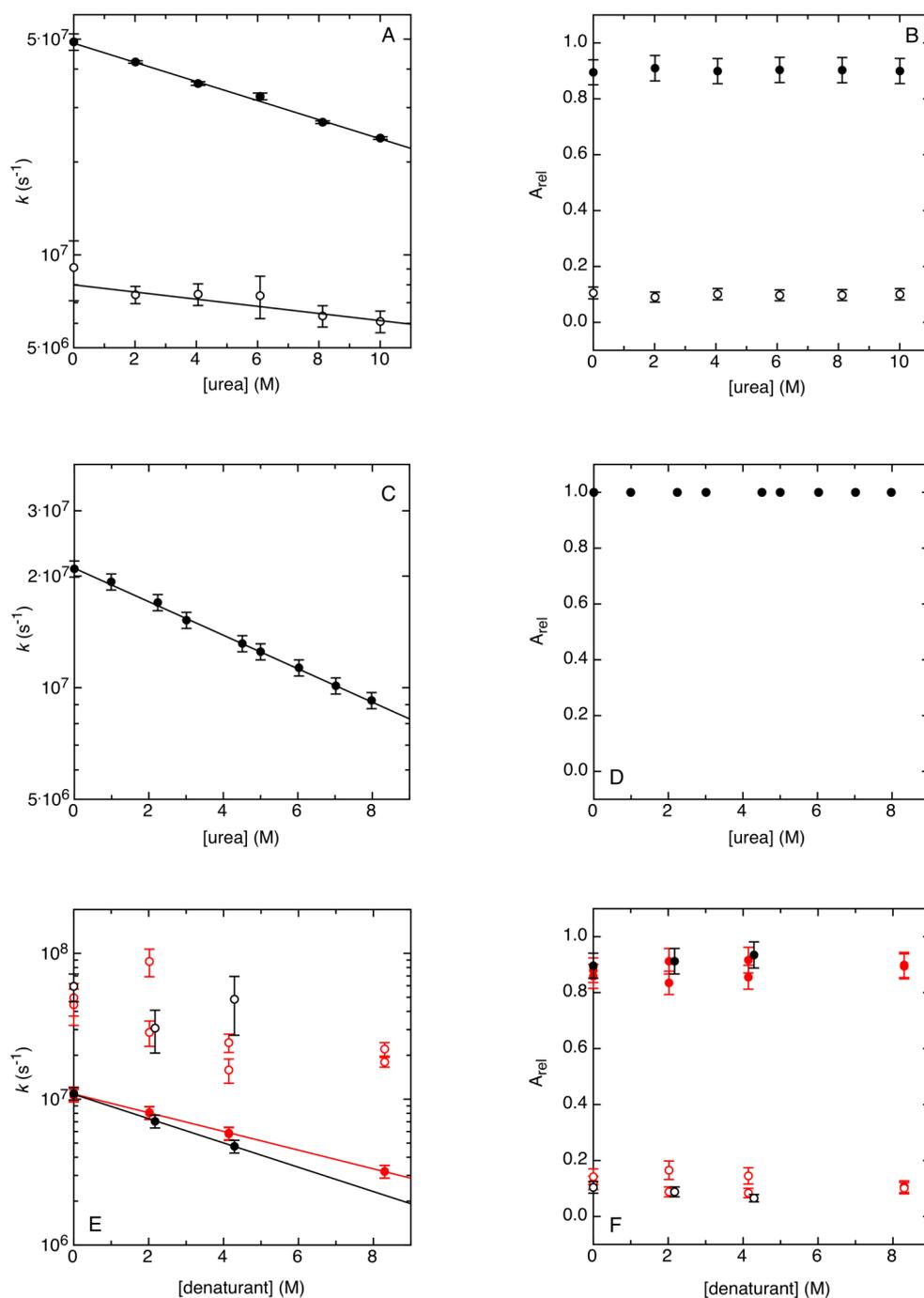
For the three studied polyglutamine peptides a linear relationship between  $\ln k_c$  and the denaturant concentration [D] is observed (Figure 42). This data can be described by

$$\ln k_c = \ln k_c^0 - \frac{m_c[D]}{RT} \quad (35)$$

Here  $k_c^0$  is the rate constant for end-to-end loop formation at zero denaturant concentration,  $k_c$  is the rate constant at the denaturant concentration  $[D]$  and  $m_c$  describes the sensitivity of  $k_c$  to the presence of denaturant. The results for (Gln)<sub>2</sub>, (Gln)<sub>4</sub> and (Gln)<sub>9</sub> are summarized in table 9. Comparing chains of similar length shows that the  $m_c$ -values for the polyglutamine peptides are similar to previous results on flexible poly(glycine-serine) chains.<sup>189</sup> However, the increase in  $m_c$  with increasing chain length is slightly more pronounced in the polyglutamine chains. The ratio of  $m_c$ -values for GdmCl and urea (1.3 for (Gln)<sub>9</sub>) is similar to the results on poly(glycine-serine) chains.<sup>189</sup> Linear extrapolation of the rate constants to zero denaturant confirms the value for  $k_c$  measured in the absence of denaturant for all three polyglutamine peptides (Figure 42 and table 9).

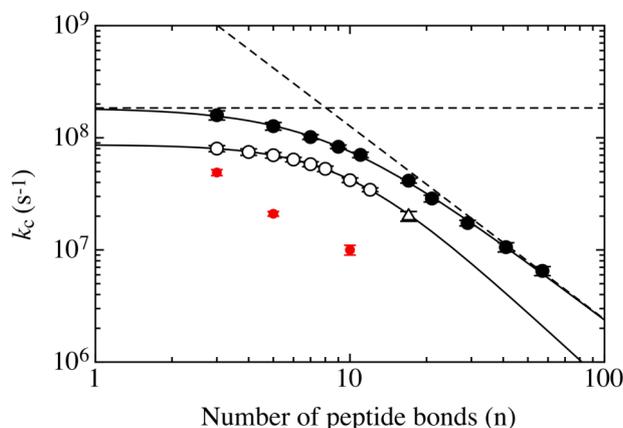
peptide	$k_{c,H_2O}$ (s <sup>-1</sup> )	$k_c^0$ (s <sup>-1</sup> )	$m_{c,urea}$ ((J/mol)/M)	$m_{c,GdmCl}$ ((J/mol)/M)
(Gln) <sub>2</sub>	4.9(±0.3)·10 <sup>7</sup>	4.9(±0.2)·10 <sup>7</sup>	175(±4)	-
(Gln) <sub>4</sub>	2.1(±0.1)·10 <sup>7</sup>	2.1(±0.1)·10 <sup>7</sup>	256(±5)	-
(Gln) <sub>9</sub>	1.0(±0.1)·10 <sup>7</sup>	1.1(±0.1)·10 <sup>7</sup>	363(±10)	472(±27)

**Table 9:** Denaturant dependence of end-to-end loop formation in polyglutamine peptides.  $m_c$  and  $k_c^0$  were determined from a fit of equation 35 to the rate constants of end-to-end loop formation in the presence of different denaturant concentrations.  $k_{c,H_2O}$  is the measured rate constant in the absence of denaturant.



**Figure 42:** A: Effect of denaturant on loop formation in  $(Gln)_2$  (A and B),  $(Gln)_4$  (C and D) and  $(Gln)_9$  (E and F). All measurements were carried out in 10mM NaCac buffer pH 7.0 at  $T = 22.5^\circ C$ . Measurements on  $(Gln)_9$  in GdmCl were carried out on disaggregated samples while samples in urea were untreated.  $(Gln)_2$  and  $(Gln)_4$  samples were not disaggregated before measurement. The relative amplitudes were determined without contributions from the intrinsic xanthone lifetime.

The kinetics of end-to-end loop formation in the studied polyglutamine chains are very slow compared to kinetics in poly(glycine-serine) and polyserine chains, (Figure 43). This effect increases with increasing chain length.



**Figure 43:** Comparison of rate constants  $k_c$  for end-to-end contact formation in polyglutamine peptides (red) with results from poly(Gly-Ser) chains (filled black circles), polySer chains (open black circles) and a natural loop sequence of the parvalbumin protein (open triangle) in water. Adapted from reference 47.

To study the unusually slow chain dynamics observed in the studied polyglutamine sequences in more detail we investigated the effect of solvent viscosity and of temperature on end-to-end loop formation in (Gln)<sub>2</sub>, (Gln)<sub>4</sub> and (Gln)<sub>9</sub>. This allows determination of contributions from solvent motion and the presence of free energy barriers. To quantify the effect of solvent viscosity  $\eta$  on the kinetics, the empirical equation

$$k_c = k_{c,0} \left( \frac{\eta}{\eta_0} \right)^{-\beta} \quad (29)$$

is fitted to rate constants determined at different solvent viscosities. Here  $\eta_0$  is the reference solvent viscosity of water at  $T = 22.5^\circ\text{C}$  and  $k_{c,0}$  is the rate constant at  $\eta_0$ . The exponent  $\beta$  reflects the sensitivity of the reaction on solvent viscosity. For reactions, which are fully diffusion-controlled  $\beta$  is 1 and rate constants decrease with  $1/\eta$ . For reactions, which are independent of solvent motion  $\beta$  is 0 and the reaction becomes fully determined by activation barriers. For all studied polyglutamine

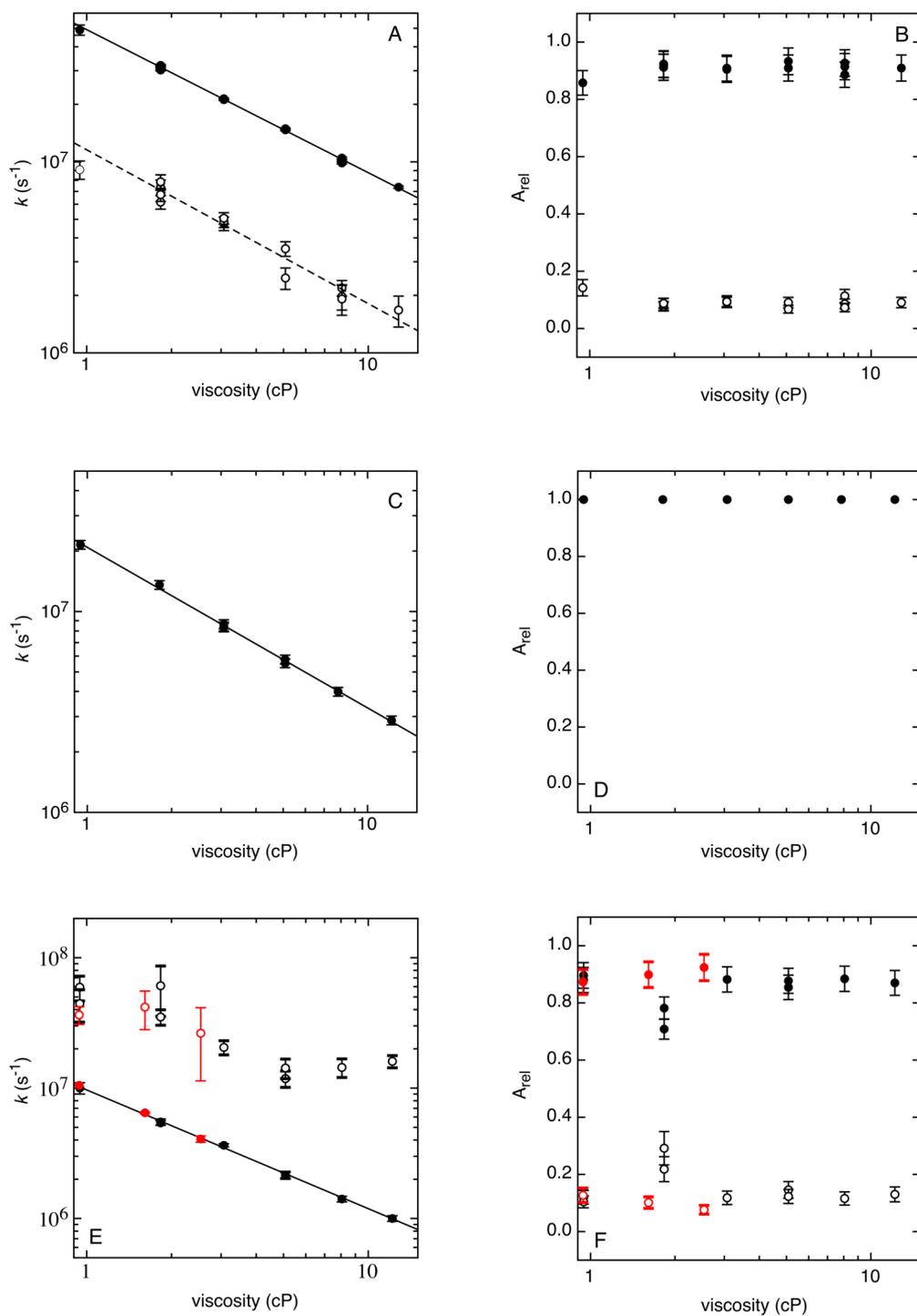
peptides we found a fractional viscosity dependence of  $k_c$  indicating that end-to-end loop formation in these chains is not fully determined by diffusion (Figure 44 and table 10). The determined  $\beta$ -values were used to correct measurements carried out at different temperatures for the effect of temperature on solvent viscosity. The corrected rate constants  $k'_c$  can be described by the Arrhenius equation

$$k'_c = A \cdot e^{\frac{-E_a}{RT}} \quad (32)$$

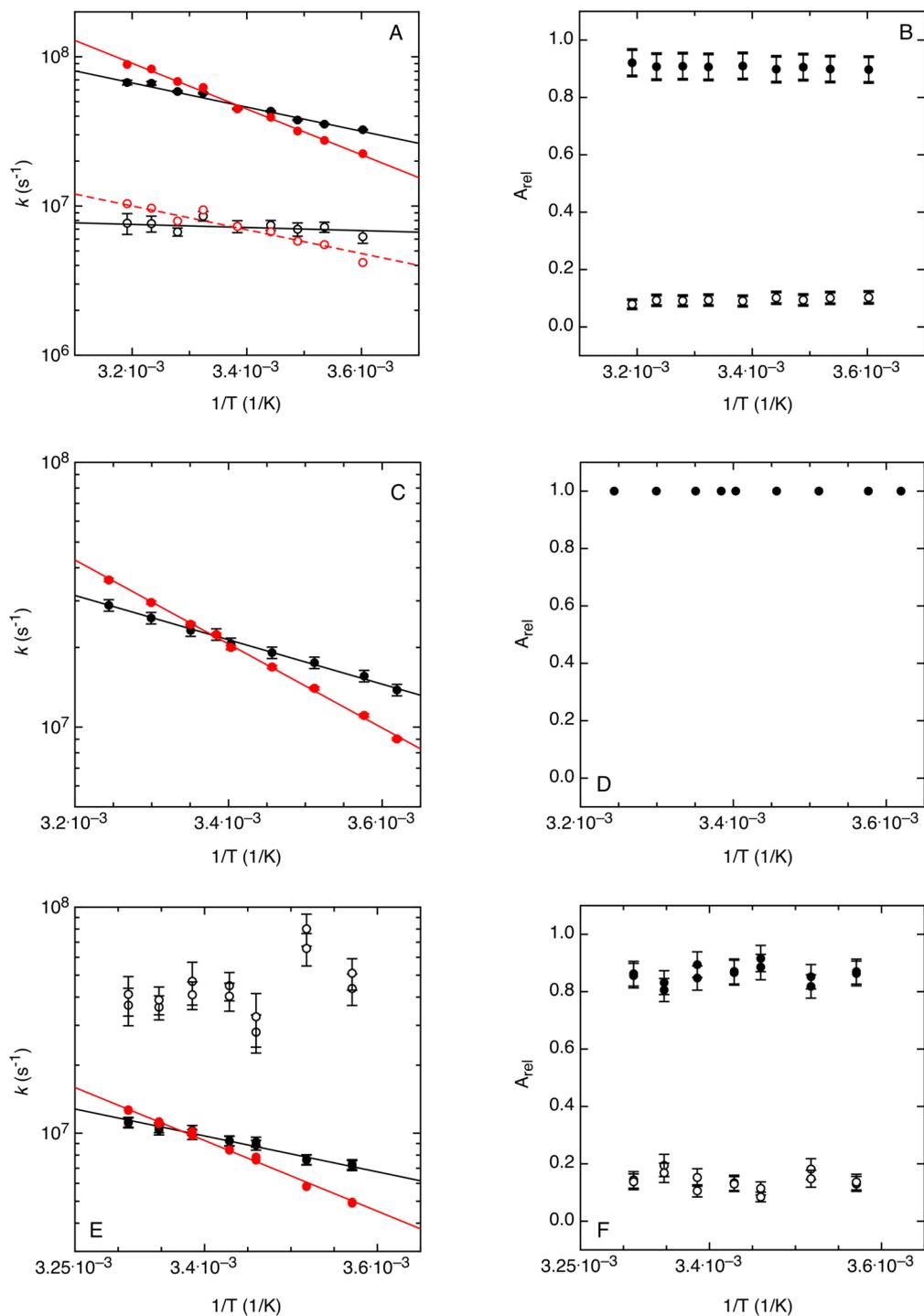
Results reveal significant activation barriers for all three studied polyglutamine peptides (Figure 45 and table 10).

peptide	$E_a$ (kJ/mol)	$A$ (s <sup>-1</sup> )	$\beta$
(Gln) <sub>2</sub>	15.5(±1.0)	2.6(±0.7) · 10 <sup>10</sup>	0.75(±0.02)
(Gln) <sub>4</sub>	16.0(±0.3)	1.5(±0.4) · 10 <sup>10</sup>	0.80(±0.03)
(Gln) <sub>9</sub>	15.1(±1.0)	4.9(±1.4) · 10 <sup>9</sup>	0.91(±0.03)

**Table 10:** Effect of temperature and of solvent viscosity on the dynamics of end-to-end loop formation in polyglutamine peptides. The activation energy  $E_a$  is corrected for the effect of temperature on solvent viscosity according to equation 31.

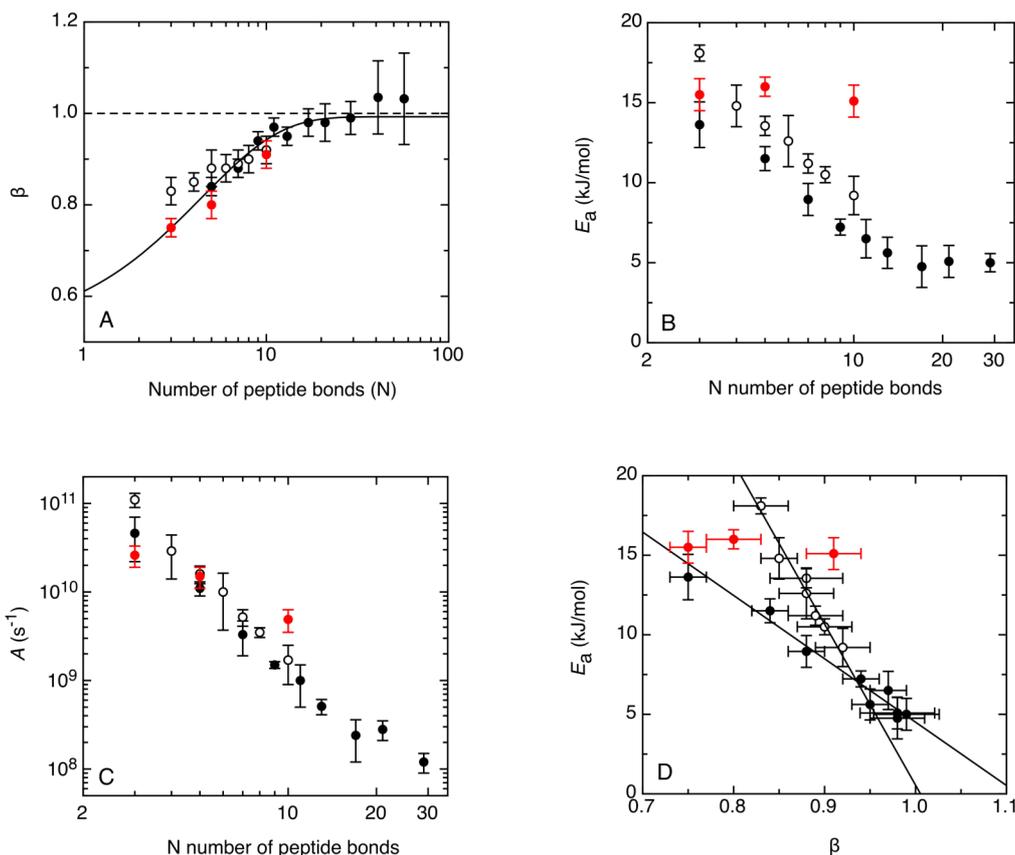


**Figure 44:** Effect of solvent viscosity on end-to-end loop formation in (Gln)<sub>2</sub> (A and B), (Gln)<sub>4</sub> (C and D) and (Gln)<sub>9</sub> (E and F). Measurements were carried out in 10 mM NaCac pH 7.0 at  $T=22.5^{\circ}\text{C}$  and viscosity was adjusted by adding glycerol. Measurements on (Gln)<sub>9</sub> were carried out either on disaggregated samples (red) or untreated samples (black) while measurements on (Gln)<sub>2</sub> and (Gln)<sub>4</sub> were carried out on untreated samples. The solid lines represent fits of equation 29 to the data. Fit results are summarized in table 10. Relative amplitudes  $A_{rel}$  were determined without contributions from the intrinsic xanthone lifetime.



**Figure 45:** Effect of temperature on end-to-end loop formation in  $(Gln)_2$  (A and B),  $(Gln)_4$  (C and D) and  $(Gln)_9$  (E and F) without correction (black) and after correction for the effect of temperature on solvent viscosity (red). Measurements were carried out in 10 mM NaCac pH 7.0 using non-disaggregated samples. The solid lines represent fits of equation 32 to the data. Fit results are summarized in table 10. Relative amplitudes  $A_{rel}$  were determined without contributions from the intrinsic xanthone lifetime.

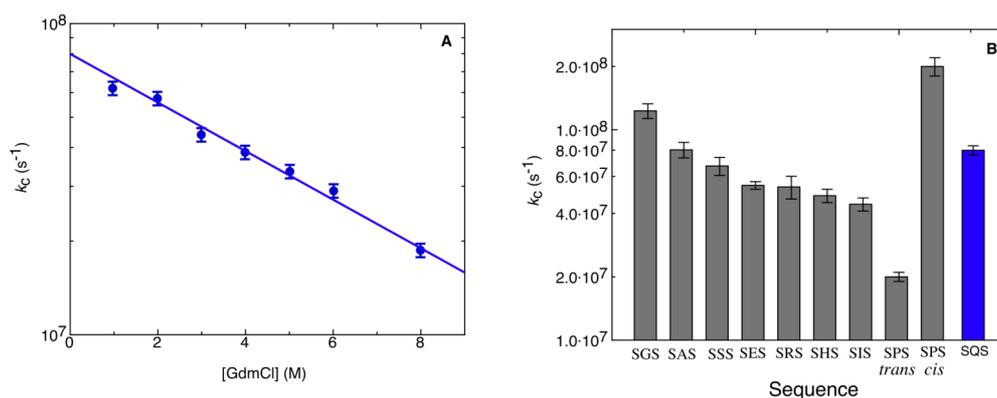
Chapter 3.1.1.1 and 3.1.1.2 describe a detailed study on the effect of solvent viscosity and of temperature on chain dynamics in poly(glycine-serine) and polyserine chains, which were used as a model for the unfolded state. A direct comparison of these measurements to the results on polyglutamine peptides studied here is shown in figure 46. Measurements on the effect of solvent viscosity on the rate constant for end-to-end loop formation reveal a similar behaviour for polyglutamine peptides compared to poly(glycine-serine) and polyserine chains (Figure 46A). In all studied polyglutamine peptides the dependence of  $k_c$  on solvent viscosity is fractional ( $k_c \sim 1/\eta^\beta$  with  $\beta < 1$ ) and  $\beta$  increases with increasing chain length. This indicates contributions from chain stiffness to the dynamics of end-to-end loop formation in these chains. Whether loop formation in long polyglutamine chains is fully diffusion-controlled ( $\beta = 1$ ) is still unclear. Measurements on the effect of temperature on  $k_c$  revealed a different behaviour for polyglutamine peptides compared to poly(glycine-serine) and polyserine chains (Figure 46B and 46C). For poly(glycine-serine) chains and polyserine chains we found that the activation energy is high for formation of short loops and decreases with increasing loop length. However, for the studied polyglutamine peptides the activation energy is rather independent of chain length with a value comparable to the results on the shortest poly(glycine-serine) and polyserine chains. (Gln)<sub>2</sub> and (Gln)<sub>4</sub> show similar Arrhenius pre-exponential factors compared to poly(glycine-serine) and polyserine chains of similar length. However, (Gln)<sub>9</sub> shows a slightly increased Arrhenius pre-exponential factor compared to poly(glycine-serine) and polyserine chains of similar length, which indicates decreased entropic cost for loop formation in (Gln)<sub>9</sub>, suggesting increased chain stiffness. For poly(glycine-serine) and polyserine chains we found a negative correlation between the activation energy for end-to-end loop formation and the sensitivity of the reaction towards solvent viscosity using small co-solutes like glycerol. Results for the studied polyglutamine peptides do not fit to these correlations (Figure 46D). In summary, these results indicate that the relatively high and chain length independent activation energy is the main source of decelerated chain dynamics in polyglutamine peptides compared to poly(glycine-serine) and polyserine chains.



**Figure 46:** Effect of solvent viscosity and of temperature on end-to-end loop formation in polyglutamine chains (red), poly(glycine-serine) chains (black) and polyserine chains (open). A: Sensitivity of loop formation dynamics to solvent viscosity as a function of chain length. Viscosity was adjusted by adding glycerol. B: Activation energy as a function of chain length. C: Arrhenius pre-exponential factor as a function of chain length. D: Correlation between the activation energy  $E_a$  and the sensitivity of loop formation dynamics to solvent viscosity.

To test whether slow chain dynamics in polyglutamine sequences are due to properties of a single glutamine residue we performed host-guest experiments. In an earlier study the effect of different amino acids on local chain dynamics was studied measuring TTET in short host guest peptides of the sequence Xan-Ser-Xaa-Ser-NAla-Ser-Gly.<sup>47</sup> Xaa denotes the guest amino acid for which Gly, Ser, Ala, Ile, His, Glu, Arg and Pro were tested. The results showed that the amino acid side chain has only a small effect on the contact formation rate constant (Figure 8). Glycine exhibits the fastest dynamics consistent with its high flexibility. Amino acids with short side chains show slightly faster loop formation than long side chains. The Ser-Pro-Ser

sequence shows two rate constants corresponding to the *cis* and *trans* state of the Ser-Pro peptide bond, respectively. We synthesized the same host guest peptide with glutamine as the guest amino acid. The obtained rate constant for loop formation is comparable to the value from results on the Ser-Ser-Ser sequence (Figure 47). This is slightly faster than expected considering the general observation of slightly decreased rate constants for large side-chains. The results show that glutamine itself has no retarding effect on local chain dynamics.



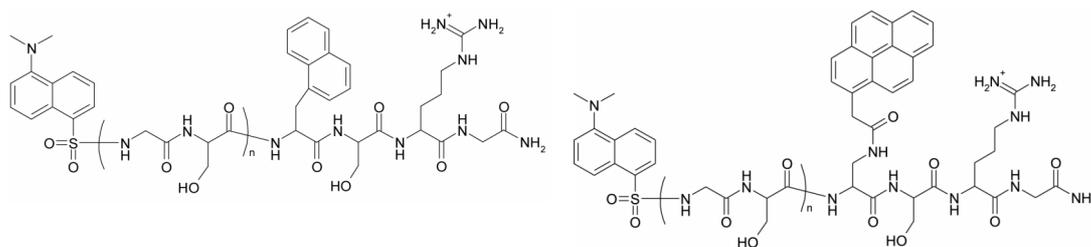
**Figure 47:** Rate constants  $k_c$  for end-to-end contact formation measured by TTET in host guest peptides to study the effect of individual amino acids on local chain dynamics, Adapted from reference 47. The result obtained for Xaa = Gln is shown in blue and was measured in 10 mM NaCac pH 7.0 at T=22.5 °C.

To get more insight into internal motion in polyglutamine sequences, we plan to measure a complete dependence of end-to-end contact formation kinetics on chain length as is was done previously for poly(glycine-serine) and polyserine chains.<sup>47</sup> Further, we want to measure the effect of denaturant, viscosity and temperature on chain dynamics in all these polyglutamine peptides.

### 3.1.4 End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains studied by FRET

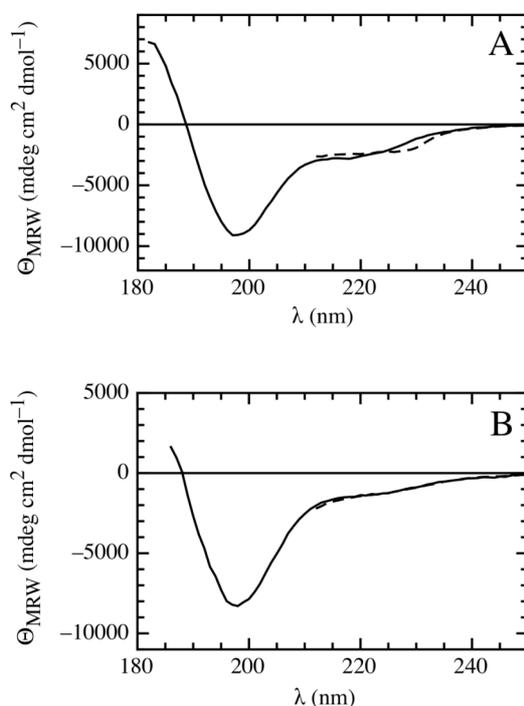
Characterization of the unfolded state is important to understand the process of protein folding in detail. Compared to the native state, the unfolded state cannot be characterized by high-resolution structure techniques due to the large ensemble of conformations in this state. Usually, very harsh conditions are used to unfold proteins *in vitro* what makes characterization under physiological conditions difficult. Many theoretical and experimental studies indicated specific and non-specific interactions as well as significant amounts of residual structure in the unfolded state.<sup>41,191-200</sup>

In previous experiments, our group applied triplet-triple energy transfer (TTET) from xanthone to naphthalene on poly(Gly-Ser) chains of different length to study the kinetics of intramolecular loop formation in unfolded polypeptide chains.<sup>47,189</sup> Results indicated that properties of these chains significantly change in the presence of denaturants.<sup>189</sup> The rate constant  $k_c$  for end-to-end contact formation decreases in the presence of denaturant due to increased solvent viscosity and due to binding of the denaturant molecules to the polypeptide chain.<sup>189</sup> To obtain more information on structure and internal dynamics of unfolded polypeptide chains we performed time resolved fluorescence resonance energy transfer (FRET) experiments on an (Gly-Ser)<sub>16</sub> chain in the absence and presence of the denaturant guanidinium hydrochloride (GdmCl).<sup>201</sup> FRET in unstructured polypeptide chains is determined both by the conformational heterogeneity and by internal dynamics on the time scale of the donor lifetime. This provides the chance to obtain information of both the dimensions and the dynamics of the polypeptide chain. In our strategy we measured FRET with two different donor-acceptor pairs with similar  $R_0$  but with different donor life times (Figure 48).



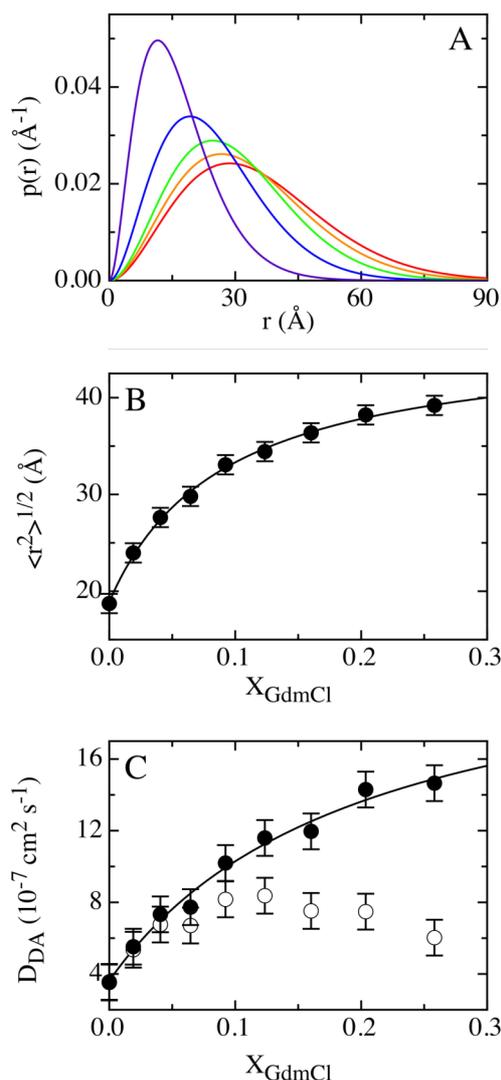
**Figure 48:** Structure of (glycine-serine)<sub>16</sub> labelled with dansyl as FRET acceptor and either naphthalene (left) or pyrene (right) as FRET donor. Taken from reference 201.

(Gly-Ser)<sub>16</sub> labelled for FRET is unstructured, indicated by a minimum of the CD signal around 200 nm (Figure 49)



**Figure 49:** CD spectra of the (Gly-Ser)<sub>16</sub> chains labelled with naphthalene/dansyl (A) and with pyrene/dansyl (B) in the absence of GdmCl (solid lines) and in the presence of 6 M GdmCl (dashed lines). The negative band at 230 nm in the naphthalene/dansyl-labelled peptide is due to contributions from naphthylalanine. Slight changes in this region between 0 and 6 M GdmCl may indicate small changes in the local environment of the naphthalene group. Taken from reference 201, supporting information.

Since the effect of intrachain diffusion on FRET efficiency increases with increasing donor lifetime, global analysis of the data allowed reliable determination of interchromophore distance distributions and intrachain diffusion constants. Results show that an increase in end-to-end distance and in diffusion constant for motion of the two fluorophores relative to each other with increasing GdmCl concentration (Figure 50).



**Figure 50:** A: Donor-acceptor distance probability distribution function  $p(r)$  of (glycine-serine)<sub>16</sub> in the presence of 0 M GdmCl (violet), 2 M GdmCl (blue), 4 M GdmCl (green), 6 M GdmCl (orange) and 8 M GdmCl (red). B: Effect of GdmCl on rms end-to-end distances  $\langle r^2 \rangle^{1/2}$ , calculated from donor-acceptor distance distribution functions. C: Effect of GdmCl on donor-to-acceptor diffusion constants  $D_{DA}$  obtained from the global analysis (open circles) and on viscosity-corrected values (closed circles). The solid lines describe fits of the Schellman weak binding model to the data. Taken from reference 201.

Both effects can be described by the Schellman's weak binding model (Figure 50). Since there are no specific side-chain interactions present in (glycine-serine)<sub>16</sub>, the significant reduction of chain dimensions in water is explained by non-specific intramolecular hydrogen bonds, which in this chain can only form between backbone amide and backbone carbonyl groups or between backbone carbonyl groups and serine side chains. The intramolecular hydrogen bond network in the absence of

denaturant leads to decreased chain flexibility as indicated by a reduced intrachain diffusion constant (Figure 50). Together with results from recent TTET measurements, which showed accelerated loop formation in the absence of denaturant<sup>189</sup> our results suggests that non-specific chain compaction might be important for fast protein folding by accelerating the conformational search for favourable interactions. Formation of non-specific intramolecular hydrogen bonds might contribute significantly to chain compaction, which is often observed early during folding and is usually attributed to a hydrophobic collapse.

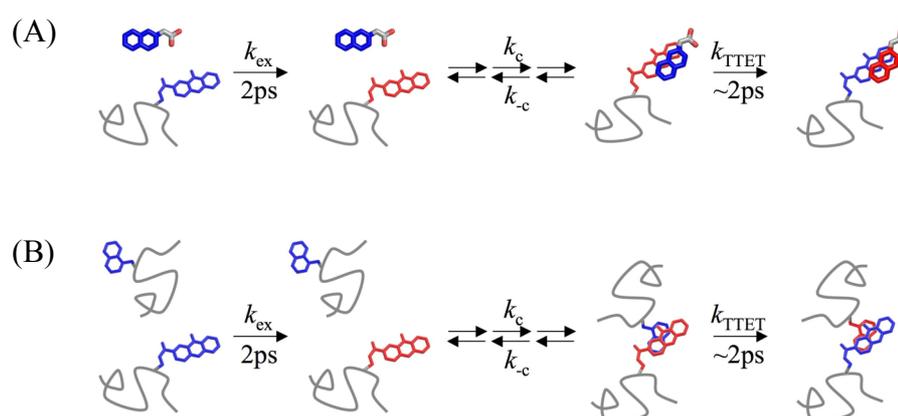
This work was carried out in collaboration with A. Möglich and published in reference 201 (see Appendix).

## 3.2 Dynamics of site-specific intermolecular interactions

Intermolecular interactions play an important role in many biological processes. The classical Smoluchowski theory models the associating species as uniformly reactive spheres and puts the maximum rate constant for association to  $5 \cdot 10^9 \text{ M}^{-1} \text{ s}^{-1}$ . The situation becomes much more complex when proteins are involved. Here, successful complex formation usually requires proper mutual orientation due to partial surface reactivity. These orientational constraints have a retarding effect on the association kinetics. The diffusion-controlled rate constant for site-specific protein-protein and protein-ligand interactions is typically assumed to be in the range of  $10^5 \text{ M}^{-1} \text{ s}^{-1}$  to  $10^6 \text{ M}^{-1} \text{ s}^{-1}$ .<sup>92-95,128</sup>

### 3.2.1 Intermolecular interactions between unfolded polypeptide chains

The method of TTET was previously applied to different model peptides and naturally occurring sequences to study intrachain contact formation between two specific points within a single polypeptide chain. Here, we apply TTET to measure diffusion-controlled formation of transient intermolecular interactions in different model systems (Figure 51). Part of the results described in this chapter was adapted from reference 202.



**Figure 51:** Scheme for intermolecular TTET from an unstructured peptide labelled with xanthone to A: 2-naphthylacetic acid (2-NAA) and B: an unstructured peptide labelled with naphthalene.

To study the kinetics of site-specific association by intermolecular TTET, we synthesized different model peptides labelled with either the triplet donor xanthone (Xan) or the triplet acceptor naphthylalanine (NAla) (Figure 52). The peptides vary either in their net charge to characterize the influence of intermolecular electrostatic forces on the bimolecular rate constant or in the number of reactive sites to study the role of orientational constraints. CD spectroscopy confirmed that the peptides do not show any stable secondary structure in solution as expected from their amino acid sequence.

peptide1: **GRSPGSRGSDpr(Xan)GSRGSFSG**  
 peptide2: **GDSFGSDGSDpr(Xan)GSDGSFSG**  
 peptide3: **GRSFGSRGSNAlaGSRGSFSG**  
 peptide4: **GRSNAlaGSRGSNAlaGSRGSNAlaSG**  
 peptide5: **GRSFGSRGSNAlaGSRGRSFSRG**  
 peptide6: **GRSNAlaGSRGSNAlaGSRGRSNAlaSRG**

*Figure 52: Amino acid sequence of model peptides synthesized for intermolecular TTET experiments. TTET labels are in green, positively and negatively charged residues in blue and red, respectively. The N-termini of all peptides are acetylated and C-termini are free.*

In order to determine the bimolecular rate constants ( $k_A$ ) for the association processes TTET experiments were performed under pseudo first-order conditions. Analysis of second-order kinetics is complex. The reaction rate  $v$  for a heterogeneous second order reaction of the type  $A + B \xrightarrow{k_A} C$  is

$$v = -k_A[A][B] \quad (36)$$

Equation 36 can be simplified if one of the reaction partners is present in large excess (minimally fivefold excess). For instance, if B is present in large excess, its concentration can be approximated to be constant throughout the reaction. Thus,  $[B]$  can be approximated by  $[B]_0$  and the rate law becomes apparent first-order.

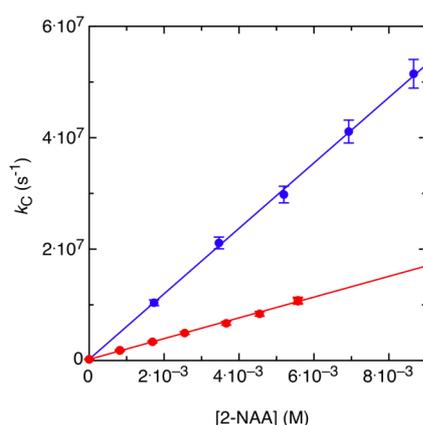
$$v = k'[A] \quad \text{with } k' = k_A[B]_0 \quad (37)$$

Reactions in which the observed kinetics is first-order but the mechanism is second-order are referred to as pseudo first-order reactions. To determine  $k_A$  in TTET experiments we used concentrations of the donor molecule between 30 to 50  $\mu\text{M}$  and an at least tenfold excess of acceptor molecules. Plotting the observed rate constant  $k_C$  ( $= k'$  in equation 37) against the acceptor concentration in a pseudo first-order plot according to

$$k_C = k_0 + k_A[\text{acceptor}] \quad (38)$$

reveals a linear relationship with the slope corresponding to the bimolecular rate constant  $k_A$ . The intersection with the y-axis corresponds to the rate constant  $k_0$  for the intrinsic decay of the donor triplet state in the absence of triplet acceptor.

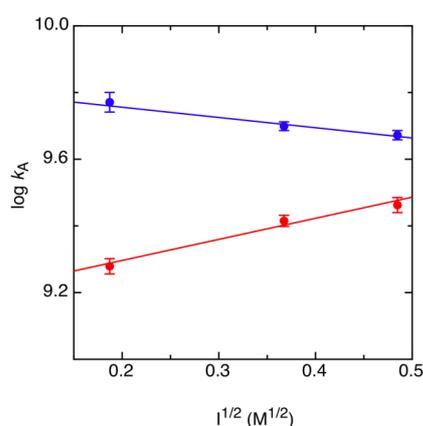
In the first part we studied association between an unfolded polypeptide chain and the small molecule 2-naphthylacetic acid (2-NAA). The results reveal bimolecular rate constants of  $k_A = 5.9(\pm 0.3) \cdot 10^9 \text{ M}^{-1}\text{s}^{-1}$  for association between peptide 1 and 2-NAA and of  $k_A = 1.9(\pm 0.2) \cdot 10^9 \text{ M}^{-1}\text{s}^{-1}$  for peptide 2 and 2-NAA in water (Figure 53 and table 11).



**Figure 53:** Pseudo first-order plot for contact formation of 2-NAA with peptide 1 (blue) and with peptide 2 (red) in 10mM KP pH7 at  $T = 22.5^\circ\text{C}$ . Adapted from reference 202.

An explanation for the difference in  $k_A$  is given by the Debye-Smoluchowski theory, which predicts that the diffusion-controlled rate constant for association of charged

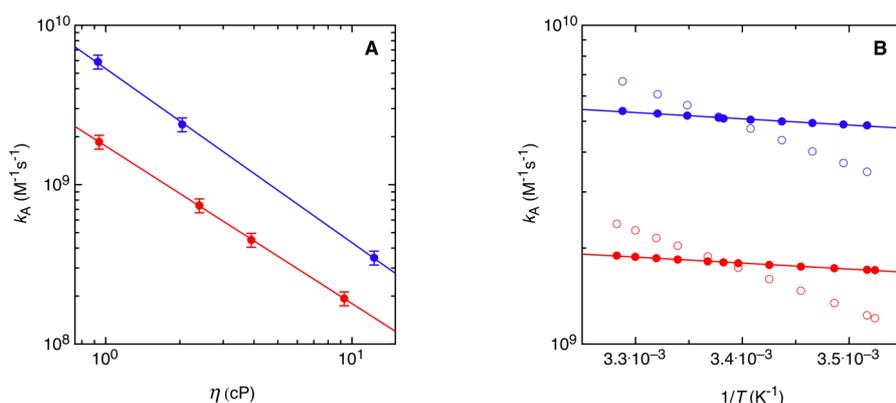
reaction partners is significantly influenced by electrostatic forces. To quantify electrostatic contributions to the association reaction we studied the effect of ionic strength on  $k_A$ . Results support the model in which attractive electrostatic interactions enhance association of peptide 1 with 2-NAA while repulsive electrostatic interactions slow down association of peptide 2 with 2-NAA. In the analyzed range we observed a linear dependence of  $\log k_A$  on the square root of the ionic strength (Figure 54) indicating that the effect of screening is not yet saturated in the presence of 200mM sodium chloride.



**Figure 54:** Effect of ionic strength on contact formation of 2-NAA with peptide 1 (blue) and with peptide 2 (red) in 10 mM KP pH7 at  $T = 22.5^\circ\text{C}$ . Ionic strength was increased by adding NaCl. Results from measurements in the absence and in the presence of 100 mM NaCl were adapted from reference 202.

The most straightforward experimental way to study whether a reaction is diffusion-controlled is to examine the viscosity dependence of the reaction. A diffusion controlled-reaction is proportional to the diffusion constant  $D$  and therefore, according to the Stokes-Einstein equation (Equation 18), inversely proportional to the viscosity of the solvent. Furthermore, in this limit the temperature dependence is weak with an apparent activation energy corresponding to the temperature dependence of the solvent viscosity. To test for diffusion control we investigated the effect of solvent viscosity on the kinetics of contact formation between peptide 1 and 2-NAA and between peptide 2 and 2-NAA (Figure 55A). Fitting equation 29 to the bimolecular rate constants measured at different solvent viscosities reveals that  $\beta = 1$  for both peptides. This indicates that the bimolecular rate constant decreases with  $1/\eta$ , as expected for a diffusion-controlled reaction. Further, we studied the effect of

temperature on  $k_A$  (Figure 55B). The bimolecular rate constants at different temperatures were corrected for the effect of temperature on solvent viscosity. Fitting the Arrhenius equation (Equation 32) to the viscosity-corrected data revealed activation energies of  $3.7(\pm 0.3)$  kJ/mol in both peptides. The small dependence of  $k_A$  on temperature supports the model of diffusion-controlled association.

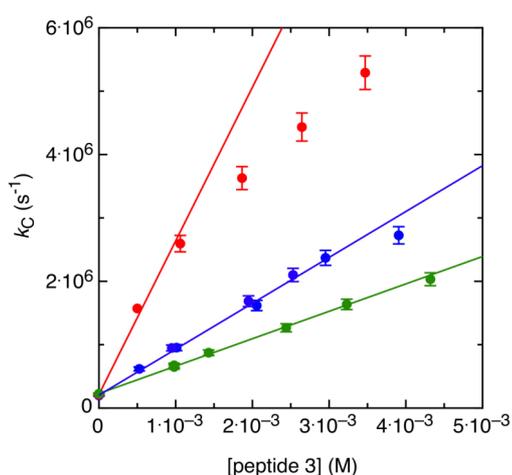


**Figure 55:** A) Effect of solvent viscosity on  $k_A$  for contact formation of 2-NAA with peptide 1 (blue) and with peptide 2 (red) in 10mM KP pH7 at  $T = 22.5^\circ\text{C}$ . The continuous lines represent fits of equation 29 to the data revealing  $\beta = 1.1(\pm 0.1)$  and  $\beta = 1.0(\pm 0.1)$ , respectively. The viscosity was modulated by adding glycerol. B) Effect of temperature on  $k_A$  for contact formation of 2-NAA with peptide 1 (blue) and with peptide 2 (red) in 10 mM KP pH7. Bimolecular rate constants determined at different temperatures (open circles) were corrected for the effect of temperature on solvent viscosity (closed circles). The continuous lines represent fits of equation 32 to the viscosity-corrected data revealing  $E_a = 3.7(\pm 0.3)$  kJ/mol for both peptides. Taken from reference 202.

All given values for bimolecular contact formation of 2-NAA with peptide 1 and with peptide 2 may be slightly too large due to uncertainties in the molar extinction coefficient for 2-NAA (see materials and methods).

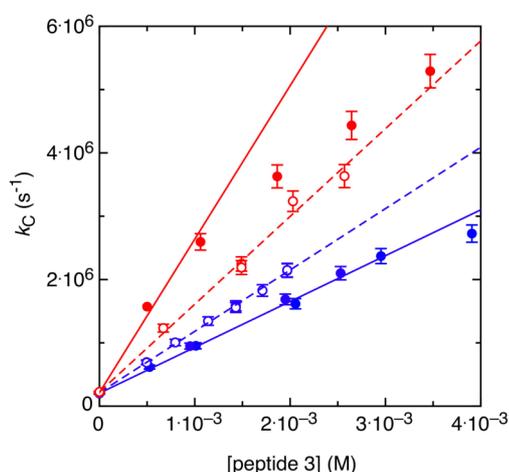
We investigated association of two unfolded polypeptide chains as a model for bimolecular steps during folding of oligomeric proteins. The TTET acceptor moiety was introduced in peptide 3 via the non-natural amino acid naphthylalanine. TTET measurements revealed a bimolecular rate constant of  $7.2 \cdot 10^8 \text{ M}^{-1}\text{s}^{-1}$  for association of peptide 1 with peptide 3 in water (Figure 56), which is almost ten times slower than contact formation of peptide 1 with the small molecule 2-NAA ( $k_A = 5.9 \cdot 10^9 \text{ M}^{-1}\text{s}^{-1}$ ). In the presence of 8 M urea, the association process is slowed down by a factor of

around two ( $k_A = 4.3 \cdot 10^8 \text{ M}^{-1} \text{ s}^{-1}$ , Figure 56). A possible explanation for the effect of urea on  $k_A$  is weakening of potential intramolecular interactions, which could be necessary to allow reorientation after initial contact formation. The pseudo first-order plot for association of peptide 2 with peptide 3 shows a nonlinear dependence of the apparent rate constant on the acceptor concentration (Figure 56). A possible explanation for this observation is a different mechanism for association of the oppositely charged polypeptide chains.



**Figure 56:** Pseudo first-order plot for association of peptide 1 with peptide 3 (blue) and for association of peptide 2 with peptide 3 (red) in 10 mM KP pH 7. Results for association of peptide 1 with peptide 3 in the presence of 8 M urea are shown in green. Adapted from reference 202.

Similar as for reaction between a small molecule and an unfolded polypeptide chain, site-specific contact formation between two unfolded peptides turned out to depend significantly on the net charge of the reactants (Figure 57). In the presence of salt electrostatic forces are screened and the charge effect becomes smaller (Figure 57). All bimolecular rate constants obtained for contact formation between two unfolded polypeptide chains are summarized in table 11.

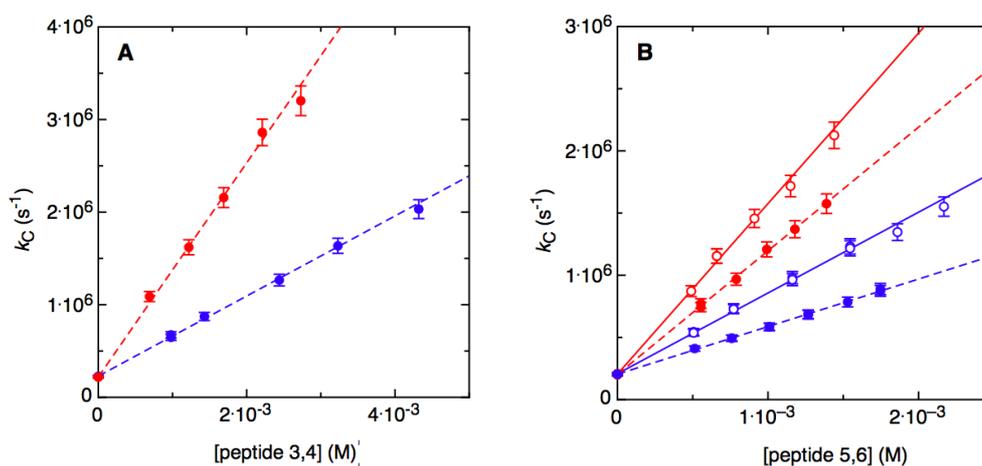


**Figure 57:** Pseudo first-order plot for association of peptide 1 with peptide 3 (blue) and of peptide 2 with peptide 3 (red) in 10mM KP pH7 at  $T = 22.5^\circ\text{C}$  in the absence (filled circles) and presence (open circles) of 100 mM NaCl. Data in the absence of NaCl was adapted from reference 202.

Oriental constraints due to partial reactivity can slow down bimolecular association significantly by decreasing the frequency of successful collisions. Reorientation during a single encounter event due to translational and rotational diffusion can attenuate this effect. In the following the extreme cases are described. The sliding model assumes that non-uniformly reactive molecules can meet in any orientation. Once the molecules are in contact they slide along their surfaces and search for the reactive site. The sliding step is a two-dimensional diffusion process and should be much faster than the three-dimensional search. Therefore, diffusion in solution and initial contact formation become rate-limiting in this model. The model of site-specific association suggests that the reaction partners have to meet directly at the reactive sites. If initial contact formation occurs somewhere else on the surface the molecules dissociate. A possibility to discriminate between the two models is to increase the number of reactive sites on one of the reaction partners. If the sliding model would be valid, this should not affect the bimolecular rate constant since only the two-dimensional diffusion step becomes faster. If initial contact formation has to occur directly at the reactive sites, the bimolecular rate constant should be increased by the factor of increased surface reactivity.

To increase the number of reactive sites in one of the used model peptides, a peptide containing three naphthylalanine residues was synthesized (peptide 4). The two phenylalanines in peptide 3 can be replaced for naphthylalanines in peptide 4 without changing chain properties significantly. Peptide 4 turned out to be poorly soluble and TTET measurements under pseudo first-order conditions could only be performed in the presence of denaturant (Figure 58A). In 8 M urea, contact formation between peptide 1 and peptide 4 occurs with  $k_A = 1.2 \cdot 10^9 \text{ M}^{-1}\text{s}^{-1}$ , which is almost three times faster than the reaction of peptide 1 with peptide 4 under the same conditions. The increase in  $k_A$  by a factor of almost three supports the model of site-specific contact formation. However, the presence of high urea concentrations might weaken potential intramolecular interactions, which could be necessary to allow reorientation after initial contact formation. To perform similar measurements in the absence of denaturant we synthesized peptides 5 and 6 whose solubility is increased by two additional arginine residues. In the presence of 8 M urea, reaction of peptide 1 with peptide 5 and with peptide 6 occurs with bimolecular rate constants of  $3.8 \cdot 10^8 \text{ M}^{-1}\text{s}^{-1}$  and  $9.9 \cdot 10^8 \text{ M}^{-1}\text{s}^{-1}$ , respectively (Figure 58B). This confirms the almost three-fold

increase in  $k_A$  comparing association of peptide 1 with peptide 3 to association of peptide 1 with peptide 4 in the presence of 8 M urea. Comparing association of peptide 1 with peptide 5 and with peptide 6 in water reveals an increase in  $k_A$  of only two (Figure 58B), indicating that reorientation/sliding during a single encounter in water can increase the probability of TTET.



**Figure 58:** A: Pseudo first-order plot for A: contact formation of peptide 1 with peptide 3 (blue) and with peptide 4 (red) in the presence of 8 M urea and B: contact formation of peptide 1 with peptide 5 (blue) and with peptide 6 (red) in water (open circles) and in the presence of 8 M urea (closed circles). All measurements were carried out in 10 mM KP pH7 at  $T = 22.5^\circ C$ . Results in panel A were adapted from reference 202.

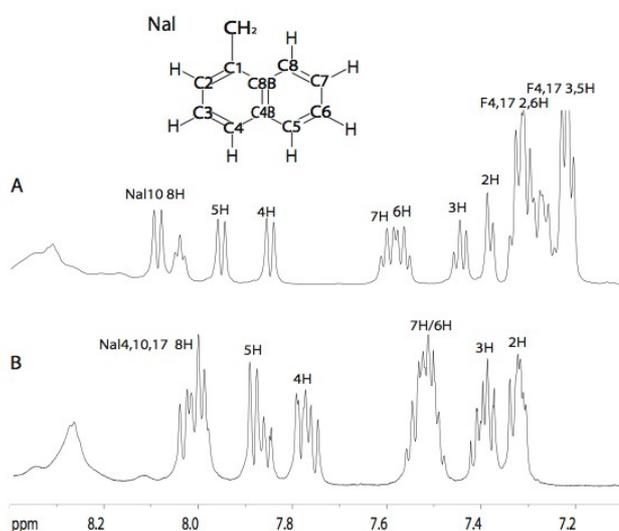
All values for  $k_A$  in the presence of urea were corrected for the effect of denaturant on solvent viscosity, which was calculated via an empirical equation introduced by Perl et al.<sup>203</sup> Table 11 summarizes all experimentally determined bimolecular rate constants obtained for association of an unfolded polypeptide chain with the small molecule 2-NAA and for association of two unfolded polypeptide chains.

Peptide	low I	100 mM NaCl	200 mM NaCl	8 M urea
1 & 2-NAA	$5.9(\pm 0.4) \cdot 10^9 \text{ M}^{-1}\text{s}^{-1}$	$5.0(\pm 0.2) \cdot 10^9 \text{ M}^{-1}\text{s}^{-1}$	$4.7(\pm 0.2) \cdot 10^9 \text{ M}^{-1}\text{s}^{-1}$	-
2 & 2-NAA	$1.9(\pm 0.1) \cdot 10^9 \text{ M}^{-1}\text{s}^{-1}$	$2.6(\pm 0.1) \cdot 10^9 \text{ M}^{-1}\text{s}^{-1}$	$2.9(\pm 0.2) \cdot 10^9 \text{ M}^{-1}\text{s}^{-1}$	-
1 & 3	$7.2(\pm 0.5) \cdot 10^8 \text{ M}^{-1}\text{s}^{-1}$	$9.7(\pm 0.5) \cdot 10^8 \text{ M}^{-1}\text{s}^{-1}$	-	$4.3(\pm 0.2) \cdot 10^8 \text{ M}^{-1}\text{s}^{-1}$
2 & 3	$(2.4(\pm 0.2) \cdot 10^9 \text{ M}^{-1}\text{s}^{-1})$	$1.4(\pm 0.2) \cdot 10^9 \text{ M}^{-1}\text{s}^{-1}$	-	-
1 & 4	-	-	-	$1.2(\pm 0.1) \cdot 10^9 \text{ M}^{-1}\text{s}^{-1}$
1 & 5	$6.5(\pm 0.4) \cdot 10^8 \text{ M}^{-1}\text{s}^{-1}$	-	-	$3.8(\pm 0.2) \cdot 10^8 \text{ M}^{-1}\text{s}^{-1}$
1 & 6	$1.4(\pm 0.2) \cdot 10^9 \text{ M}^{-1}\text{s}^{-1}$	-	-	$9.9(\pm 0.6) \cdot 10^8 \text{ M}^{-1}\text{s}^{-1}$

**Table 11:** Bimolecular rate constants in ( $\text{M}^{-1}\text{s}^{-1}$ ) for association between two unstructured model peptides and between an unstructured model peptide and 2-NAA. All measurements were carried out at pH 7 and  $T = 22.5^\circ\text{C}$ . Stern-Volmer analysis for contact formation of peptide 2 with peptide 3 did not reveal a linear relation over the studied concentration range (Figures 56 and 57,  $k_A$  was determined from a linear fit to the data at low acceptor concentrations). Results from measurements in the presence of 8 M urea were corrected for the effect of urea on solvent viscosity.

The naphthalene groups in peptide 6 might form intramolecular hydrophobic interactions and by this lead to a deviation from flexible random coil structure. To detect putative hydrophobic clusters we investigated peptides 5 and 6 by NMR spectroscopy using chemical shift analysis of the aromatic side chain residues and measurements of nuclear Overhauser effects (NOE). The observed chemical shifts represent an average of the chemical shifts over all conformations present in solution weighted by their population. In contrast, the NOE intensity decreases with the inverse sixth power of the distance between two protons and only protons with a distance of less than 5 Å contribute to the overall intensity of the corresponding NOE cross peak.

It was possible to assign all proton resonances of the naphthylalanine residue of peptide 5 (Figure 59A) except for the amide backbone proton, which is in exchange with the solvent. The proton resonances of all amino acid residues are degenerate. In peptide 6 the naphthylalanine residues are also degenerate (Figure 59B).

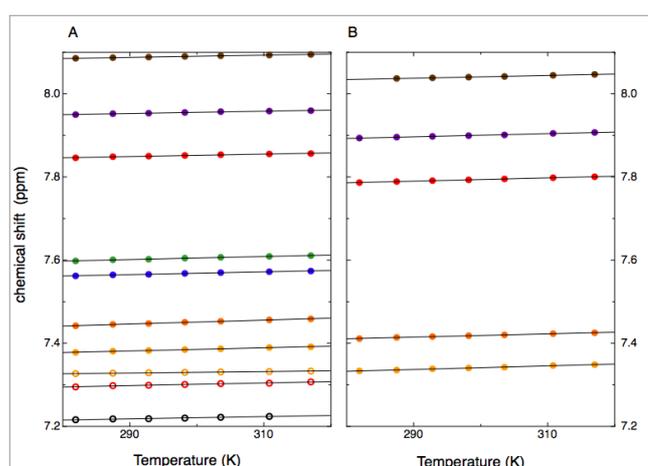


**Figure 59:** Aromatic region of the 1D  $^1\text{H}$  NMR spectrum for peptide 5 (A) and peptide 6 (B). Spectra were recorded at peptide concentrations of 1.6 mM in 10 mM KP pH 7.0 (95%  $\text{H}_2\text{O}$ , 5%  $\text{D}_2\text{O}$ ) at  $T = 22.5^\circ\text{C}$ . DSS was used as reference for  $^1\text{H}$  chemical shift. The spectra were acquired by averaging 256 scans.

All amide protons, except for two glycines, show line broadening due to exchange with water. The temperature coefficient for the amide proton of one of these glycines was determined to be  $-6.6$  ppb/K, which is typical for an amide proton in a random coil conformation.<sup>204</sup> Above 300 K all amide protons are in fast exchange with water. These observations suggest that the peptide does not contain stable hydrogen bonds

that are necessary for stable elements of secondary structure. The spectrum of peptide 6 did not show any detectable concentration dependence. A spectrum obtained with a peptide concentration of 0.016 mM was essentially identical to that of 1.6 mM peptide concentration, indicating that the spectral dispersion is not the result of aggregation.

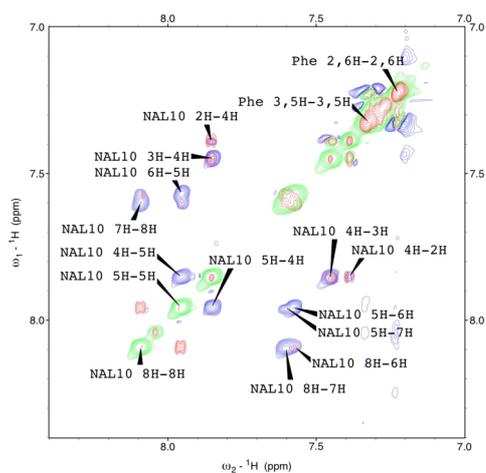
Further experiments based on the temperature dependence of 1D  $^1\text{H}$  NMR spectra for peptide 5 and peptide 6 showed that the aromatic resonances of both peptides are not significantly affected by temperature (Figure 60). The resonances of all side chain protons in both peptides shift linearly downfield towards the random coil values as the temperature is increased. The temperature coefficients of all aromatic proton resonances of peptide 5 were in the range of 0.1 to 0.5 ppb/K (Figure 60A). Due to degeneracy of the side chain resonances in peptide 6 the temperature coefficients for only one naphthylalanine could be determined (Figure 60B), whereas an upper limit of 1 ppb/K was estimated for the other two naphthylalanines.



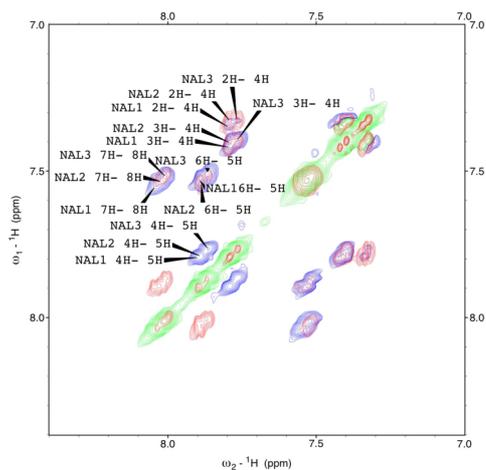
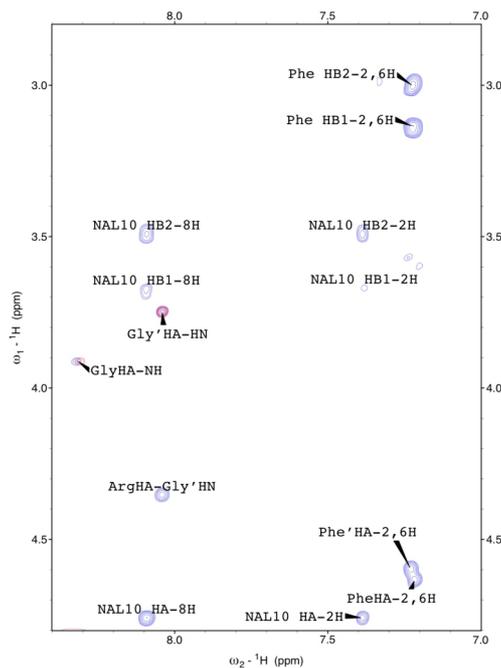
**Figure 60:** Temperature dependence of the  $^1\text{H}$  NMR chemical shifts of the aromatic protons in peptide 5 (A) and peptide 6 (B). The symbols correspond to the following assignments: 8H (●), 7H (●), 6H (●), 5H (●), 4H (●), 3H (●) and 2H (●) of naphthylalanines and 3,5H (○,○) 2,4H (○) of the phenylalanines.

The relatively small chemical shift deviation of the aromatic proton resonances from random coil and the small temperature dependences of these resonances suggest that even if the aromatic rings are involved in some kind of hydrophobic interaction there is substantial relative motion of the residues on a rapid timescale.

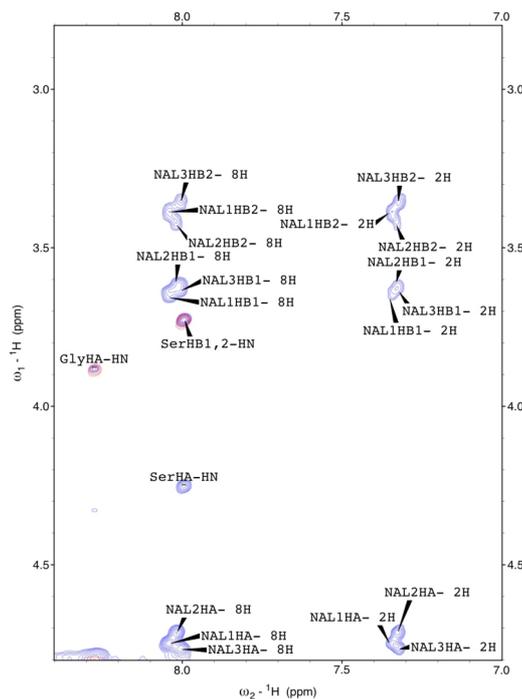
For both peptides there were no NOE cross peaks observed which are not in agreement with a random coil structure. Particularly no intramolecular NOE cross signals were detected between the ring protons of the aromatic groups, neither in the NOESY nor in the ROESY experiments.



**Figure 61A:** Sections from the ROESY spectrum of peptide 5. Spectra were recorded at a peptide concentration of 1.6 mM in 10 mM KP pH 7.0 (95% H<sub>2</sub>O, 5% D<sub>2</sub>O) at T = 22.5 °C.



**Figure 61B:** Sections from the ROESY spectrum of peptide 6. Spectra were recorded at a peptide concentration of 1.6 mM in 10 mM KP pH 7.0 (95% H<sub>2</sub>O, 5% D<sub>2</sub>O) at T = 22.5 °C.



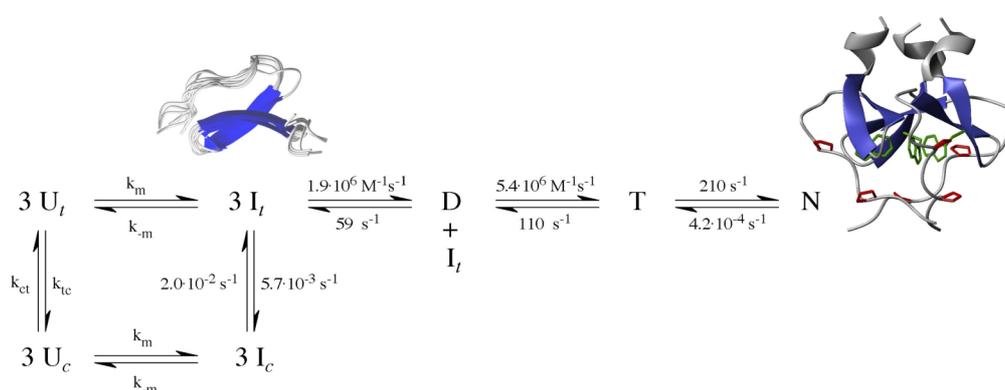
Neri et al. estimated that NOESY/ROESY cross peaks would be detectable if the population of a folded conformation were more than 10%,<sup>193</sup> indicating that there are no significant populations of conformations of hydrophobic clusters present in solution for peptide 5 and peptide 6.

Aromatic ring current effects<sup>205</sup> are the main source for chemical shift deviations of side chain protons in native proteins from random coil values measured in short peptides.<sup>206</sup> The observed tendency of side chain protons for upfield shifts in native protein spectra can be explained by the fact that in closely packed structures protons experiencing upfield ring current shifts are on average closer to the aromatic ring than those experiencing downfield ring current shifts.<sup>207</sup> Therefore the net effect of random encounters of side chain protons with aromatic rings in a random structure would be expected to be an upfield shift. Compared to random coil chemical shifts, we observe upfield shifts for all side chain protons in peptide 5 and peptide 6. The aromatic protons of the two phenylalanines in peptide 5 show an upfield shift of 0.08ppm compared to random coil chemical shifts. Replacing the two phenylalanines for naphthylalanines in peptide 6 increases the upfield shifts. The aromatic proton resonances of the three naphthylalanines shift within 0.09ppm upfield, relative to those of peptide 5. Because the ring current effects of the double ring of naphthylalanine are larger than those of a single ring these additional upfield shifts would be expected for a random structure. In summary, the NMR measurements proof that peptide 6 is unfolded. NMR measurements and NMR data evaluations were performed in collaboration with J. Habazettl (Biozentrum Basel).

Compared to the generally assumed diffusion limit for site-specific protein-protein interactions of  $10^5 \text{ M}^{-1}\text{s}^{-1}$  to  $10^6 \text{ M}^{-1}\text{s}^{-1}$ ,<sup>92-95,128</sup> our TTET measurements revealed rate constants in the range of  $10^8 \text{ M}^{-1}\text{s}^{-1}$  to  $10^9 \text{ M}^{-1}\text{s}^{-1}$  for association between an unfolded polypeptide chain and a small molecule and between two unfolded polypeptide chains. The results further revealed that electrostatic interactions significantly contribute to the bimolecular rate constant. Similar effects were observed for many protein-protein interactions.<sup>128</sup> Testing for orientational constraints during association between two unfolded polypeptide chains showed that not every random encounter is successful. However, reorientation of the interacting polypeptide chains during a single encounter can increase the rate constant of productive association.

### 3.2.2 Role of electrostatic interactions in the assembly of the foldon domain

In many cases protein-protein interactions are assisted by electrostatic forces, which may drastically accelerate bimolecular reactions. Beside a general acceleration of protein-protein association through electrostatic attractive forces, electrostatic interactions may guide the interaction partners into proper mutual orientation and thus increase the frequency of successful collisions.<sup>119</sup> We used the small homotrimeric foldon domain from bacteriophage T4 fibritin as a model to investigate the formation of site-specific protein-protein interactions during folding and assembly of a trimeric protein. Previous results from kinetic folding studies showed that the isolated 27 amino acid foldon domain spontaneously folds into a very stable trimer in a series of very fast folding and association steps (Figure 62). During folding from the GdmCl-unfolded state a monomeric intermediate is formed on the submillisecond time scale.<sup>111</sup> This intermediate has native structure but largely increased flexibility and reduced stability compared to a subunit in the native trimer.<sup>112</sup> Assembly of the intermediate into a trimeric state occurs in two consecutive bimolecular reactions via a dimeric intermediate with rate constants of  $1.9 \cdot 10^6 \text{ M}^{-1}\text{s}^{-1}$  and  $5.4 \cdot 10^6 \text{ M}^{-1}\text{s}^{-1}$ , respectively. The resulting trimer is rapidly converted to the final native state in a first order reaction with  $k = 210 \text{ s}^{-1}$ .<sup>111</sup> In addition to this fast folding pathway, about 20% of the molecules fold in a slow process that is limited by prolyl isomerization reactions at the prolyl residues at positions 4 and/or 7.<sup>111</sup>



**Figure 62:** Folding/assembly mechanism of the foldon domain in the presence of 0.58 M GdmCl.<sup>111</sup>

The NMR structure of the monomeric folding intermediate revealed oppositely charged sites, which suggests an important role of electrostatic interactions for association.<sup>112</sup> Previous studies on the assembly kinetics of the foldon domain were carried out at a residual GdmCl concentration of 0.58 M, which largely shields electrostatic interactions.

To investigate the contributions of electrostatic interactions to the stability of the foldon domain we measured urea-induced equilibrium unfolding transitions at low salt concentrations (10 mM NaCac) and in the presence of additional 150 mM NaCl, which resembles physiological ionic strength. The transitions were monitored by changes in intrinsic Trp fluorescence at 320 nm and by changes in the far UV CD signal at 228 nm to test for the validity of the two-state model. Figure 63 shows the urea-induced unfolding transition at low ionic strength performed at monomer concentrations ( $[M]_0$ ) of 5  $\mu$ M and 13  $\mu$ M. The transition is shifted to higher urea concentrations with increasing  $[M]_0$  but is independent of the probe used to monitor the transition indicating a two state-transition between folded trimers (N) and unfolded monomers (U)

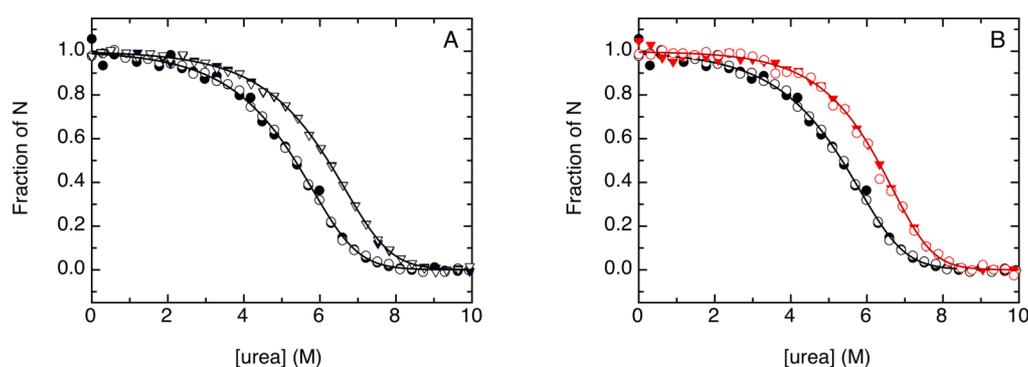


The concentration dependence of the equilibrium between N and U can be described by

$$K_{eq} = \frac{3f_u^3[M]_0^2}{1-f_u} \quad (40)$$

where  $K_{eq}$  is the equilibrium constant,  $[M]_0$  is the total monomer concentration ( $[M]_0 = [U] + 3[N]$ ) and  $f_u$  is the fraction of unfolded monomer in equilibrium.<sup>111</sup> A global fit to the transitions at 5  $\mu$ M and 13  $\mu$ M yields a free energy of folding in the absence of urea of  $\Delta G^0(\text{H}_2\text{O}) = -89.2(\pm 0.3)$  kJ/mole, which is identical to the value obtained from GdmCl-induced transitions ( $\Delta G^0(\text{H}_2\text{O}) = -89.2(\pm 0.6)$  kJ/mol).<sup>111</sup> The change in free energy with urea concentration ( $m_{eq} = \partial\Delta G^0/\partial[\text{urea}]$ ) is  $5.4(\pm 0.1)$  (kJ/mol)/M, which is the value expected from the change in SASA upon unfolding of the native

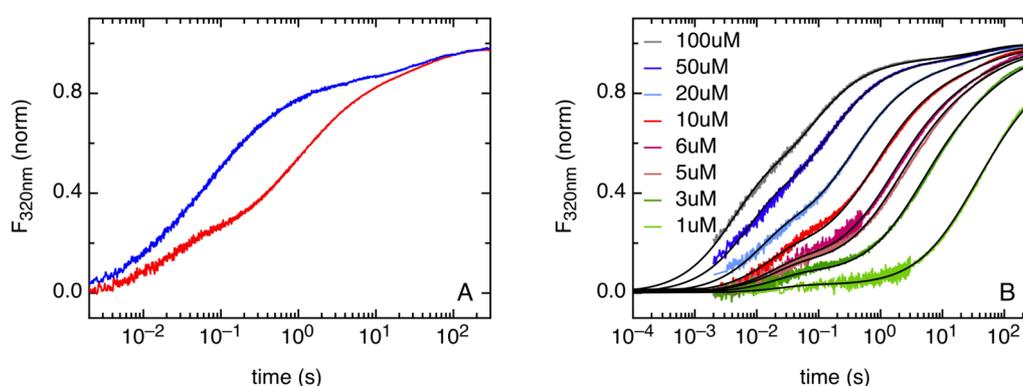
trimer.<sup>3</sup> Figure 63B shows that the addition of 150 mM NaCl increases the stability of the foldon domain. A fit of the transitions in the presence of 150 mM NaCl gives  $\Delta G^0(\text{H}_2\text{O}) = -98.8(\pm 0.5)$  kJ/mol and  $m_{\text{eq}} = 6.1(\pm 0.1)$  (kJ/mol)/M, indicating a 9.6 kJ/mol stabilization compared to low ionic strength. This increase in trimer stability in the presence of 150 mM NaCl is not expected from the position of NaCl as a neutral Hofmeister salt regarding protein stabilization and from the previously observed significant stabilization of the native trimer by intermolecular salt bridges between Gly5 and Arg15 of adjacent subunits.<sup>112,208</sup>



**Figure 63:** A: Equilibrium unfolding transitions at low ionic strength measured at  $[M]_0 = 5 \mu\text{M}$  (circles) and  $[M]_0 = 13 \mu\text{M}$  (triangles). B: Equilibrium unfolding transitions at low ionic strength at  $[M]_0 = 5 \mu\text{M}$  (black) and in the presence of 150 mM NaCl at  $[M]_0 = 5 \mu\text{M}$  (red). The transitions were monitored by changes in Trp fluorescence at 320 nm (closed symbols) and by changes in far UV CD-signal at 228 nm (open symbols).

Previous kinetic studies on the folding/assembly mechanism of the foldon domain were performed in the presence of a residual salt concentration of 0.58 M GdmCl<sup>111</sup>, which largely shields electrostatic interactions. To investigate the role of electrostatic interactions for the two bimolecular association reactions we decreased the ionic strength by unfolding the foldon domain in 6 M urea, pH 3.0 and monitored refolding after stopped-flow mixing into final conditions of 1 M urea at pH 7.0 in 10 mM NaCac buffer. At this urea concentration the foldon domain shows the same stability as in the presence of 0.58 M GdmCl ( $\Delta G_0 = -83.5$  kJ/mole), which eliminates effect from differences in protein stability. Figure 64A compares stopped-flow folding kinetics of the foldon domain at a monomer concentration of  $[M]_0 = 10 \mu\text{M}$  in the presence of 0.58 M GdmCl and of 1.0 M urea monitored by the change in intrinsic

Trp fluorescence at 320 nm. In the presence of 0.58 M GdmCl the majority of foldon monomers assemble in an apparent third order reaction reflecting a monomer-dimer pre-equilibrium followed by a second-order reaction between a monomer and a dimer to form a trimer (Figure 62).<sup>111</sup> Since the stability of the dimer is low and trimer formation is faster than dimer formation, the dimeric intermediate populates only to very low amounts, which leads to the observed apparent third order kinetics for formation of the native trimer when the concentration-dependence of the half-life of the assembly reaction is analyzed.<sup>111</sup> Additionally, about 20% of the fluorescence changes in a parallel slow reaction that is limited by prolyl isomerization in unfolded molecules that contain non-native *cis* prolyl isomers at the prolyl peptide bonds Thr3-Pro4 and/or Ala6-Pro7.<sup>111</sup> Figure 64A shows that assembly of the foldon domain is slowed down at low ionic strength and an additional kinetic step is resolved resulting in three distinct reactions. About 10% of the fluorescence changes in a slow process that becomes concentration-independent at very high protein concentrations (Figure 64B), compatible with the prolyl isomerization-limited reaction observed in the presence of GdmCl. To further characterize the three reactions observed during folding/assembly of the foldon domain we measured the assembly kinetics at different initial monomer concentrations ( $[M]_0$ ). Figure 64B shows that the two faster processes become faster with increasing protein concentration over the whole concentration range and the amplitude of the fastest step increases with  $[M]_0$ . This behaviour indicates that the two faster reactions reflect two consecutive second order steps in the assembly process and suggests the transient population of the dimeric intermediate (D; see Figure 62) in the fastest reaction.

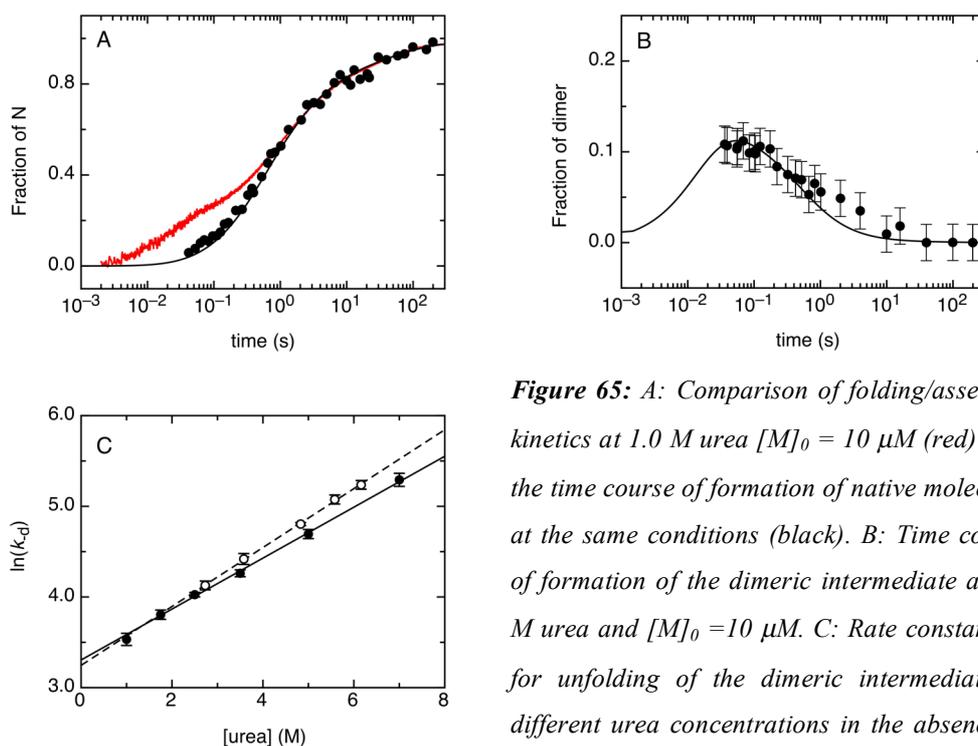


**Figure 64:** A: Refolding/assembly kinetics at 0.58 M GdmCl pH 7.0 (red) and at 1.0 M urea pH 7.0 (blue). B: Concentration-dependence of the refolding/assembly kinetics at 1.0 M urea pH 7.0.

The model of transient population of the dimeric intermediate in the fastest reaction during refolding at low ionic strength is confirmed by the results from interrupted refolding experiments, which monitor the time course of formation of native molecules (Figure 65). Unfolded foldon monomers (in 6.0 M urea, pH 3.0) were diluted into final conditions of 1 M urea, pH 7.0 at  $[M]_0 = 10 \mu\text{M}$  to initiate refolding and assembly under the same conditions as in the fluorescence-detected experiments (Figure 65A). After various times,  $t_i$ , refolding is interrupted by a second mixing step into final conditions of 6.7 M GdmCl, 0.17 M urea, pH 7.0. Under these conditions the native foldon domain unfolds with a time constant of  $9.0(\pm 1.0)$  s. The amplitude of this unfolding reaction is a measure for the amount of native molecules that was formed after the time  $t_i$ , when refolding was interrupted. Allowing the refolding process to go to completion ( $t_\infty$ ) gives the unfolding amplitude of 100% native molecules, which is used for a normalization of the data to obtain the absolute fraction of N formed after  $t_i$ . Figure 65A compares the time course of formation of native trimer at low salt conditions determined in interrupted refolding experiments with the fluorescence detected assembly kinetics under identical conditions. Only the two slower processes produce native foldon trimer, whereas the faster concentration-dependent reaction is not leading to the native trimer. This supports the model that the fastest observable reaction reflects formation of a dimeric intermediate, whereas the two slower reactions monitor formation of the native state on two parallel pathways limited by trimerization or by prolyl isomerization. A second, faster unfolding process is observed in the interrupted refolding experiments with  $\tau = 116(\pm 10)$  ms. The amplitude of this reaction increases up to  $t_i = 20$ s and decreases to zero on the same time scale as the slow, prolyl-isomerization-limited assembly process occurs. This observation indicates that molecules with non-native *cis* prolyl bonds can form an intermediate, which is only slowly converted to the native all-*trans* state.

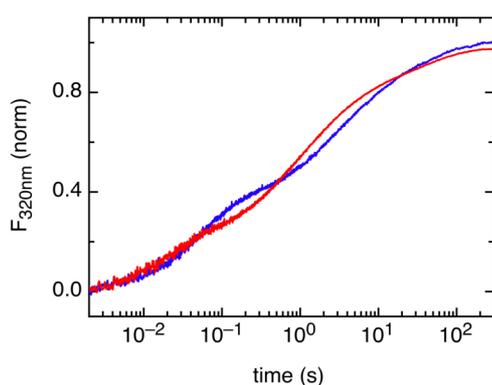
Comparison of the fluorescence-detected kinetics with the results from interrupted refolding experiments reveals the transient population of a dimeric intermediate. Unfolding of this intermediate is not observed in the interrupted refolding experiments, which may be due to a very fast unfolding reaction under the applied strong unfolding conditions. To resolve the unfolding reaction of the dimeric intermediate we performed interrupted refolding experiments with less destabilizing

conditions in the final unfolding step. In the first mixing step the intermediate was allowed to populate for various times,  $t_i$ , and in the second mixing step unfolding is induced in the presence of 2.5 M urea. At this urea concentration the native foldon trimer is still about 98% populated (Figure 63). Figure 65B shows that a fast unfolding reaction with  $\tau = 18(\pm 2)$  ms is observed in these experiments. The amplitude of this reaction increases to its maximum value within the dead time of the interrupted refolding experiments (35 ms) and decreases on the same time scale as native trimer is formed, indicating that this reaction reflects the dimer to trimer interconversion. The observed fast unfolding reaction at 2.5 M urea does not reflect unfolding of the monomeric intermediate, since the monomeric E5R variant, which is structurally identical to the transient monomeric folding intermediate, unfolds within the dead time of the experiment under the same experimental conditions. This result is in agreement with NMR studies on the E5R variant, which showed fast folding/unfolding dynamics on the 100  $\mu$ s time scale. Varying the urea concentration in the unfolding step of the interrupted refolding experiments allows a further characterization of the dimeric intermediate (Figure 65C). Unfolding becomes faster with increasing urea concentration with a kinetic  $m_u$ -value ( $m_u = \partial\Delta G^{0\ddagger}/\partial[\text{urea}]$ ) of  $-0.69(\pm 0.03)$  (kJ/mol)/M. These data yield a rate constant for dimer unfolding ( $k_d$ ) of  $36(\pm 1)$  s<sup>-1</sup> under the conditions applied in the refolding experiments (1.0 M urea). This information is used in the global analysis for the determination of all rate constants for the assembly process described below.



**Figure 65:** A: Comparison of folding/assembly kinetics at 1.0 M urea  $[M]_0 = 10 \mu\text{M}$  (red) with the time course of formation of native molecules at the same conditions (black). B: Time course of formation of the dimeric intermediate at 1.0 M urea and  $[M]_0 = 10 \mu\text{M}$ . C: Rate constant  $k_d$  for unfolding of the dimeric intermediate at different urea concentrations in the absence of salt (filled circles) and in the presence of 150 mM NaCl (open circles).

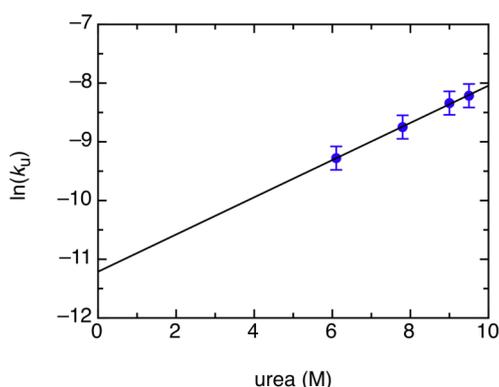
Refolding experiments in the presence of glycerol support the model of transient population of the dimeric intermediate in the fast concentration-dependent reaction observed at low ionic strength. The addition of glycerol slows down the two concentration-dependent kinetic phases as expected for bimolecular reactions (Figure 66 and table 12).



**Figure 66:** Refolding/assembly kinetics of the foldon domain at 1.0 M urea (red) and at 1.0 M urea in the presence of 35% (w/v) glycerol (blue). Measurements were carried out at  $[M]_0 = 10 \mu\text{M}$  in 10 mM NaCac pH 7.0 at  $T = 20 \text{ }^\circ\text{C}$ .

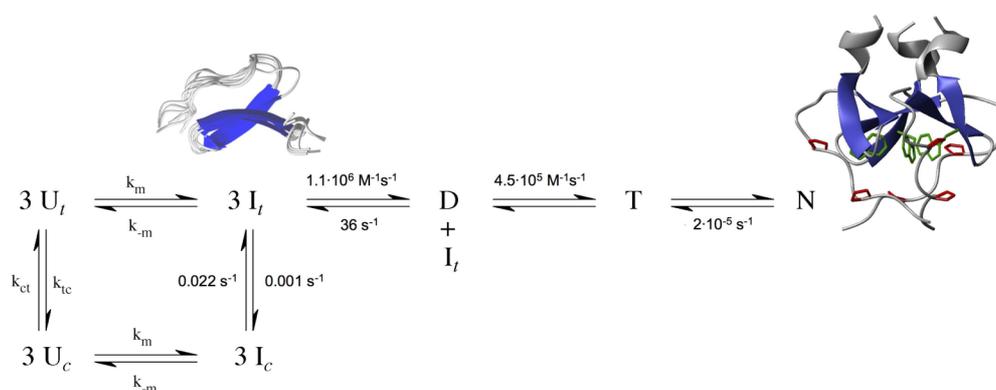
To determine the rate constants for all reactions involved in the folding/assembly process of the foldon domain (Figure 68) at low-salt conditions, we globally fitted the fluorescence detected refolding experiments at various  $[M]_0$  (Figure 64B) together with the time course of formation of native molecules (Figure 65A) and of the dimeric

folding intermediate (Figure 65B). In the global analysis the rate constant for unfolding of the dimeric intermediate was fixed to the value of  $36 \text{ s}^{-1}$ , as determined in the interrupted refolding experiments shown in figure 65C. Additionally, unfolding of the native trimer was measured at various urea concentrations which yielded a rate constant  $k_u = 2.0(\pm 0.2) \cdot 10^{-5} \text{ s}^{-1}$  at 1.0 M urea (Figure 67) which was also fixed in the global analysis. The fractions of native and unfolded molecules at 1.0 M urea calculated from the urea-induced equilibrium transition were used as additional constraint.



**Figure 67:** Unfolding of the native foldon trimer in the presence of different urea concentrations. Measurements were carried out in 10 mM NaCac pH 7.0 at  $T = 20^\circ\text{C}$ .

The results from the global fit are shown in figures 64B, 65A and 65B and the rate constants for the different reactions are given in figure 68. Comparison of the assembly process at isostability conditions in the presence of 0.58 M GdmCl (Figure 62) and of 1.0 M urea (Figure 68) shows that a change in ionic strength affects formation of the trimer ( $k_t$ ) more than formation of the dimer ( $k_d$ ). As a result, trimer formation is faster than dimer formation at high ionic strength and slower than dimer formation at low ionic strength.



**Figure 68:** Folding/assembly process of the foldon domain in the presence of 1.0 M urea.

The global fit of the assembly kinetics additionally yields the spectroscopic parameter for the dimeric intermediate ( $S_D$ ), which represents the fractional change in fluorescence upon dimer formation compared to the fluorescence change upon formation of the native trimer. For refolding in the presence of 1.0 M urea a  $S_D$ -value of 0.44 is obtained, which is virtually identical to the  $S_D$ -value of 0.43 obtained for assembly in the presence of 0.58 M GdmCl. The rate constant for folding of the trimeric intermediate to the native trimer ( $k_f$ , see figure 68) cannot be obtained for assembly in the presence of 1.0 M urea, since trimerization is slower than the subsequent folding step at all applied monomer concentrations.

The transient population of the dimeric intermediate during assembly of the foldon domain at low ionic strength allows the independent determination of the rate constant for dimerization ( $k_d$ ) by analyzing the concentration-dependence of the initial slope of the fluorescence change according to

$$\left(\frac{d[D]}{dt}\right)_{t=0} = k_d[M]_0^2 \quad (41)$$

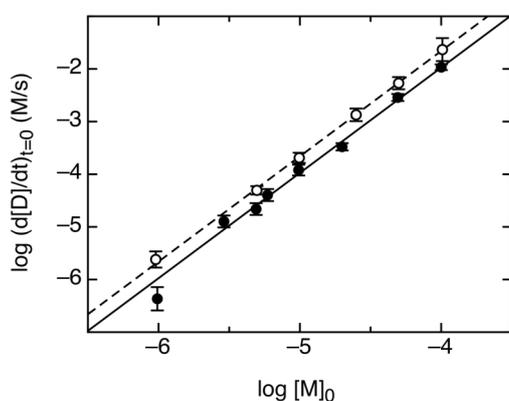
To normalize the fluorescence detected assembly kinetics to the actual amount of dimer we normalized the final fluorescence by the concentration of native trimer present in equilibrium at any given protein concentrations and by the spectroscopic parameter ( $S_D$ ) of the dimer. According to the equation

$$\log\left(\frac{d[D]}{dt}\right)_{t=0} = 2\log[M]_0 + \log k_d \quad (42)$$

a plot of  $\log(d[D]/dt)_{t=0}$  vs.  $\log[M]_0$  yields the rate constant for the initial fast process and should have a slope of 2 for a second-order reaction. Figure 69 shows a linear dependence of  $\log(d[D]/dt)_{t=0}$  vs.  $\log[M]_0$  with a slope of  $2.0(\pm 0.1)$ , which confirms that the fast process represents a second-order reaction leading to the dimeric intermediate. The linear fit of equation 42 to the data additionally yields  $k_d = 1.1(\pm 0.5) \cdot 10^6 \text{ M}^{-1}\text{s}^{-1}$ , in good agreement with the results from the global fit (Figure

68). These results confirm the validity of the model shown in figure 68 and indicate the initial formation of a dimeric intermediate in a second order reaction.

In the previous study on refolding in the presence of 0.58 M GdmCl the reaction order was determined by analyzing the half-life of the reaction at different  $[M]_0$  and the rate constant for dimerization was obtained from a global fit of the refolding kinetics at various  $[M]_0$  in combination with results from interrupted refolding experiments.<sup>111</sup> Since the analysis of the initial slope of the reaction at  $t = 0$  should give information on the initial dimerization step even when the dimer is not highly populated, we plotted  $\log(d[D]/dt)_{t=0}$  vs.  $\log[M]_0$  for the fluorescence-detected assembly kinetics in the presence of 0.58 M GdmCl. The kinetic traces were again normalized to the concentration of dimer as described above. For refolding in the presence of 0.58 M GdmCl the slope also yields a reaction order of  $2.0(\pm 0.1)$  and the fit gives a value of  $k_d = 2.1(\pm 0.2) \cdot 10^6 \text{ M}^{-1}\text{s}^{-1}$  (Figure 69), which is in good agreement with the result previously obtained from a global fit of the concentration-dependence of assembly (Figure 62) and confirms the result that dimer formation is slightly accelerated with increasing ionic strength.



**Figure 69:** Concentration-dependence of the logarithm of the initial slope of fluorescence-detected assembly kinetics in the presence of 0.58 M GdmCl (open circles) and in the presence of 1.0 M urea (closed circles).

Comparison of the assembly kinetics of the foldon domain in the presence of urea and GdmCl at identical protein stability shows that assembly is slowed down at low ionic strength. To investigate the effect of ionic strength on the assembly kinetics in more detail we monitored refolding in the presence of different concentrations of NaCl at a residual denaturant concentration of 1.0 M urea. Figure 70A shows that increasing ionic strength accelerates both dimer and trimer formation with a stronger effect on the trimerization reaction, which becomes faster than dimer formation at high ionic

strength. The dimerization rate constant ( $k_d$ ) can be calculated from the initial slope of the assembly reaction at a given ionic strength ( $I$ ) using the equation

$$k_d = \left( \frac{d[D]}{dt} \right)_{t=0} \cdot \frac{1}{[M]_0^2} \quad (43)$$

Plotting  $\log k_d$  vs.  $I^{1/2}$  yields information on the contributions from electrostatic interactions to dimer formation. According to the equation

$$\log k_d = \log k_d^0 + 1.02 z_A z_B I^{1/2} \quad (44)$$

the slope of this plot yields the primary salt effect, which depends on the effective charges that contribute to the electrostatic of dimer formation. Figure 70B reveals a slope of  $1.0(\pm 0.1)$ , indicating that a single charge on each monomer is responsible for the observed salt effect, in agreement with a single negative net charge of a foldon monomer. The interaction energy of the observed effect of charges on the dimerization reaction can be obtained from the equation

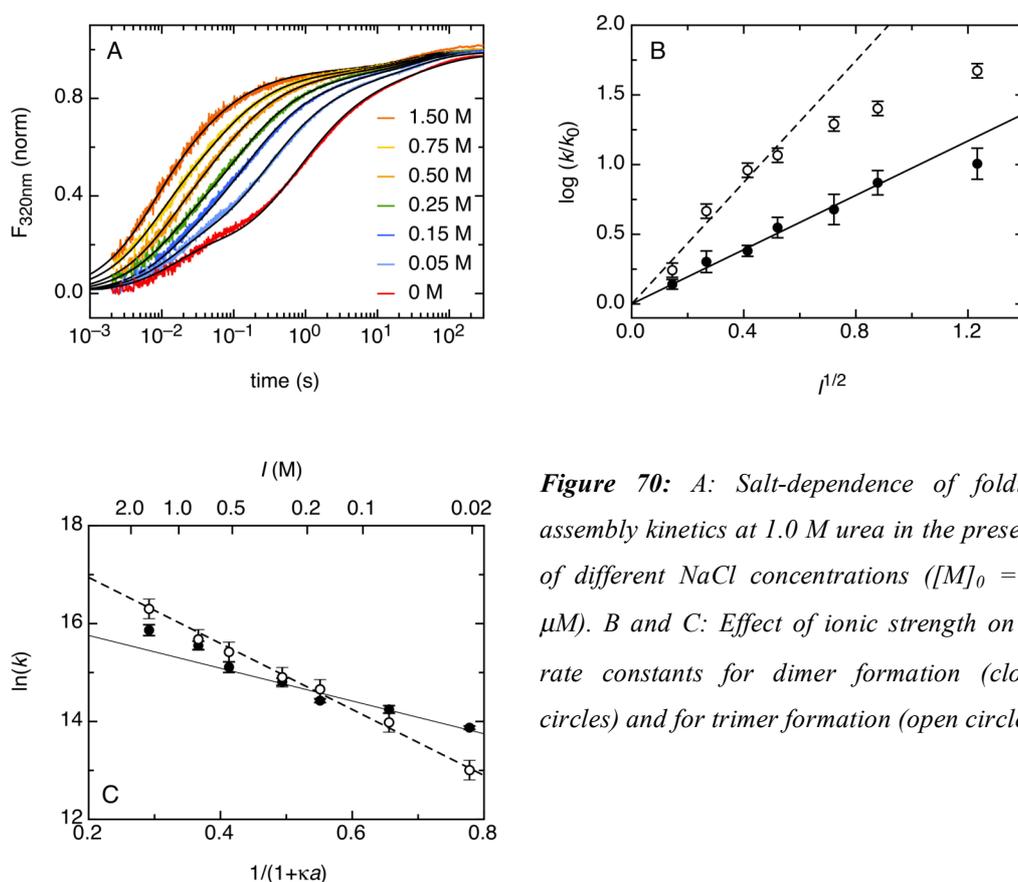
$$\ln k_d = \ln k_d^0 - \frac{U_0 / k_B T}{1 + \kappa a} \quad (45)$$

where  $U_0$  is the electrostatic interaction energy,  $a$  is the minimal distance of approach between the two molecules, which is commonly assumed to be around 6 Å for protein-protein interactions.<sup>128</sup>  $\kappa$  is the Debye-Hückel screening parameter, which is given by

$$\kappa = \left( \frac{4\pi e^2 I}{\epsilon k_B T} \right)^{1/2} \quad (46)$$

Figure 70C shows the analysis of the salt dependence of dimerization according to equation 45, which yields values of  $U_0 = 8.2(\pm 0.4)$  kJ/mol and  $k_d^0 = 1.0(\pm 0.1) \cdot 10^6 \text{ M}^{-1} \text{ s}^{-1}$ .

The rate constant for trimer formation ( $k_t$ ) can be obtained from the assembly kinetics at different ionic strength using the  $k_d$ -values determined from the initial slopes of the reactions at different ionic strength (Figure 70B) and using the observation that ionic strength has no effect on  $k_d$  at 1.0 M urea (Figure 65C). Applying these constraints the fluorescence detected assembly kinetics could be fitted with the kinetic mechanism shown in figure 68 to obtain  $k_t$ -values at different concentrations of NaCl (Figure 70A). The results confirm the previous conclusion that the trimerization step is more strongly accelerated by increasing ionic strength compared to the dimerization step (Figure 70B). Analysis of the effect of ionic strength on trimer formation according to equations 44 and 45 yields values of  $z_A z_B = 2.2(\pm 0.2)$ ,  $U_0 = 16(\pm 1)$  kJ/mol and  $k_t^0 = 4.6(\pm 0.2) \cdot 10^5 \text{ M}^{-1} \text{ s}^{-1}$  (Figures 70B and 70C). Since the net charge of a monomer was determined to be -1 this result indicates a net charge of -2 for the dimer. The ionic strength dependence of  $k_t$  is weaker at high ionic strength. A similar non-linearity of bimolecular rate constants on ionic strength is also observed for reactions between small ions.



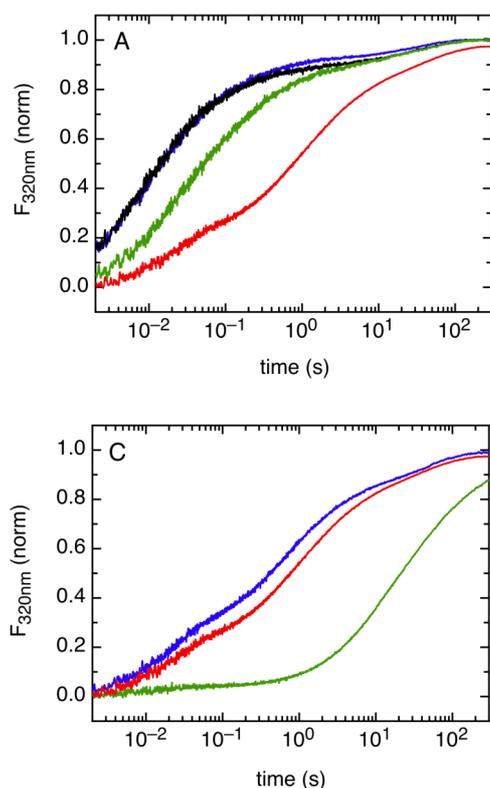
**Figure 70:** A: Salt-dependence of folding/assembly kinetics at 1.0 M urea in the presence of different NaCl concentrations ( $[M]_0 = 10 \mu\text{M}$ ). B and C: Effect of ionic strength on the rate constants for dimer formation (closed circles) and for trimer formation (open circles).

In summary, the effect of ionic strength on dimer and trimer formation indicates that the assembly process is exclusively influenced by the net charge of the interacting molecules. This result is not compatible with an electrostatic guidance of the monomer-monomer and the monomer-dimer interactions by the asymmetric charge distributions on the surface of the monomeric intermediate (Figure 10).

A possible source for the observed slower trimer formation at low ionic strength is an increased dimer stability, which could make it either difficult for a third monomer to be incorporated into a stable dimer or could decrease the concentration of free monomer. To test for the effect of ionic strength on the stability of the dimeric intermediate we compared the equilibrium constant  $K_d = [D]/[I] = k_d/k_{-d}$  for dimer formation at 1.0 M urea in the absence of NaCl and in the presence of 150 mM NaCl. The unfolding rate constant of the dimer,  $k_{-d}$ , in the presence of 150 mM NaCl was determined in interrupted refolding experiments as described above. Figure 65C shows that NaCl has no effect on  $k_{-d}$  at 1.0 M urea. As a result,  $K_d$ -values of  $2.9(\pm 0.2) \cdot 10^4 \text{ M}^{-1}$  in the absence of NaCl and  $5.3(\pm 0.2) \cdot 10^4 \text{ M}^{-1}$  in the presence of 150 mM NaCl are obtained, which shows that dimer stability is only slightly increased at high ionic strength.

The effect of increased dimer stability on the rate constant of trimer formation was tested by performing assembly kinetics in the presence of  $\text{Na}_2\text{SO}_4$ , which generally stabilizes folded and partially folded structures. Figure 71A shows that the addition of  $\text{Na}_2\text{SO}_4$  accelerates both dimer and trimer formation during assembly. In the presence of 0.5 M  $\text{Na}_2\text{SO}_4$  virtually the same kinetics are observed as in the presence of 1.5 M NaCl (Figure 71A), which gives the same ionic strength, but does not lead to a general stabilization of folded structures. These results argue against partial unfolding of the dimer as a rate-limiting step in trimer formation. This conclusion is further supported by the results from assembly kinetics in the presence of various concentrations of the stabilizing osmolyte sarcosine, which leads to an increased transient population of dimer intermediate and accelerates trimer formation (Figure 71B). Further, increasing the urea concentration to 3.0 M, which destabilizes the dimeric intermediate, leads to a deceleration of trimer formation, whereas more stabilizing conditions (0.55 M urea) accelerate assembly (Figure 71C). These results

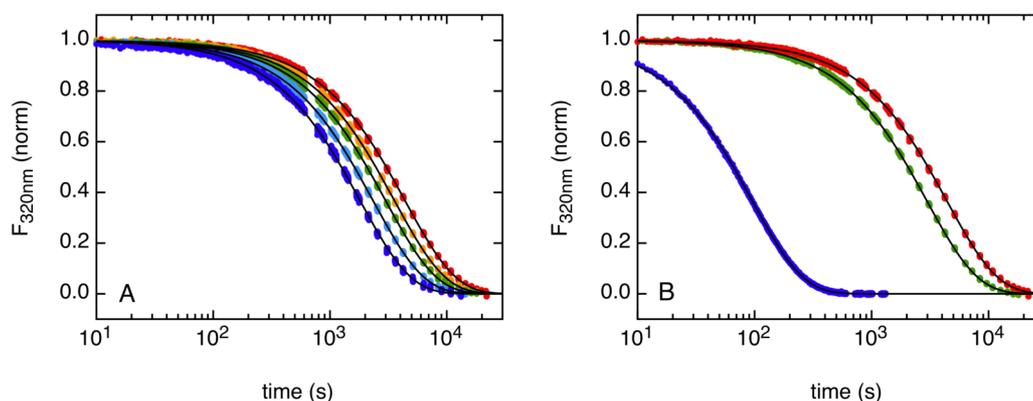
show that unfolding reactions of partially folded intermediates are not rate-limiting for assembly of the foldon domain. They further reveal that the concentration of free monomer is not limiting for trimer formation, even under conditions, where the dimer becomes increasingly populated.



**Figure 71:** Folding/assembly kinetics at 1.0 M urea (red) and at 1.0 M urea in the presence of A: 0.25 M Na<sub>2</sub>SO<sub>4</sub> (green), 0.5 M Na<sub>2</sub>SO<sub>4</sub> (blue) and 1.5 M NaCl (black), B: 0.35 M sarcosine (blue) and 0.70 M sarcosine (green). C: Folding/assembly kinetics at 0.55 M urea (green), 1.0 M urea (red) and 3.0 M urea (blue). Measurements were carried out at  $[M]_0 = 10 \mu\text{M}$  in 10 mM NaCac at  $T=20^\circ\text{C}$ .

The effect of salt on the assembly kinetic of the foldon domain showed that both dimer and trimer formation are accelerated with increasing ionic strength, indicating unfavorable electrostatic interactions in the transition states for both reactions. To test the effect of ionic strength on the dissociation and unfolding of the native trimer we monitored unfolding kinetics at 8.5 M urea in the presence of various concentrations of NaCl. Figure 72A shows that unfolding/dissociation of the trimer is accelerated with increasing concentrations of NaCl, despite stabilization of the native state by salt (Figure 63), indicating that the transition state is more strongly destabilized by electrostatic interactions than the native state. To determine the effect of ionic strength on the unfolding/disassembly process at identical protein stability we compared unfolding at 8.5 M urea with unfolding at 9.1 M urea in the presence of 150 mM NaCl and at 4.43 M GdmCl, which results in a  $\Delta G^0 = 43.1 \text{ kJ/mol}$  for all conditions. Figure 72B shows that increasing ionic strength strongly accelerates the

unfolding/disassembly process under isostability conditions, indicating unfavorable electrostatic interactions in the transition state. Figure 72 further shows that the effect of ionic strength is not yet saturated at  $I = 150$  mM. A further strong acceleration is observed between  $I = 150$  mM and  $I = 4.43$  M.



**Figure 72:** A: Unfolding of the native foldon trimer at 8.5 M urea in the absence of salt (red) and in the presence of different NaCl concentrations: 0.15 M (yellow), 0.30 M (green), 0.75 M (light blue) and 1.50 M (dark blue). B: Unfolding of the native protein at isostability conditions in the presence of 8.5 M urea (red), at 9.1 M urea in the presence of 0.15 M NaCl (green) and in the presence of 4.43 M GdmCl (blue).

Our results show that the transition state for trimer formation and unfolding/dissociation of the trimeric foldon domain is destabilized by unfavorable electrostatic interactions. Also the native state is slightly destabilized by electrostatic interactions, as indicated by the increased stability in the presence of 150 mM NaCl. However, the transition state is more strongly destabilized by electrostatics than the native state, since unfolding becomes faster at high ionic strength, despite the stabilization of the native trimer. The observations are surprising since salt bridges between Glu5 and Arg15 of adjacent subunits were shown to be important for stability of the native trimer and asymmetric charge distribution on the surface of the monomeric intermediate suggested an important role of electrostatic interactions in the assembly process. Our results show that recognition between the subunits does not involve electrostatic interactions and argue against an important role of the asymmetric charge distribution and of the Glu5-Arg15 salt bridge during neither dimer formation nor trimer formation. The deceleration of the two association steps by electrostatic interactions is in agreement with Debye-Smoluchowski theory, which

takes only the net charges of the interacting molecules into account and predicts a bimolecular rate constant of

$$k_{DS} = \frac{N_L}{1000} \cdot \frac{4\pi r_c (D_A + D_B)}{\exp\left(\frac{r_c}{a}\right) - 1} \quad \text{with } r_c = \frac{z_A z_B e^2}{4\pi\epsilon_0 \epsilon k_B T} \quad (47)$$

The net charge of a foldon monomer is -1 and of a foldon dimer -2. Thus, the Debye-Smoluchowski theory predicts a stronger deceleration of trimer formation compared to dimer formation in agreement with our experimental results.

These results raise the question of the importance of the native topology of the monomeric intermediate for the assembly process. In addition to the asymmetric charge distribution the formation of a native-like topology in the monomer also creates a hydrophobic cluster involving residues Pro4, Ala6, Pro7, Tyr13, Trp20 and Leu22.<sup>112</sup> Our results indicate that hydrophobic interactions between these clusters may be important for recognition between the subunits during dimer and trimer formation. This model is in agreement with results on the recognition between S-peptide and S-protein during formation of ribonuclease S, which is also solely mediated by hydrophobic interactions.<sup>188</sup>

Unfavorable electrostatic interactions were also shown to slow down dimerization of the homodimeric Arc repressor leading to an increase in the rate constant for association with increasing ionic strength. In this system dimerization is drastically accelerated when the charged residues at the interaction site are replaced by hydrophobic side chains.<sup>103</sup>

Refolding conditions	$k_d$ ( $M^{-1}s^{-1}$ )	$k_{-d}$ ( $s^{-1}$ )	$k_t$ ( $M^{-1}s^{-1}$ )	$k_{-t}$ ( $s^{-1}$ )
1.0 M urea + 0.25 M Na <sub>2</sub> SO <sub>4</sub>	$2.6(\pm 0.4) \cdot 10^6$ <sup>A</sup>	-	-	-
1.0 M urea + 0.50 M Na <sub>2</sub> SO <sub>4</sub>	$8.1(\pm 1.2) \cdot 10^6$ <sup>A</sup>	-	-	-
1.0 M urea + 0.35 M sarcosine	$1.1(\pm 0.1) \cdot 10^6$ <sup>A</sup>	-	-	-
1.0 M urea + 0.70 M sarcosine	$1.2(\pm 0.2) \cdot 10^6$ <sup>A</sup>	-	-	-
1.0 M urea + 35% (w/v) Glycerol (2.5cP)	$7.4(\pm 0.8) \cdot 10^5$ <sup>A</sup>	-	-	-
0.55 M urea	$1.4(\pm 0.2) \cdot 10^6$ <sup>A</sup>	$32(\pm 2)$ <sup>B</sup>	-	$1.6(\pm 0.4) \cdot 10^{-5}$ <sup>B</sup>
1.0 M urea	$1.1(\pm 0.1) \cdot 10^6$ <sup>C</sup>	$36(\pm 1)$ <sup>B</sup>	$4.5(\pm 0.5) \cdot 10^5$ <sup>C</sup>	$2.0(\pm 0.4) \cdot 10^{-5}$ <sup>B</sup>
3.0 M urea	$2.2(\pm 0.3) \cdot 10^5$ <sup>A</sup>	$63(\pm 2)$ <sup>B</sup>	$8.2(\pm 1.4) \cdot 10^4$ <sup>C</sup>	$3.5(\pm 0.7) \cdot 10^{-5}$ <sup>B</sup>

**Table 12:** Rate constants for folding/assembly of wild type foldon. *A*: determined from initial slope, *B*: experimentally measured, *C*: determined from a fit to the mechanism shown in figure 68. A fit of equation 29 to  $k_d$  in 1.0 M urea and in 1.0 M urea + 35% (w/v) Glycerol reveals  $\beta = 0.42$ .

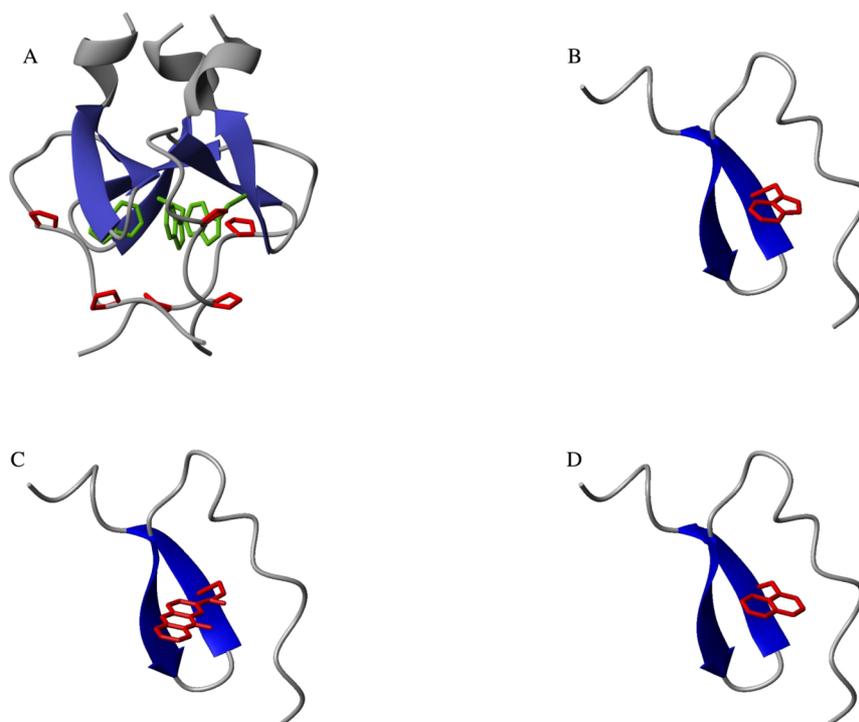
### 3.2.3 The speed-limit for site-specific formation of an encounter complex during assembly of the foldon domain

Previous results from kinetic folding studies showed that the isolated 27 amino acid foldon domain from bacteriophage T4 fibrinogen spontaneously folds into a very stable trimer in a series of very fast folding and association steps (Figure 62).<sup>111</sup> At low ionic strength formation of the dimeric intermediate (D) and trimeric intermediate (T) occur with rate constants of  $k_d = 1.1(\pm 0.1) \cdot 10^6 \text{ M}^{-1}\text{s}^{-1}$  and  $k_t = 4.5(\pm 0.9) \cdot 10^5 \text{ M}^{-1}\text{s}^{-1}$ , respectively (see chapter 3.2.2). Dimer formation becomes about two times faster at physiological ionic strength while trimer formation increases about 5-fold at physiological ionic strength (see chapter 3.2.2). The rate constants of the association reactions are in the range of the expected diffusion limit for site-specific protein-protein interactions. This comparison raises the question of diffusion control of the bimolecular steps during folding and assembly of the foldon domain.

Site-specific protein-protein interactions are initiated by the formation of an encounter complex, which is typically followed by internal structural rearrangements. To study the process of site-specific encounter complex formation for dimerization during assembly of the foldon domain we apply TTET to the foldon E5R variant. The E5R variant of the foldon domain was shown to be a good model system for the early folding intermediate,<sup>112</sup> which provides basis for the two association steps during assembly. In the E5R variant an intermolecular salt-bridge between Glu5 and Arg15 is replaced by a repulsive interaction, which results in a monomeric protein that structurally resembles the monomeric assembly intermediate. The E5R variant is thus suited to apply TTET for measuring diffusion-controlled formation of transient protein-protein interactions between folded monomeric subunits, since the subsequent formation of a stable dimer does not occur.

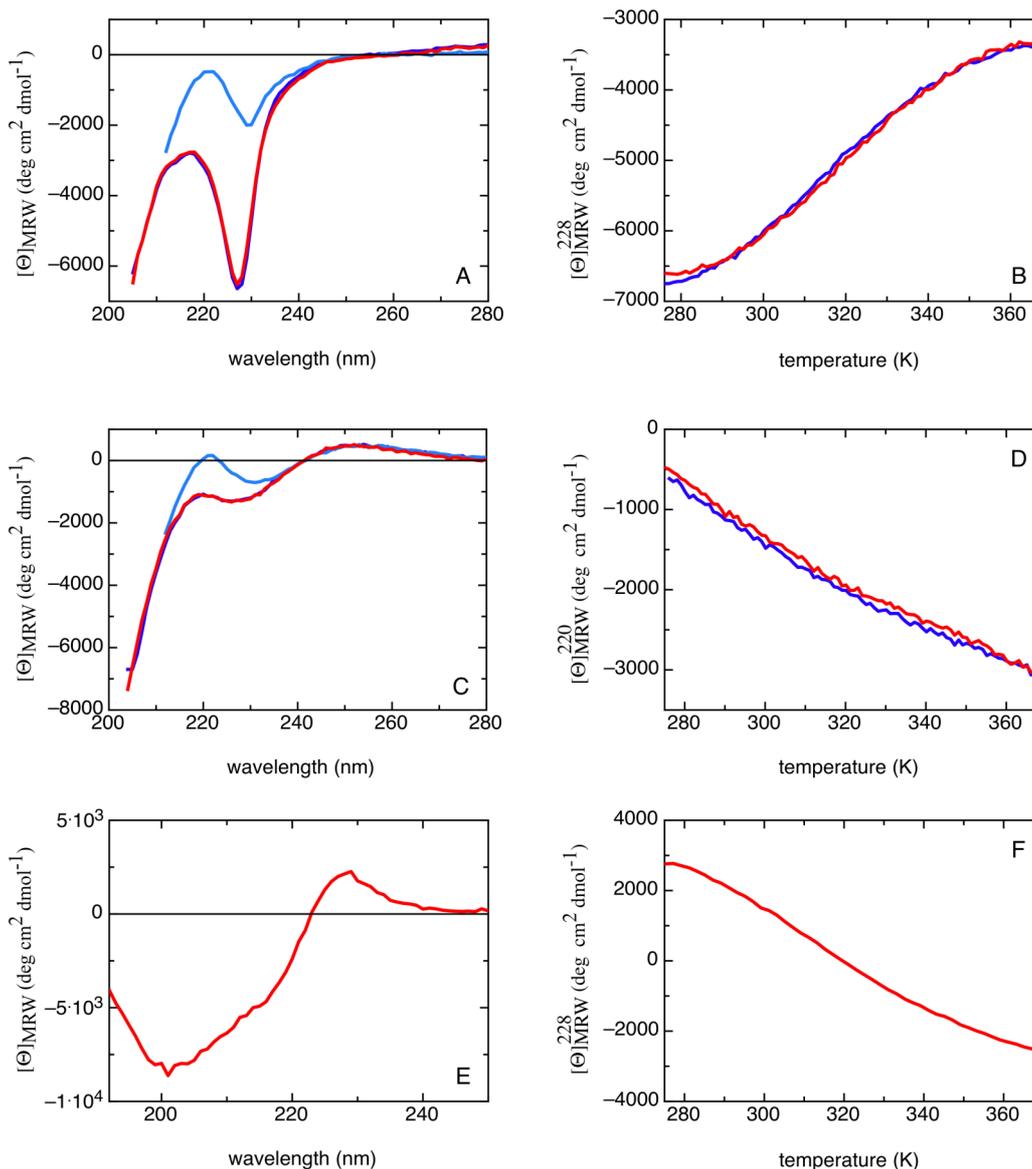
To study site-specific contact formation between two foldon subunits we replaced tryptophane at position 20 by either the triplet donor xanthone (Trp20Xan) or the acceptor naphthalene (Trp20Nal) in the foldon E5R variant (Figure 73). This position is suitable for introducing the TTET labels since (i) in the native trimer close-packed interactions between Trp20 of each subunit are formed (Figure 73A), (ii) in the dimeric folding intermediate intermolecular interactions involving Trp20 are already

formed as judged by changes in fluorescence properties of Trp20 during dimer formation and (iii) the side chain of Trp is chemically similar to the aromatic triplet labels.



**Figure 73:** *A: Structure of the native foldon trimer taken from reference 111. Trp residues at position 20 of each chain are shown in green and prolyl residues at positions 4 and 7 are shown in red. B: Structure of the foldon E5R variant taken from reference 112. Trp at position 20 is shown in red. C and D: E5R TTET variants where either the triplet donor (C) or the triplet acceptor (D) was introduced into the structure of the E5R variant at position of Trp20.*

To test whether introducing the TTET labels at the position of Trp20 changes the structural properties of the E5R variant we performed far UV CD and solution NMR spectroscopy experiments. Comparison of the far UV CD spectra of labelled monomer and unlabelled monomer with Trp at position 20 is difficult due to significant contributions from the TTET labels to the CD signal (Figure 74A, 74C and 74E). However, CD-detected melting curves indicate that the stability of the labelled monomer is similar to the stability of the unlabelled monomer (Figure 74B, 74D and 74F).



**Figure 74:** Far UV CD spectra and temperature-induced equilibrium transitions of Trp20Nal (A and B), Trp20Xan (C and D) and the unlabelled E5R variant (E and F). Measurements in A-D were carried out in 10 mM KP pH 7.0 8.0 M GdmCl (light blue), in 10 mM KP pH 7.0 0.58 M GdmCl (dark blue) and in 10 mM NaCac pH 7.0 1.0 M urea (red).  $T=20^{\circ}\text{C}$ . Measurements on the unlabelled E5R variant were performed in the absence of denaturant. The spectrum and the equilibrium transition of the unlabelled E5R variant are taken from reference 112.

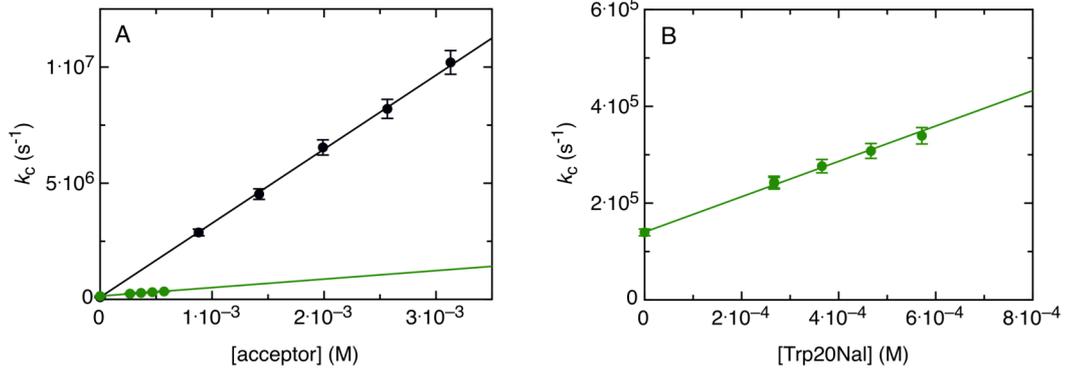
Comparison of the 2-D NOESY spectra for the Trp20Nal variant and the unlabelled E5R variant showed that the  $\beta$ -hairpin is still formed in the E5R TTET variant. In this region similar NOEs are present in the labelled monomer and in the unlabelled monomer. Further, Nal20 in the labelled monomer has similar contacts as Trp20 in the unlabelled monomer. In the unlabelled foldon E5R variant the N-terminal region

participates in a hydrophobic cluster with residues from the hairpin and from the C-terminal region.<sup>112</sup> Corresponding NOEs were not observed in the Trp20Nal variant, indicating that the hydrophobic cluster is not formed in the labelled monomer. In the unlabelled foldon E5R variant the presence of a defined structure in the C-terminal region was found that is similar to the  $3_{10}$ -helical turn in the native wild-type trimer.<sup>112</sup> Results on the Trp20Nal variant indicate that this structure is also present in the labelled monomer. However, for the unlabeled monomer more NOEs are found in the C-terminal region compared to the labelled monomer, which can partly be explained by spectral overlap of resonances in this region. However, this observation also indicates that the structure in the C-terminal region is more loosely formed in the Trp20Nal variant compared to the unlabelled E5R variant.

Intermolecular TTET experiments were performed under pseudo first-order conditions with donor concentrations between 20 and 50  $\mu\text{M}$  and a minimally 10-fold excess of the acceptor molecule. The respective bimolecular rate constants  $k_A$  for TTET can then be obtained by analyzing the data according to the equation

$$k_c = k_0 + k_A \cdot [\textit{acceptor}] \quad (38)$$

The experiments were carried out in the presence of 1.0 M urea to enable comparison with results on the kinetics of the folding/assembly reaction in the wild-type protein (see chapter 3.2.2). Figure 75 shows that the foldon monomers can form this site-specific contact at position 20 with a rate constant of  $k_A = 3.7(\pm 0.2) \cdot 10^8 \text{ M}^{-1}\text{s}^{-1}$ . Encounter complex formation is diffusion-controlled as indicated by the 2-fold decrease in  $k_A$  if the viscosity is increased by a factor of 2 (Figure 78). As a reference, bimolecular TTET between the free labels xanthone and 1-naphthylacetic acid under the same conditions yields  $k_A = 3.2(\pm 0.2) \cdot 10^9 \text{ M}^{-1}\text{s}^{-1}$ , in a process that is also diffusion-controlled.<sup>47,140,209</sup>



**Figure 75:** Intermolecular TTET under pseudo-first order conditions between Trp20Xan and Trp20NaI (green,  $k_A = 3.7 \cdot 10^8 \text{ M}^{-1} \text{ s}^{-1}$ ) and between the free labels xanthone and 1-naphthylacetic acid (black,  $k_A = 3.2 \cdot 10^9 \text{ M}^{-1} \text{ s}^{-1}$ ) at 10 mM NaCac pH 7.0 1.0 M urea at  $T = 20^\circ \text{C}$ .

The Smoluchowski theory allows to calculate the diffusion-controlled rate constant for formation of a random contact between two foldon monomers according to<sup>88,210,211</sup>

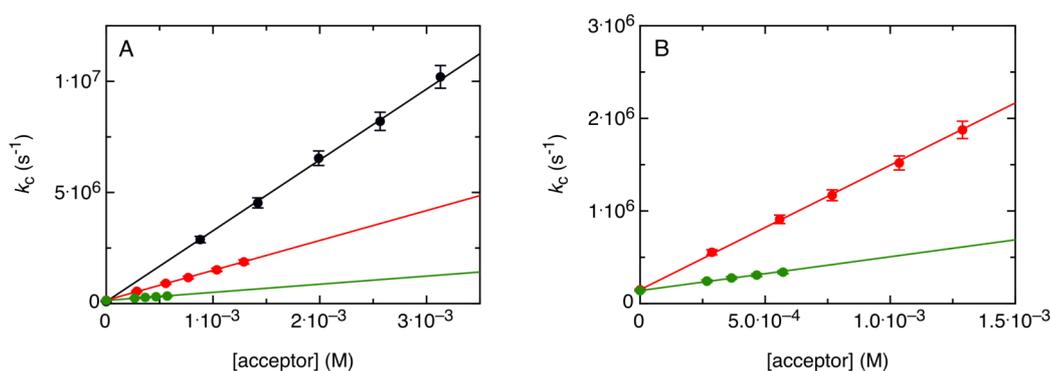
$$k_S = \frac{N_L}{1000} 4\pi a (D_A + D_B) \quad (17)$$

Determination of the diffusion coefficients ( $D_{A/B}$ ) for the labelled monomeric assembly intermediate by dynamic light scattering (DLS) yielded values of  $1.47 \cdot 10^{-6} \text{ cm}^2 \text{ s}^{-1}$  for both variants in the presence of 1.0 M urea at  $20^\circ \text{C}$  (see chapter 3.2.4). The contact distance for TTET between donor and acceptor,  $a$ , was calculated from the NMR-structure of the monomeric intermediate (see chapter 3.2.4).<sup>112</sup> Assuming a spherical molecule yields an effective radius of the monomeric intermediate of  $r = 12.0 \text{ \AA}$  resulting in  $a = 2r = 24.0 \text{ \AA}$ . These values give a maximum rate constant for random contact between the subunits of  $k_S = 5.3 \cdot 10^9 \text{ M}^{-1} \text{ s}^{-1}$ , which is 14-times faster than the observed value for site-specific interaction at position 20. A slightly lower value of  $4.6 \cdot 10^9 \text{ M}^{-1} \text{ s}^{-1}$  is obtained by applying the Debye-Smoluchowski equation

$$k_{DS} = \frac{N_L}{1000} \cdot \frac{4\pi r_c (D_A + D_B)}{\exp\left(\frac{r_c}{a}\right) - 1} \quad \text{with } r_c = \frac{z_A z_B e^2}{4\pi \epsilon_0 \epsilon k_B T} \quad (47)$$

which takes into account the effect of the net charge of +1 of a foldon subunit.<sup>90</sup> This comparison shows that about 8 % of random contacts between the subunits lead to a

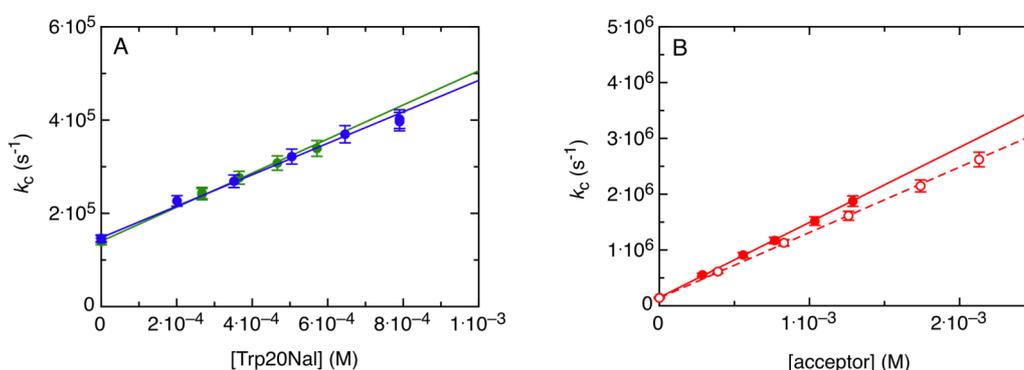
site-specific interaction at position 20, which indicates that the large majority of subunit contacts occur in orientation that do not allow interactions at position 20. This conclusion is confirmed by TTET experiments between a triplet donor foldon subunit (Trp20Xan) and the small triplet acceptor 1-naphthylalanineamide (Nal<sub>amid</sub>), which carries the same net charge of +1 as a foldon subunit but should not have major orientational constraints for TTET. Compared to the site-specific interactions of two foldon subunits the small molecule acceptor shows a nearly 4-fold acceleration of site-specific encounter complex formation ( $k_A = 1.4(\pm 0.1) \cdot 10^9 \text{ M}^{-1}\text{s}^{-1}$ , Figure 76A), which represents the speed-limit for interaction of a small molecule with position 20 on a foldon subunit. This value is significantly lower than the predicted value of  $k_S = 8.8 \cdot 10^9 \text{ M}^{-1}\text{s}^{-1}$  or  $k_{DS} = 6.7 \cdot 10^9 \text{ M}^{-1}\text{s}^{-1}$  using equations 17 and 47, respectively. This result shows that about 20 % of all randomly formed encounter complexes lead to an interaction of the small molecule Nal<sub>amid</sub> with position 20 on a foldon monomer.



**Figure 76:** Intermolecular TTET under pseudo-first order conditions between Trp20Xan and Trp20Nal (green,  $k_A = 3.7 \cdot 10^8 \text{ M}^{-1}\text{s}^{-1}$ ), between the free labels xanthone and 1-naphthylacetic acid (black,  $k_A = 3.2 \cdot 10^9 \text{ M}^{-1}\text{s}^{-1}$ ) and between Trp20Xan and Nal<sub>amid</sub> (red,  $k_A = 1.4 \cdot 10^9 \text{ M}^{-1}\text{s}^{-1}$ ). Measurements were carried out in 10mM NaCac pH 7.0 1.0 M urea at  $T = 20^\circ\text{C}$ .

The maximum rate constant determined for encounter complex formation between the subunits ( $k_A = 3.7 \cdot 10^8 \text{ M}^{-1}\text{s}^{-1}$ ) is significantly faster than formation of the dimeric intermediate on the assembly pathway of the wild-type protein under the same experimental conditions ( $k_A = 1.1 \cdot 10^6 \text{ M}^{-1}\text{s}^{-1}$ ). This comparison reveals that only one out of 300 encounter complexes that form at this native interaction site proceeds to yield a stable dimer. Since only 8 % of all random contacts between the subunits lead to a site-specific encounter complex, only one out of 4000 random complexes leads to

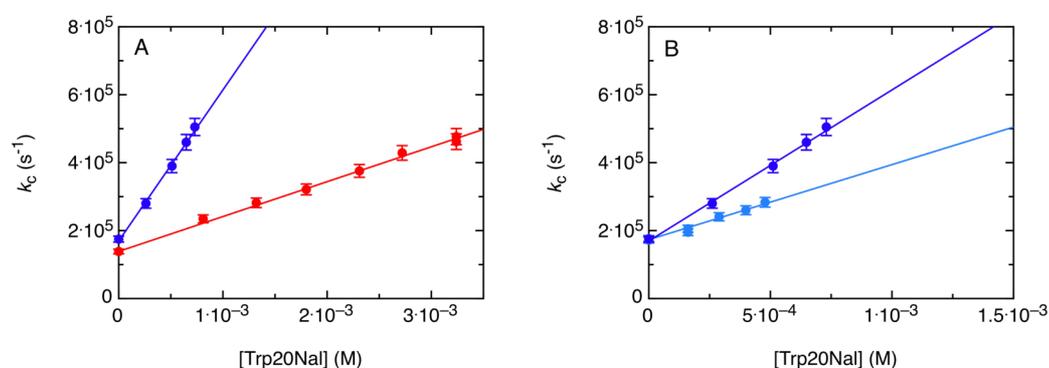
formation of a stable dimer, indicating significant enthalpic and/or entropic barriers for dimerization. Our results further show that site-specific encounter complex formation between two foldon subunits occurs faster than expected in the absence of attractive interactions (the basal rate constant for site-specific protein-protein interactions is set to around  $10^5$  to  $10^6$   $M^{-1} s^{-1}$ ).<sup>128</sup> To test for the contributions of electrostatic interactions to encounter complex formation we performed TTET between donor and acceptor subunits in the presence of 0.58 M GdmCl. The increased ionic strength has virtually no effect on encounter complex formation ( $k_A = 3.4(\pm 0.2) \cdot 10^8$   $M^{-1} s^{-1}$ ), which demonstrates that electrostatic effects do not contribute to the fast encounter complex formation (Figure 77A). This result is in agreement with experimentally observed minor effect of ionic strength effects on the rate constant for dimer formation in the assembly process of the wild-type protein (see chapter 3.2.2). In agreement with this result decreasing the net charge of the small molecule from  $z = +1$  in  $Nal_{amid}$  to  $z = 0$  in  $Nal$  has no significant effect on the rate constant for diffusion-controlled interaction with  $Trp20Xan$  (Figure 77B).



**Figure 77:** A: Intermolecular TTET under pseudo-first order conditions between  $Trp20Xan$  and  $Trp20Nal$  in 10 mM  $NaCac$  pH 7.0 1.0 M urea (green,  $k_A = 3.7(\pm 0.2) \cdot 10^8$   $M^{-1} s^{-1}$ ) and in 10 mM  $KP$  pH 7.0 0.58 M  $GdmCl$  (blue,  $k_A = 3.4(\pm 0.2) \cdot 10^8$   $M^{-1} s^{-1}$ ). B: Intermolecular TTET under pseudo-first order conditions between  $Trp20Xan$  and  $Nal_{amid}$  (red,  $k_A = 1.4(\pm 0.1) \cdot 10^9$   $M^{-1} s^{-1}$ ) and between  $Trp20Xan$  and  $Nal$  (red open,  $k_A = 1.2(\pm 0.1) \cdot 10^9$   $M^{-1} s^{-1}$ ) in 10 mM  $NaCac$  pH 7.0 1.0 M urea. Measurements were carried out at  $T = 20^\circ C$ .

Small single domain proteins with less than 100 amino acids typically do not populate intermediate states so significant amounts during folding.<sup>6</sup> In this respect, the rapid formation of a monomeric intermediate by a 27 amino acid foldon subunit is

surprising. It was proposed that the native-topology of the intermediate facilitates assembly by providing a template for site-specific interaction between the subunits.<sup>112</sup> We tested the importance of native topology for encounter complex formation by measuring intersubunit TTET in the presence of 8.0 M GdmCl, which completely unfolds the foldon domain (Figure 74A and 74C). Figure 78A shows that intermolecular TTET between positions 20 of unfolded monomers in 8.0 M GdmCl is only decreased about 2-fold compared to TTET between intermediates in the presence of 0.58 M GdmCl. This result shows that the native topology in the intermediate only slightly accelerates site-specific subunit-subunit recognition but it has a strong effect on the probability of an encounter complex to proceed and yield a stable dimer.



**Figure 78:** A: Intermolecular TTET under pseudo-first order conditions between Trp20Xan and Trp20Nal at  $T=25^{\circ}\text{C}$  A: in 10 mM KP pH 7.0 in the presence of 0.58 M GdmCl (blue,  $k_A=4.4(\pm 0.2)\cdot 10^8 \text{ M}^1\text{s}^{-1}$ ,  $\eta=0.91 \text{ cP}$ ) and in the presence of 8.0 M GdmCl (red,  $k_A=1.0(\pm 0.1)\cdot 10^8 \text{ M}^1\text{s}^{-1}$ ). For comparison,  $k_A$  in the presence of 8.0 M GdmCl ( $\eta=2.16 \text{ cP}$ ) was corrected for viscosity effects, which yields  $k_A=2.4(\pm 0.2)\cdot 10^8 \text{ M}^1\text{s}^{-1}$  at  $\eta=0.91 \text{ cP}$ . B: in 10 mM KP pH 7.0 in the presence of 0.58 M GdmCl (blue,  $k_A = 4.4(\pm 0.2)\cdot 10^8 \text{ M}^1\text{s}^{-1}$ ) and in the presence of 0.58 M GdmCl, 24 % (w/v) glycerol (light blue,  $k_A=2.2(\pm 0.1)\cdot 10^8 \text{ M}^1\text{s}^{-1}$ ), which corresponds to a 1.9-fold viscosity increase from  $\eta=0.91 \text{ cP}$  to  $\eta=1.72 \text{ cP}$ .

### 3.2.4 Analytical models for calculation of the reactive surface on a monomeric foldon subunit

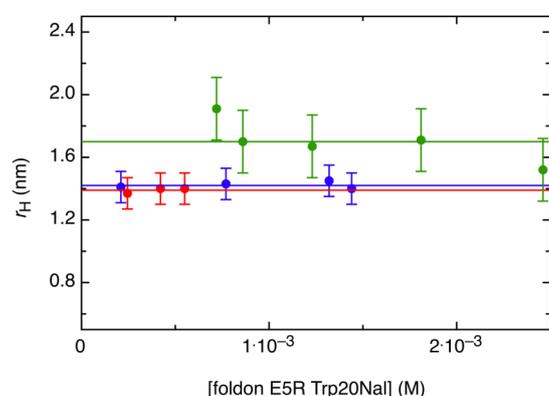
The diffusion-controlled rate constant  $k_A$  for site-specific interaction between two monomeric foldon subunits was determined in triplet-triplet energy transfer (TTET) experiments to be  $3.7 \cdot 10^8 \text{ M}^{-1}\text{s}^{-1}$  (see chapter 3.2.3). Comparing this value to the Smoluchowski rate constant  $k_S$  (Equation 17), which assumes uniformly reactive particles, indicates that the monomeric foldon subunit is not reactive over its entire surface. This is supported by an increase in  $k_A$  to  $1.4 \cdot 10^9 \text{ M}^{-1}\text{s}^{-1}$  when one foldon subunit is replaced for uniformly reactive 1-naphthylalanineamide (see chapter 3.2.3). Applying analytical models for site-specific bimolecular reactions allows quantification of the partial surface reactivity of a monomeric foldon subunit from the comparison of  $k_A$  with  $k_S$ .

Analytical models for bimolecular reactions require knowledge of the reaction distance  $a$ . The association reaction is defined to be successful as soon as the radial separation between the reactants reaches the reaction distance. Since TTET measures formation of van-der-Waals contacts, we determined the reaction distance to be the sum of the van-der-Waals radii of the interacting molecules. For the free triplet labels this is 0.4 nm. Assuming a spherical molecule, the effective radius  $r$  of a foldon monomer was determined from the radius of gyration ( $r_G$ ) according to

$$r_G = \sqrt{\frac{5}{3}} r \quad (48)$$

The radius of gyration for a foldon monomer was calculated from the NMR-structure of the monomeric E5R variant to be 9.3 nm resulting in an effective radius of  $r = 1.2$  nm and a reaction distance for interaction between two foldon monomers of  $a = 2.4$  nm. For the foldon TTET variants, this value is probably somewhat underestimated since the introduction of the triplet labels at the position of tryptophane 20, especially the introduction of the relatively large xanthone (Xan), might lead to an increased radius due to less tight packing of the N-terminus to the  $\beta$ -hairpin. The analytical models further require knowledge of the translational and rotational diffusion coefficients of the reactants. Translational diffusion coefficients  $D_X$  have been

determined experimentally via dynamic light scattering (DLS) and were used to calculate the hydrodynamic radius  $r_{X,H}$  for molecule X via the Stokes-Einstein equation (Equation 18 and figure 79). Values are summarized in table 13.



**Figure 79:** Hydrodynamic radius  $r_H$  of the foldon triplet acceptor variant (Trp20Nal) as a function of protein concentration in 10 mM KP pH 7.0 in the presence of 0.58 M GdmCl (red), 1.1 M GdmCl (blue) and 8 M GdmCl (green) at  $T = 25^\circ\text{C}$ . The apparent  $r_H$  is independent of protein concentration. Therefore, the effective  $r_H$  was calculated from the average of the individual measurements (represented by the continuous lines).

Molecule	[GdmCl] (M)	$D$ ( $\text{cm}^2/\text{s}$ )	$D_{20,w}$ ( $\text{cm}^2/\text{s}$ )	$\eta$ (cP)	$r_H$ (nm)
Trp20Nal	0.58 M	$1.73 \cdot 10^{-6}$	$1.57 \cdot 10^{-6}$	0.91	$1.39(\pm 0.1)$
Trp20Nal	1.1 M	$1.67 \cdot 10^{-6}$	$1.54 \cdot 10^{-6}$	0.92	$1.42(\pm 0.1)$
Trp20Nal	8.0 M	$5.94 \cdot 10^{-7}$	$1.28 \cdot 10^{-6}$	2.16	$1.70(\pm 0.1)$
2-NAA	0 M	$8.18 \cdot 10^{-6}$	$7.28 \cdot 10^{-6}$	0.89	$0.30(\pm 0.05)$

**Table 13:** Diffusion coefficients  $D$  of Trp20Nal and 2-naphthylacetic acid (2-NAA) measured by DLS in 10 mM KP pH 7 at  $25^\circ\text{C}$ . Viscosities of all GdmCl buffers were measured by falling ball viscosimetry.  $D_{20,w}$  is the diffusion constant normalized to the viscosity of water at  $20^\circ\text{C}$ .

Since DLS measurements on the foldon triplet donor variant (Trp20Xan) were not possible due to relatively small solubility, we assumed identical diffusion coefficients as for the foldon triplet acceptor variant (Trp20Nal). Similarly, the diffusion coefficient of 1-naphthylalanineamide (Nal<sub>amid</sub>) and 1-NAA were assumed from the measurements on 2-NAA. Rotational diffusion coefficients of the reactants have not been determined experimentally. Assuming a spherical shape with hydrodynamic no-slip boundary conditions on the surface we determined them according to

$$D_X^{rot} = \frac{k_B T}{8\pi\eta r_{X,H}^3} = \frac{3D_X}{4r_{X,H}^2} \quad \text{with X = A, B} \quad (49)$$

resulting in

$$\frac{D_A^{rot}}{D_A + D_B} = \frac{3r_{B,H}}{4r_{A,H}^2(r_{A,H} + r_{B,H})} \quad (50)$$

Taking the Smoluchowski rate constant  $k_S$  as a reference, the reduction in rate constant  $\Gamma$  due to geometrical constraints can be described according to

$$k_a = \Gamma k_S \quad (51)$$

The Smoluchowski theory assumes spherical shape of the interacting molecules and neglects any specific intermolecular interactions. The corresponding values are summarized in table 14.

Type of reaction	Conditions	$k_A$ ( $M^{-1}s^{-1}$ )	$k_S$ ( $M^{-1}s^{-1}$ )	$k_S/k_A$	$\Gamma$
Trp20Xan / Trp20Nal	0.58 M GdmCl	$3.4 \cdot 10^8$	$5.5 \cdot 10^9$	16.2	0.06
Trp20Xan / Trp20Nal	1 M urea	$3.7 \cdot 10^8$	$5.3 \cdot 10^9$	14.3	0.07
Trp20Xan / Nal <sub>amid</sub>	1 M urea	$1.4 \cdot 10^9$	$8.8 \cdot 10^9$	6.3	0.16
Xan / 1-NAA	1 M urea	$3.2 \cdot 10^9$	$4.1 \cdot 10^9$	1.3	0.78

**Table 14:** Summary of measured bimolecular rate constants ( $k_A$  in 10 mM KP) and Smoluchowski rate constants ( $k_S$ ) calculated via equation 17.

Using the constant flux approximation, Schlosshauer et al. presented a general expression for the bimolecular rate constant for association of spherical particles bearing reactive patches on the surface considering two limiting angles for each particle.<sup>127</sup> Within the following calculations we assume that reactive patches on the surface of the interacting particles are spherically symmetric. Further, each correctly oriented particle pair should react instantaneously ( $\kappa \rightarrow \infty$ ). The corresponding rate reduction factor then simplifies considerably and reduces to the expression found by Zhou,<sup>113</sup> who also considered the effect of centrosymmetric interaction potentials.

Using spherical coordinates we assume that the surface of particle X is reactive for  $0 \leq \theta < \theta_X^0$  and inert for  $\theta_X^0 \leq \theta \leq \pi$ . The fraction of reactive surface then is

$$F_X = \frac{1 - \cos \theta_X^0}{2} \quad (52)$$

The diffusion-controlled rate constant  $k_A$  for association between molecule A and molecule B is reduced by the factor

$$\Gamma = -(1 - \cos \theta_A^0)^2 (1 - \cos \theta_B^0)^2 \left( \sum_{l, l_1, l_2} S(l, l_1, l_2) \right)^{-1} \quad (53)$$

compared to the Smoluchowski rate constant  $k_S$  due to partially reactive surfaces of both molecules.<sup>127</sup> The sum over the indices  $l, l_1, l_2$  runs from 0 to  $\infty$  and the summands are defined as

$$S(l, l_1, l_2) \equiv \frac{(2l+1)K_{l+1/2}(\xi^*)}{lK_{l+1/2}(\xi^*) - \xi^* K_{l+3/2}(\xi^*)} (2l_1+1)(2l_2+1) I_{l_1, A}^2 I_{l_2, B}^2 \begin{pmatrix} l & l_1 & l_2 \\ 0 & 0 & 0 \end{pmatrix}^2 \quad (25)$$

$$\text{with } \xi^* \equiv \sqrt{\frac{D_A^{rot} l_1(l_1+1) + D_B^{rot} l_2(l_2+1)}{D_A + D_B}} \cdot a \quad (26)$$

$$\text{and } I_{l, X} \equiv \int_0^{\theta_X^0} d\theta \sin \theta P_l(\cos \theta) = \frac{P_{l-1}(\cos \theta_X^0) - P_{l+1}(\cos \theta_X^0)}{2l+1} \quad (27)$$

with X = A, B

The matrix  $\begin{pmatrix} l & l_1 & l_2 \\ 0 & 0 & 0 \end{pmatrix}$  denotes the Wigner 3- $j$  symbol.  $K_l$  is a modified Bessel function of the second kind and  $P_l(x)$  is the  $l$ th order Legendre polynomial. The reduction in rate constant is less than expected from a naïve geometric argument, i.e.

$\Gamma \geq \Gamma^{\text{geom}} \equiv F_A \cdot F_B$ . The reason is that molecules, which do not react at their initial encounter, have a certain probability to react after mutual reorientation and therefore can still contribute to the overall reaction rate constant  $k_A$ .

In the case where one of the interacting molecules is uniformly reactive over its surface, i.e.  $\theta_B^0 = \pi$ , we arrive at the model introduced by Solc and Stockmayer,<sup>122</sup> for which within the constant flux approximation the rate reduction factor was previously shown to be<sup>126</sup>

$$\Gamma^{\text{uni}} = -(1 - \cos \theta_A^0)^2 \left( \sum_{l=0}^{\infty} \frac{K_{l+1/2}(\xi^{*'})}{lK_{l+1/2}(\xi^{*'}) - \xi^{*'} K_{l+3/2}(\xi^{*'})} \cdot \frac{(P_{l-1}(\cos \theta_A^0) - P_{l+1}(\cos \theta_A^0))^2}{2l+1} \right)^{-1} \quad (54)$$

with

$$\xi^{*'} \equiv \sqrt{\frac{D_A^{\text{rot}} l(l+1)}{D_A + D_B}} \cdot a \quad (22)$$

It is important to consider that the following effects are not taken into account by the applied models for bimolecular reactions: i) Intermolecular hydrodynamic interactions, which correlate the motion of the reactants at small distances. ii) Any electrostatic interactions between the molecules. This assumption seems to be appropriate since the experimentally determined bimolecular rate constants do not depend on ionic strength. iii) Deviations of a perfectly spherical shape of the reactants.

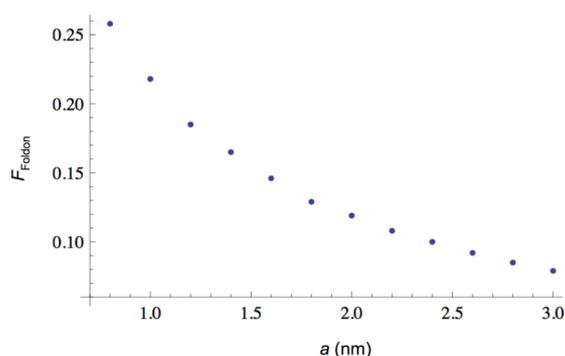
The obtained values for surface reactivity in a foldon subunit and  $\text{Nal}_{\text{amid}}$  as well as for the free labels are summarized in table 15. These calculations were performed in collaboration with Y. von Hansen (TU München). For association of 1-NAA with xanthone we assume both reactants to possess the same surface reactivity. The reactive fraction of the surface is chosen such that equation 53 matches the experimentally observed reduction in rate constant. Results reveal  $F_{1\text{-NAA}} = F_{\text{Xan}} \approx 0.7$ . Here, the deviation of  $F$  from unity might be rather due to hydrodynamic effects than from reduced surface reactivity. Using equation 53 and  $F_{\text{Nal(amid)}} = 0.7$  for the foldon- $\text{Nal}_{\text{amid}}$  reaction we find  $F_{\text{foldon}} \approx 0.05$ . Assuming a uniform reactivity of the

chromophore ( $F_{\text{Nal(amid)}} = 1$ ) and applying equation 54 one obtains  $F_{\text{foldon}} \approx 0.04$ . The small dependence of the obtained  $F_{\text{foldon}}$  on  $F_{\text{Nal(amid)}}$  can be explained by very fast rotational diffusion of  $\text{Nal}_{\text{amid}}$  during a single encounter. Using equation 53 for the association of two foldon subunits and assuming that both foldon variants have identical surface reactivity ( $\theta_A^0 = \theta_B^0$ ) one obtains  $F_{\text{foldon}} \approx 0.098$  in the presence of 0.58 M GdmCl and  $F_{\text{foldon}} \approx 0.104$  in the presence of 1.0 M urea. Thus, the effective reactive surface of a foldon subunit is larger in the foldon-foldon reaction compared to the foldon-Nal(amid) reaction. This is expected due to specific protein-protein interactions, which are not present in the foldon-Nal<sub>amid</sub> reaction.

Reaction	Conditions	$F_{\text{reactant1}}$	$F_{\text{reactant2}}$
Trp20Xan / Trp20Nal	0.58M GdmCl	0.098	0.098
Trp20Xan / Trp20Nal	1M urea	0.104	0.104
Trp20Xan / Nal <sub>amid</sub>	1M urea	0.05	(0.7)
Trp20Xan / Nal <sub>amid</sub>	1M urea	0.04	(1.0)
Xan / 1-NAA	1M urea	0.7	0.7

**Table 15:** Summary of calculated surface reactivities of a monomeric foldon subunit,  $\text{Nal}_{\text{amid}}$  and the free labels, respectively. Values in brackets were hold fixed during calculation. Calculations were performed in collaboration with Y. von Hansen (TU München).

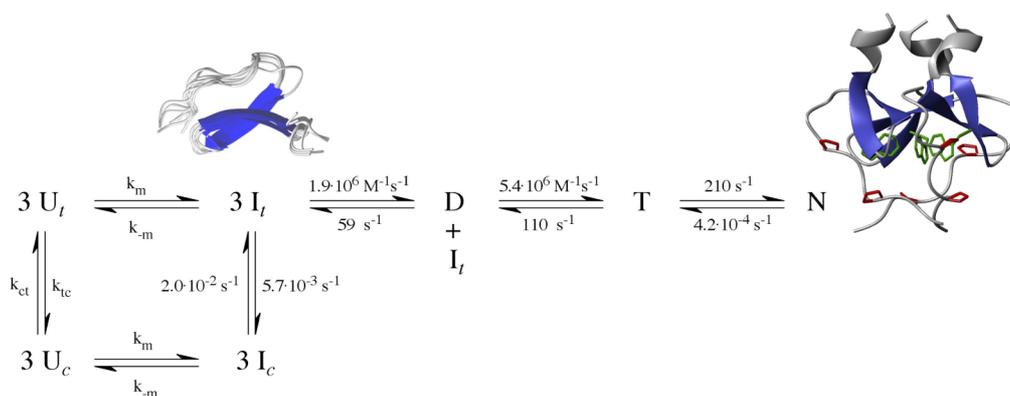
Since determination of the reactive distance  $a$  is subject to some uncertainties, we determined how the calculated surface reactivity changes with varying  $a$  for association of two foldon subunits (Figure 80).



**Figure 80:** Reactive surface on a foldon subunit vs. the reaction distance  $a$ . For a given reaction distance,  $F_{\text{foldon}}$  was adjusted such that the relation  $k_S(a) \cdot \Gamma(F_{\text{foldon}}) = k_a = 3.5 \cdot 10^8 \text{ M}^{-1} \text{ s}^{-1}$  is fulfilled. We thereby evaluate equation 53 using experimental parameters in table 13 and 14. Calculations were performed in collaboration with Y. von Hansen (TU München).

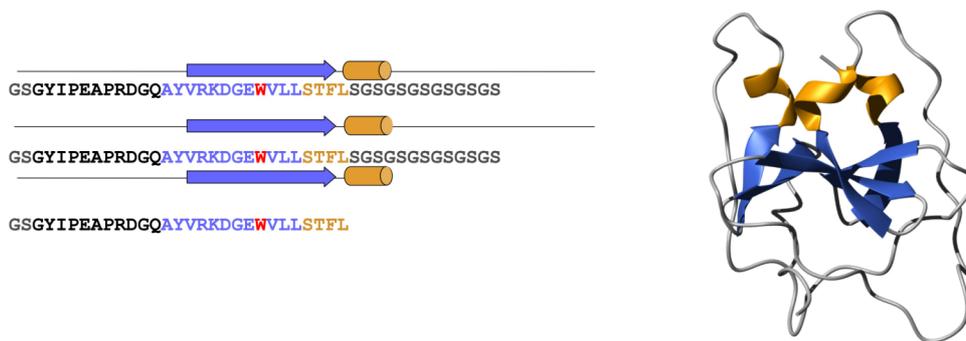
### 3.2.5 Elimination of association reactions during folding of the trimeric foldon domain

Previous studies on the trimeric foldon domain from bacteriophage T4 fibritin yielded detailed information on the mechanism of coupled folding and association during the assembly process.<sup>111</sup> During folding from the GdmCl-unfolded state a monomeric intermediate with native-like topology is formed on the submillisecond time scale. This intermediate assembles into a partially folded trimer via two very fast consecutive bimolecular reactions with rate constants of  $k_d=1.9(\pm 0.5)\cdot 10^6 \text{ M}^{-1} \text{ s}^{-1}$  and  $k_t=5.4(\pm 0.3)\cdot 10^6 \text{ M}^{-1} \text{ s}^{-1}$ . A final folding step yields the native trimer. In addition to this fast folding pathway, about 20% of the molecules fold in a slow process that is limited by prolyl isomerization reactions (Figure 62).<sup>111</sup>



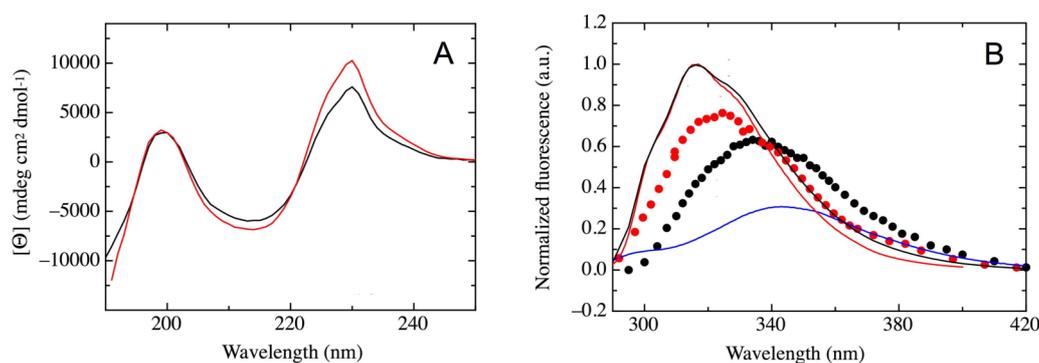
**Figure 62:** Folding/assembly mechanism of the foldon domain in the presence of 0.58 M GdmCl.<sup>111</sup>

At estimated cellular concentrations (1  $\mu$ M to 5  $\mu$ M foldon monomer) foldon assembly is limited by the association reactions. To investigate folding in the absence of the association reactions we engineered two different monomeric foldon variants in which the elementary step of free diffusion of the subunits is eliminated and thus intrachain diffusion becomes the elementary step for folding. In the single-chain foldon variant three repeats of the amino acid sequence of a monomeric foldon subunit are connected in N- to C-terminal direction to form a single continuous amino acid chain (Figure 81). The amino acid sequences of the subunits are separated by flexible 15 amino acid poly(glycine-serine) linkers (Figure 81).



**Figure 81:** Amino acid sequence (left) and model of the structure (right) of the single-chain foldon variant.

The single-chain foldon variant and the wild type trimer have identical far UV CD spectra (Figure 82A) and fluorescence spectra (Figure 82B), indicating that the single-chain foldon folds into the same structure as the trimeric wild type protein. This conclusion is supported by solution NMR experiments, which revealed identical HSQC spectra for the two proteins except for the very C- and N-terminal regions of two of the three “subunits” in single-chain foldon, in accordance with the linkers attached to these regions (Figure 83). The structural assignment by 3D TOCSY HSQC spectra confirmed the identical structures of single-chain foldon and wild type foldon (J. Habazettl, unpublished results).



**Figure 82:** A: Comparison of far UV CD spectra of the native wild type foldon domain (black) and of native single-chain foldon (red). B: Comparison of the kinetic burst phase intermediate fluorescence spectrum of wild type foldon (black circles) and single-chain foldon (red circles). The spectra of native wild type foldon and native single-chain foldon are shown as lines in black and red, respectively. The fluorescence spectrum of GdmCl-unfolded wild type foldon is shown in blue. Spectra of wild type foldon are taken from reference 111 while spectra of single-chain foldon are from M. D. Crespo, unpublished results.



Denaturant-induced equilibrium unfolding transition showed that the thermodynamic stability of the single-chain foldon variant is independent of protein concentration, as expected for a monomeric protein (Figure 84). Additionally, the transition is independent of the probe used to monitor unfolding indicating two-state behavior. The folding free energy for the single-chain foldon variant ( $\Delta G^0(\text{H}_2\text{O})$ ) is  $-66.7(\pm 0.1)$  kJ/mol. Applying the concept of an effective concentration ( $C_{\text{eff}}$ ) with

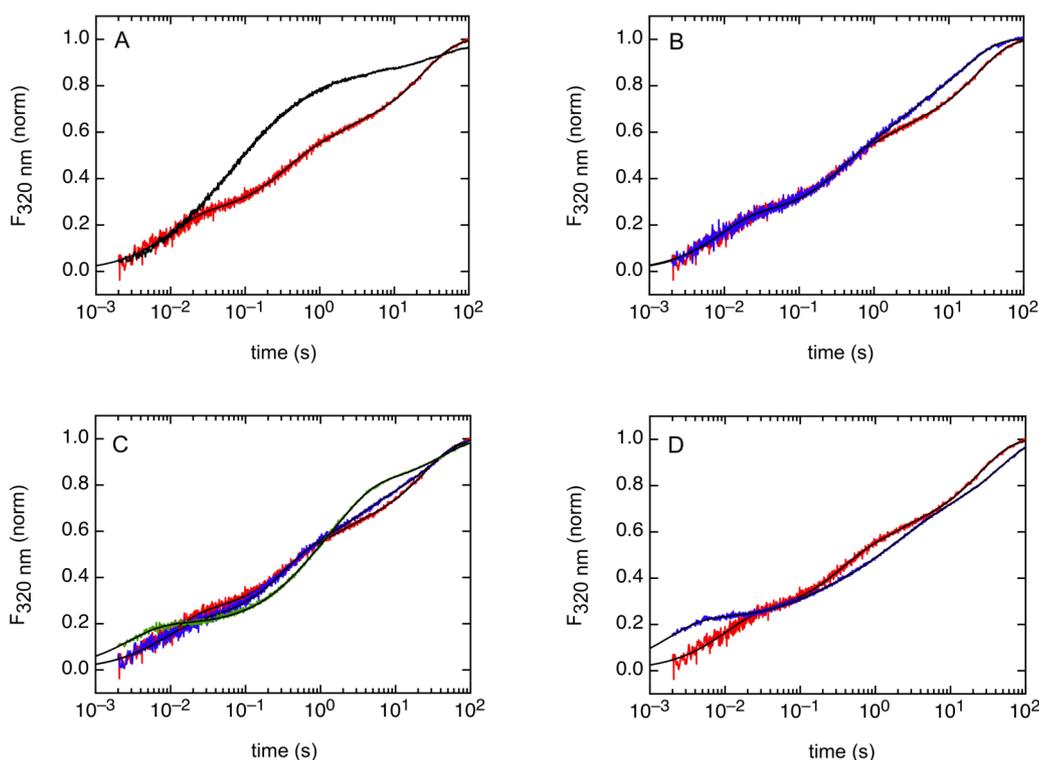
$$C_{\text{eff}} = \sqrt{\frac{K_{\text{eq}}(\text{monomer})}{K_{\text{eq}}(\text{trimer})}} \quad (55)$$

yields  $C_{\text{eff}}=22$  mM for the single-chain foldon variant.

To study the effect of very high local monomer concentrations in the single-chain foldon variant on the folding kinetics, we performed stopped-flow refolding experiments detected by the change in intrinsic tryptophane fluorescence at 320 nm. Similar to folding of the wild type foldon domain,<sup>111</sup> a burst phase intermediate is formed during folding of monomeric single-chain foldon within the dead time of the stopped-flow experiment (M. D. Crespo, unpublished results). The fluorescence spectrum of the burst phase intermediate reveals a blue shift and an increased intensity of the fluorescence maximum in the single-chain foldon variant compared to the wild type intermediate (Figure 82). This indicates that tryptophane has lower solvent accessibility in the burst phase intermediate of the single-chain foldon variant than in the monomeric intermediate observed for wild type foldon. Refolding kinetics in the single-chain foldon variant is concentration independent, as it is expected for folding of a monomeric protein. Folding in this variant turned out to be complex and very slow (Figure 85A). Folding of single-chain foldon ( $C_{\text{eff}}=22$  mM) is slower than folding of the wild type protein at a monomer concentration  $[\text{M}]_0=10\mu\text{M}$  (Figure 85A). The observed fluorescence change can be described by the sum of four exponentials (Table 16). Results from interrupted refolding experiments showed that native protein is formed in all four kinetic phases (M. D. Crespo, unpublished results). Similar as in the trimeric wild type protein the slowest kinetic phase can be attributed to isomerization reactions at the prolyl residues. This kinetic phase is accelerated in

the presence of cyclophilin 18, a peptidyl-prolyl *cis-trans* isomerase (Figure 85B). Refolding of single-chain foldon is accelerated with increasing GdmCl concentrations (Figure 85C). This effect can be attributed to both the increase in the relative contribution ( $A_{2,rel}$ ) of the second kinetic phase to the observed fluorescence change and to the significant increase in the rate constant ( $k_4$ ) of the fastest observable kinetic phase with increasing GdmCl concentration (Table 16). This observation indicates the transient formation of stable incorrectly folded structures, which have to partially unfold before they can reach the native state. This is in agreement with results from interrupted refolding experiments, which indicate the formation of two stable folding intermediates early in folding (M. D. Crespo, unpublished results). Compared to measurements at neutral pH, refolding kinetics at pH 2 are even more complex (Figure 85D). Here, the sum of five exponentials is necessary to describe the kinetic data (Table 16). An interesting observation is the significant increase in  $k_4$  at low pH compared the value determined at neutral pH. This observation is in agreement with the formation of a stable partially folded intermediate, which needs to unfold and rearrange before folding into the native conformation can occur.

These results show that, despite high local concentration of monomeric foldon subunits, folding of single-chain foldon is very slow and complex. The main problem in this variant seems to be the difficulty in finding the native topology. The results indicate that this might be attributed to non-native association from the intermediate formed on the submillisecond time scale. Only a small amount of the molecules in this intermediate state fold fast into the final native state. However, many molecules seem to be trapped in intermediates representing incorrectly folded structures stabilized by non-native interactions. These stable incorrectly folded structures have to partially unfold before they can reach the native state what slows down the folding process in single-chain foldon.

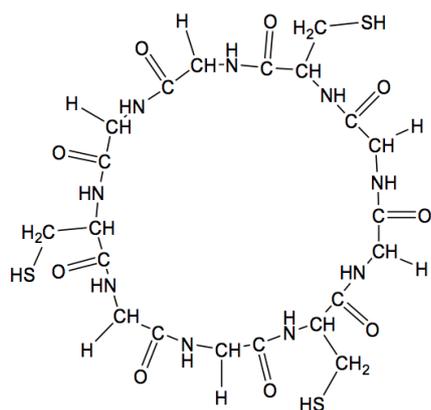


**Figure 85:** A: Folding kinetics of single-chain-foldon variant. A: in the presence of 0.58 M GdmCl pH 7.0 (red) in comparison to folding kinetics of wild type foldon (black,  $[M]_0=10\mu\text{M}$ ) at the same conditions. B: in the presence of 0.58 M GdmCl pH 7.0 (red), 0.95 M GdmCl pH 7.0 (blue) 3 M GdmCl pH 7.0 (green). C: in 0.58 M GdmCl pH 7.0 in the absence (red) and in the presence (blue) of 3.3  $\mu\text{M}$  cyclophilin 18. D: in 0.58 M GdmCl at pH 7.0 (red) and at pH 2.0 (blue). Final protein concentration for refolding kinetics of single-chain-foldon was 5  $\mu\text{M}$ . Measurements were carried out in 10 mM NaCac pH 7.0 or 10 mM Glyc pH 2.0 at  $T = 20^\circ\text{C}$ .

Folding conditions	$A_{1,\text{rel}}$	$k_1 (\text{s}^{-1})$	$A_{2,\text{rel}}$	$k_2 (\text{s}^{-1})$	$A_{3,\text{rel}}$	$k_3 (\text{s}^{-1})$	$A_{4,\text{rel}}$	$k_4 (\text{s}^{-1})$	$A_{5,\text{rel}}$	$k_5 (\text{s}^{-1})$
0.58 M GdmCl pH 7	0.39 ( $\pm 0.04$ )	0.038 ( $\pm 0.004$ )	0.12 ( $\pm 0.02$ )	0.57 ( $\pm 0.15$ )	0.26 ( $\pm 0.03$ )	3.3 ( $\pm 0.2$ )	0.24 ( $\pm 0.03$ )	107 ( $\pm 10$ )	-	-
0.95 M GdmCl pH 7	0.31 ( $\pm 0.04$ )	0.033 ( $\pm 0.004$ )	0.19 ( $\pm 0.02$ )	0.37 ( $\pm 0.08$ )	0.31 ( $\pm 0.04$ )	3.3 ( $\pm 0.4$ )	0.20 ( $\pm 0.02$ )	124 ( $\pm 12$ )	-	-
3.0 M GdmCl pH 7	0.21 ( $\pm 0.03$ )	0.024 ( $\pm 0.003$ )	0.42 ( $\pm 0.05$ )	0.53 ( $\pm 0.11$ )	0.18 ( $\pm 0.02$ )	2.8 ( $\pm 0.3$ )	0.19 ( $\pm 0.02$ )	360 ( $\pm 36$ )	-	-
0.58 M GdmCl pH 7 + 3.3 $\mu\text{M}$ cyclophilin	0.33 ( $\pm 0.04$ )	0.062 ( $\pm 0.006$ )	0.20 ( $\pm 0.02$ )	0.58 ( $\pm 0.12$ )	0.23 ( $\pm 0.03$ )	3.4 ( $\pm 0.4$ )	0.24 ( $\pm 0.03$ )	117 ( $\pm 12$ )	-	-
0.58 M GdmCl pH 2	0.34 ( $\pm 0.04$ )	0.023 ( $\pm 0.003$ )	0.21 ( $\pm 0.03$ )	0.32 ( $\pm 0.07$ )	0.14 ( $\pm 0.02$ )	1.5 ( $\pm 0.2$ )	0.22 ( $\pm 0.03$ )	559 ( $\pm 60$ )	0.09 ( $\pm 0.01$ )	10.3 ( $\pm 1.1$ )

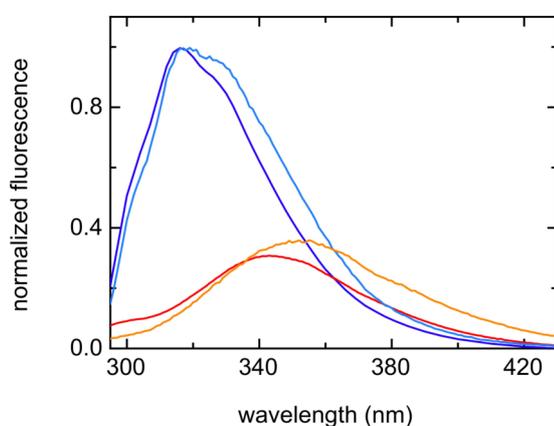
**Table 16:** Summary of the kinetic phases for refolding of single-chain-foldon obtained from fitting the kinetic traces with the sum of four or five exponentials, respectively.

In the second strategy we used a foldon variant in which three monomeric foldon subunits are attached via their N-terminus to a cyclic template, which orients the three subunits correctly relative to each other (foldon-on-template variant, Figure 86). The cyclic template mimics the situation in fibritin, where foldon is located at the C-terminal end of a long trimeric coiled-coil.



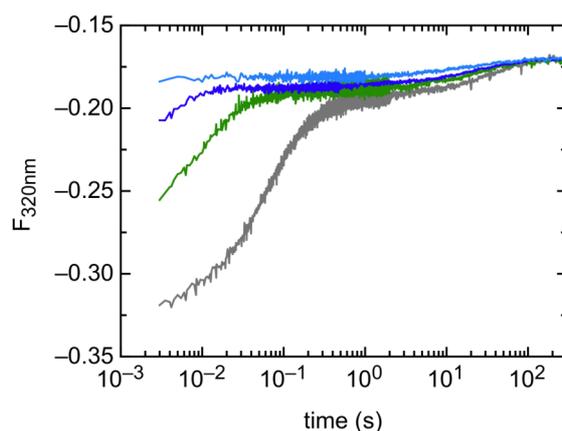
**Figure 86:** Structure of the cyclic (Gly-Gly-Cys)<sub>3</sub> template used to orient monomeric foldon subunits correctly relative to each other. Foldon subunits were attached via their N-terminus to the cysteines of the template (see materials and methods). Synthesis of the template and coupling of the foldon subunits to Cys residues were performed by W. Haehnel (Universität Freiburg).

The fluorescence spectrum of the foldon-on-template variant is similar to the spectrum of the native wild type trimer at low GdmCl concentration (Figure 87), indicating a native-like structure. In the presence of 6.38 M GdmCl pH 2.0 the fluorescence spectrum of the foldon-on-template variant is similar to the spectrum of unfolded wild type foldon (Figure 87).



**Figure 87:** Fluorescence spectra of the foldon-on-template variant in 10 mM NaCac at  $T = 20$  °C in the presence of 0.58 M GdmCl pH 7.0 (light blue) and in the presence of 6.38 M GdmCl pH 2.0 (orange). Protein concentrations were 4.2  $\mu$ M. Spectra of native (blue) and GdmCl-unfolded wild type foldon (red) are taken from reference 111.

We studied the folding kinetics of the foldon-on-template variant in stopped-flow experiments starting from unfolded protein in 6.38 M GdmCl at pH 2. At a residual GdmCl concentration of 0.58 M GdmCl the major folding reaction occurs within the dead time of the experiment ( $\sim 2$  ms), indicating very fast folding. This experiment only resolves a very slow phase (Figure 88 and table 17). Comparison with folding of wild type foldon and single-chain foldon shows that this phase can be attributed to a prolyl isomerization reaction. Refolding experiments in the presence of increasing GdmCl concentrations reveal that folding is resolved only at denaturant concentrations above 3 M (Figure 88). The increase in total amplitude of the observable fluorescence change with increasing GdmCl concentration (Figure 88) indicates that folding of foldon-on-template is significantly slowed down with increasing denaturant concentration.



**Figure 88:** Folding kinetics of foldon-on-template in 10 mM NaCac pH 7.0 in the presence of 0.58 M GdmCl (light blue), 3 M GdmCl (dark blue), 4 M GdmCl (green) and 5 M GdmCl (grey) in 10 mM NaCac pH 7.0 at  $T = 20^\circ\text{C}$ . Final protein concentration was  $4.2 \mu\text{M}$ .

Folding kinetics in the presence of 3 M, 4 M and 5 M GdmCl can be described with the sum of three exponentials. Fit results are summarized in table 17. The slow *cis-trans* isomerization reaction is not significantly affected by the denaturant concentration. The rate constant of the second kinetic phase does not show a significant dependence on the denaturant concentration either. The small contribution of this phase to the fluorescence change upon folding is a possible explanation why this phase is not observed in refolding kinetics monitored in the presence of 0.58 M GdmCl. The rate constant of the third kinetic phase is significantly decelerated with

increasing denaturant concentration, which is the typical behaviour of a folding reaction

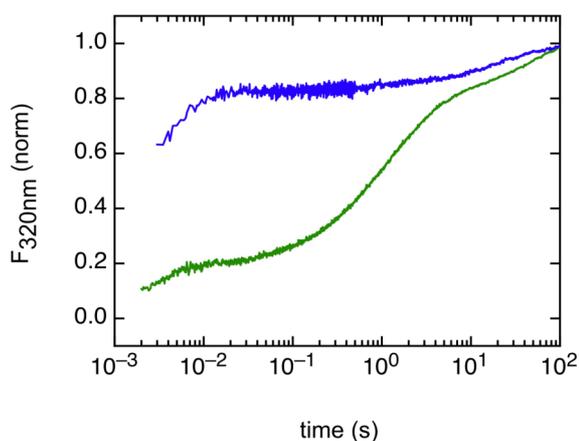
Folding conditions	$A_{1, \text{rel}}$	$k_1$ (s <sup>-1</sup> )	$A_{2, \text{rel}}$	$k_2$ (s <sup>-1</sup> )	$A_{3, \text{rel}}$	$k_3$ (s <sup>-1</sup> )
0.58 M GdmCl pH 7.0	1	0.050 (±0.020)	-	-	-	-
3.0 M GdmCl pH 7.0	0.25 (±0.04)	0.034 (±0.001)	0.05 (±0.03)	1.2 (±0.2)	0.71 (±0.07)	254 (±11)
4.0 M GdmCl pH 7.0	0.17 (±0.03)	0.036 (±0.002)	0.05 (±0.03)	3.35 (±0.4)	0.77 (±0.05)	79 (±2)
5.0 M GdmCl pH 7.0	0.15 (±0.02)	0.027 (±0.002)	0.13 (±0.03)	2.6 (±0.1)	0.72 (±0.03)	16 (±1)

**Table 17:** Results from an exponential fit to non-normalized refolding kinetics of foldon-on-template at  $T = 20$  °C.

The two folding reactions are either decelerated with increasing GdmCl concentration or are virtually independent on the denaturant concentration. This observation indicates that there is no evidence for formation of misfolded structures in both reactions. This is in contrast to results on the single-chain foldon variant, where increasing GdmCl concentrations speed up the fastest observable rate constant, indicating formation of incorrectly folded structures, which need to unfold and rearrange before folding into the native conformation can occur. In the presence of 3 M GdmCl, folding of the foldon-on-template variant is completed after around 30 ms and thus occurs around 1000 times faster compared to folding of single-chain foldon variant (Figure 89).

The results on single-chain foldon revealed that finding the native topology in this variant is difficult. This can be explained by stable incorrectly folded intermediates, which are formed early in folding. These structures have to partially unfold before they can reach the native state what drastically slows down the folding process. In contrast to single-chain foldon, the foldon subunits in foldon-on-template are oriented correctly relative to each using a cyclic template. Folding in this variant turned out to be very fast and without any evidence for incorrectly folded structures, indicating that

the correct relative orientation of the subunits is important for fast and efficient folding.



**Figure 89:** Comparison of normalized refolding kinetics for single-chain foldon (green) and foldon-on-template (blue) in 10 mM NaCac pH 7.0 in the presence of 3 M GdmCl,  $T = 20^{\circ}\text{C}$ . Final protein concentrations were  $5\ \mu\text{M}$  for single-chain foldon and  $4.2\ \mu\text{M}$  for foldon-on-template, respectively.

In future experiments, we plan to study folding of a foldon variant in which the monomeric subunits are attached via their C-terminus to a cyclic (Gly-Gly-Cys)<sub>3</sub> template. This is in contrast to the situation in nature where the foldon subunits are attached via their N-terminus to the trimeric coiled-coil domain of fibrin. These experiments allow investigation of whether the N-terminal attachment to fibrin is important for fast and efficient folding of the foldon domain.

### 3.3 Materials and Methods

#### *Peptide synthesis and purification*

All peptides were synthesized by standard fluorenylmethoxycarbonyl (Fmoc) chemistry on a 433A peptide synthesizer from Applied Biosystems. The following solid supports from Rapp polymere were used: TentaGel S RAM for GB1<sub>54-52</sub>, TentaGel R RAM (poly(glycine-serine) peptides, poly(threonine-serine)<sub>7</sub> peptide, polyglutamine peptides and Brk<sub>66-89</sub>), TentaGel R Trt-Gly Fmoc (polyserine, poly(glycine-serine) and Gln-host-guest peptide), TentaGel S Trt/PHB-Gly Fmoc (model peptides for intermolecular TTET experiments) and TentaGel R PHB-Leu Fmoc (foldon). Generally, we applied the *FastMoc* 0.10 mmol chemistry option for which employed amounts of resin and amino acids were 0.10 mmol and 1.00 mmol, respectively. We used HBTU(/HOBt) as the activation agent, 2.0 M DIPEA in NMP as the coupling agent and 22% v/v piperidine in NMP for Fmoc deprotection. Special procedures were applied in the following cases:

- For the synthesis of poly(threonine-serine)<sub>7</sub> only double building blocks (Fmoc-Ser-Thr( $\psi^{\text{Me,Me}}$ )-OH) were used and HATU was used as coupling agent.
- Several pseudoprolines (Fmoc-Gly-Ser( $\psi^{\text{Me,Me}}$ )-OH) were used for the synthesis of poly(glycine-serine) peptides.
- For synthesis of Brk<sub>66-89</sub> the use of HATU as coupling agent was essential for coupling of residues Gly66 to Phe80.
- Synthesis of peptides 4 and 6 for bimolecular TTET experiments were only successful if pseudoprolines (Fmoc-Gly-Ser( $\psi^{\text{Me,Me}}$ )-OH) were introduced.
- The yield of pure foldon E5R variants for TTET experiments was increased significantly if Fmoc-Leu-Ser( $\psi^{\text{Me,Me}}$ )-OH was introduced at position Leu23-Ser24.
- Couplings using a threefold excess of amino acid (diaminopropionic acid and pseudoprolines) were carried out manually using PyBOP as the activation agent and 44% NMM in DMF for coupling.

Peptide products were purified by reversed-phase HPLC. Purity of the peptides was analyzed by analytical HPLC and mass spectroscopy.

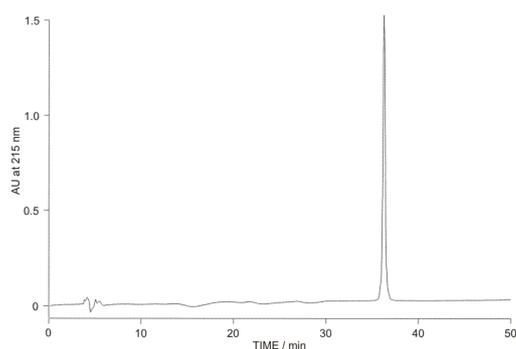
### *Synthesis of xanthonic acid and peptide labelling*

9-Oxoxanthen-2-carboxylic acid was synthesized as described<sup>212</sup> and either attached to diaminopropionic acid (Dpr) or to the N-terminus of the peptides using Fmoc chemistry. The triplet acceptor group naphthalene was introduced by the non-natural amino acid naphthylalanine.

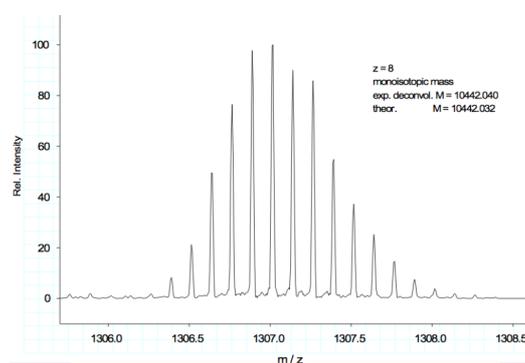
### *Synthesis and purification of the foldon-on-template variant*

Synthesis of the cyclic template (Gly-Gly-Cys)<sub>3</sub> and coupling of monomeric foldon subunits to Cys residues were performed by W. Hachnel (Universität Freiburg). The use of StBu as protection group for the cysteines revealed the best results for synthesis of the template. In contrast, the Trt protection group turned out to be too bulky.

The foldon monomers used in the foldon-on-template variant contained an additional Gly-Ser sequence at the N-terminus to increase flexibility near the template. Multiple coupling of Br-AcBr to the N-terminus of the foldon sequence was necessary to obtain a good yield in AcBr-(Gly-Ser)-Foldon. Conversion was controlled via Kaiser test. AcBr-(Gly-Ser)-Foldon was coupled in solution to the cysteines of (Gly-Gly-Cys)<sub>3</sub>. Successful coupling was confirmed by mass spectrometry (Figure 90A) and analytical HPLC (Figure 90B).



**Figure 90A:** Analytical HPLC analysis after purification of (Gly-Ser)-foldon coupled to the (Gly-Gly-Cys)<sub>3</sub> template.



**Figure 90B:** High resolution MS spectrum of (Gly-Ser)-foldon coupled to the (Gly-Gly-Cys)<sub>3</sub> template.

Optimizing the coupling protocol revealed the following important aspects. To decrease the formation of side products, the pH has to be controlled during the coupling reaction. To obtain good yields it is necessary to use an excess of BrAc-(Gly-Ser)-Foldon. Nevertheless, small amounts of templates with only two foldon monomers coupled are produced. Due to easy formation of disulfide bridges between cysteines of the template the reaction has to be performed in complete absence of oxygen. Further, the reaction has to be fast in order to prevent reaction between BrAc-(Gly-Ser)-Foldon molecules.

#### *Cloning, expression and purification of single-chain foldon*

A synthetic gene encoding single-chain foldon (SCF) was prepared with optimal codon usage for *E. coli* by Richard Kammerer (University of Manchester). Three separate PCR amplification reactions were performed using primer pairs and the codon-optimized foldon cDNA as a template. The full-length construct was assembled sequentially in pUC18 (PCR fragment SCF1 ligated into the HindII/EcoRI site of pUC19; SCF2 ligated into BamHI/EcoRI site of pUC18-SCF1; SCF3 ligated into BspEI/EcoRI site of pUC18-SCF1-SCG2). For expression, the insert coding for the SCF protein was digested with BglII/EcoRI and cloned into the BamHI/EcoRI site of pHis, a derivative of pET-15b encoding an N-terminal 6xHis tag followed by a thrombin cleavage site. Recombinant insert DNA was verified by DNA sequencing. SCF carries two additional, N-terminal residues (Gly and Ser), which originate from the thrombin cleavage site and are not part of the coding sequences.

Sequence:

```
1-42  GSGYIPEAPRDGQAYVRKDGEWVLLSTFL SGSGSGSGSGSGS  
43-84 GSGYIPEAPRDGQAYVRKDGEWVLLSTFL SGSGSGSGSGSGS  
85-113 GSGYIPEAPRDGQAYVRKDGEWVLLSTFL
```

The recombinant proteins were expressed in *E. coli* BL21(DE3) host strain after isopropyl  $\beta$ -D-thiogalactoside induction to a final concentration of 1mM. Purification of the 6xHis-tagged fusion protein was carried out by immobilized metal affinity chromatography on Chelating Sepharose Fast Flow (Amersham Biosciences) and

separation of single-chain foldon from the 6×His-tagged carrier protein, thioredoxin, by thrombin cleavage was performed as described in the manufacturer's instructions.

#### *Disaggregation of polyglutamine peptides*

Polyglutamine peptides were disaggregated according to the protocol introduced by O'Nuallain et al.:<sup>190</sup>

1. Lyophilized purified peptide is dissolved in a 1:1 mixture (v/v) of TFA and HFIP to 0.5 mg/ml in a sealed Erlenmeyer flask and incubated at room temperature for 5 hours.
2. The peptide solution is dried under a stream of argon gas and then immediately placed under vacuum for around 30 minutes to remove any residual volatile solvents. The vacuum is vent with argon gas.
3. The dry peptide is dissolved in pH 3 aqueous TFA to a concentration in the range of 50 to 100  $\mu\text{M}$  and centrifuged at 386000 g for 3 hours.
4. The upper two thirds of the solution are removed and adjusted to final buffer conditions by a tenfold dilution with 100 mM NaCac.

The lower third of the solution from the centrifugation step was adjusted to final buffer conditions and used as a positive control in the test for successful disaggregation (Figure 39).

#### *Determination of peptide and protein concentrations*

Concentrations of all peptides, proteins and 2-NAA were determined spectrophotometrically using the Lambert-Beer relationship. The following molar absorption coefficients were used:  $\epsilon_{343} = 3900 \text{ M}^{-1}\text{cm}^{-1}$  for xanthone,  $\epsilon_{282} = 5550 \text{ M}^{-1}\text{cm}^{-1}$  for naphthalene and 2-NAA,  $\epsilon_{280} = 8250 \text{ M}^{-1}\text{cm}^{-1}$  for a foldon monomer,  $\epsilon_{280} = 24750 \text{ M}^{-1}\text{cm}^{-1}$  for single-chain foldon and foldon-on-template. For 1-NAA and 2-NAA we used the same molar absorption coefficient as for naphthalene. However, for 2-NAA the absorption maximum is shifted to 276 nm and  $\epsilon$  is probably smaller (experimental determination:  $\epsilon_{276} = 4742(\pm 200) \text{ M}^{-1}\text{cm}^{-1}$ ). This would lead to a

decrease in bimolecular rate constants  $k_A$  given in chapter 3.2.1 by a factor of 1.3( $\pm$ 0.1) for reactions in which 2-NAA is involved.

#### *Viscosity measurements*

Solvent viscosity was measured in a HAAKE falling-ball viscosimeter Type C from Thermo Scientific.

#### *Conductivity measurements*

Ion movement was studied by measuring electrical conductivity of a 50 mM NaCl solution using a 217 conductometer from Metrohm. All measurements were performed at 22.5 °C.

#### *CD spectroscopy and determination of the helix content in Brk<sub>66-89</sub>*

All CD experiments were carried out on an AVIV 62ADS or an AVIV 410 spectropolarimeter (AVIV, USA). The helicity of Brk<sub>66-89</sub> was calculated from the mean residual ellipticity at 222 nm ( $[\theta]_{222}$ ). For the calculation we used values of  $[\theta]_H = -44000 \cdot (1-3/N) \text{ deg cm}^2 \text{ dmol}^{-1}$ , where N is the number of residues, and  $[\theta]_C = 2200 \text{ deg cm}^2 \text{ dmol}^{-1}$  for the 100% helical and 100% unfolded state, respectively.<sup>213</sup> Due to the contribution of the TTET labels to  $[\theta]_{222}$  all calculated helix propensities are slightly too large.

#### *Fluorescence spectroscopy*

Equilibrium fluorescence measurements were carried out on an AMINCO Bowman series 2 spectrofluorimeter (SLM Aminco, USA).

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. The excitation wavelength was 280 nm (4.5 nm bandwidth) for final foldon monomer concentrations up to 10  $\mu$ M and 295 nm (4.5 nm bandwidth) for higher concentrations. Emission was detected at 320 nm (12 nm bandwidth).

### *Laser flash photolysis*

Laser flash photolysis provides a method that allows rapid production of excited states. This makes it a perfect tool to study energy transfer reactions such as TTET. Kinetics of nanosecond-laser flash experiments was collected using a laser flash photolysis reaction analyser model LKL.60 from Applied Photophysics. Here, TTET was induced by a 4 ns laser pulse at 355 nm from a Nd:Yag laser. In the picosecond-laser flash experiments TTET was induced by a 120 ps laser flash. Peptide concentrations for intramolecular TTET measurements were in the 20-100  $\mu\text{M}$  range. Transient triplet absorbance of xanthone was measured at 590 nm and the kinetics were analyzed using ProFit (Quansoft, Zurich, Switzerland).

### *Dynamic light scattering measurements*

DLS experiments were carried out on an ALV DLS/SLS-5000 system equipped with a compact goniometer system. For all samples, the apparent hydrodynamic radius  $r_{\text{H,app}}$  was measured as a function of the solute concentration and the effective hydrodynamic radius  $r_{\text{H}}$  was determined from extrapolation of  $r_{\text{H,app}}$  to  $c \rightarrow 0$ .

### *NMR measurements*

NMR experiments were carried out on a BRUKER DRX 600 spectrometer using a broadband inverse probe head equipped with z-axis pulsed field gradients. The probe temperature was calibrated using 100% methanol for temperatures below 305 K and 100% ethylene glycol for higher temperatures.<sup>214</sup> A continuous wave spinlock was used during the mixing time (416 ms) of the ROESY sequence.<sup>215</sup> The homonuclear pulse programs incorporated the WATERGATE pulse sequence for solvent signal suppression.<sup>216</sup> TOCSY experiments used the MLEV17 sequence with gradients.<sup>217</sup> NMR data were processed using the Bruker software XWINNMR and analyzed with the program SPARKY.<sup>218</sup>

### *Data fitting*

Data evaluation was carried out using the program ProFit (Quantumsoft, Zurich, Switzerland). For the global fit of concentration-dependent assembly kinetics of the foldon domain and interrupted refolding experiments (chapter 3.2.2) we used KinTek

Explorer.<sup>219,220</sup> Rate constants and spectroscopic parameters of the different kinetic species were fitted as global parameters. The rate constants for unfolding of the dimeric intermediate and the native protein were fixed to experimentally determined values. The equilibrium constant between unfolded monomer and native protein measured by CD and fluorescence was used as an additional constraint for the global fit.

## 4. Summary

The first part of this thesis addresses intramolecular chain diffusion processes during protein folding. An unfolded polypeptide chain has to explore a vast conformational space to search for favourable intramolecular interactions in order to find its native state. Formation of site-specific interactions within a polypeptide chain is thus regarded as the elementary step during conformational search and sets an upper limit to protein folding rate constants. We applied triplet-triplet energy transfer (TTET) to determine the rate constants for formation of site-specific intramolecular interactions in polypeptide chains. TTET from the triplet donor xanthone to the triplet acceptor naphthalene is a diffusion-controlled process that is based on van-der-Waals contact between the chromophores. This method thus allows to directly monitor the kinetics of formation of site-specific contacts on the time scale from picoseconds to tens of microseconds. Earlier experiments on the dynamics of end-to-end loop formation in unfolded poly(glycine-serine) and polyserine chains indicated different rate-limiting steps for the formation of short and long loops. To understand the dynamics in the different regimes in more detail we investigated the effect of solvent viscosity and of temperature on the kinetics of end-to-end loop formation in different polypeptide chains. The rate constant for formation of large loops is inversely proportional to solvent viscosity in the presence of small co-solutes like ethylene glycol and glycerol, indicating a diffusion-controlled reaction. In the presence of large co-solutes loop formation kinetics show a weaker viscosity-dependence, indicating that the macroscopic solvent viscosity and the microscopic viscosity experienced by a polypeptide chain are identical only for small co-solutes. Small co-solutes can thus be used to determine the actual viscosity-dependence of loop formation. The weaker effects of large co-solutes on loop formation indicates that molecular crowding inside cells, which is due to the presence of high molecular weight compounds, has only little effect on the dynamics of loop formation. Formation of short loops shows a fractional viscosity-dependence even for small co-solutes, suggesting the presence of intrinsic energy barriers. This model is supported by the observation of significant activation energies of up to 17 kJ/mol for formation of short loops, whereas formation of large loops has a low activation energy around 5 kJ/mol in accordance with a diffusion-controlled reaction.

To investigate the dynamics of loop formation in an unfolded natural amino acid sequence we performed TTET experiments on a 24 amino acid fragment from the natively unfolded brinker protein. The results showed that end-to-end loop formation in this fragment is significantly slowed down due to repulsive interactions between positively charged residues. Possible explanations for this effect are increased internal friction and/or increased chain dimension. Inorganic phosphate or sodium chloride induces formation of helical structure in the brinker fragment, which enabled us to test for the effect of secondary structure on the dynamics of loop formation. The experiments revealed that loop formation between the ends of flexible helical structures is decelerated. We further studied loop formation in sequences derived from the natural protein  $\beta$ -crystallin and the B1 domain of protein G. In all studied natural protein sequences we found fractional viscosity dependencies and significant intrinsic barriers for loop formation, indicating that chain dynamics in these fragments is not fully diffusion-controlled. TTET experiments cannot completely resolve loop formation kinetics in the eight amino acid  $\beta$ -hairpin fragment from protein G, indicating subnanosecond reactions of loop formation in this peptide, which suggests that this sequence is optimized for rapid and efficient formation of tight turns.

Polyglutamine stretches beyond a critical length tend to form aggregates, which are assumed to play a central role in various neurodegenerative diseases. TTET was used to study whether unfolded polyglutamine chains exhibit different dynamical and structural properties compared to other unfolded polypeptide chains. The polyglutamine peptides showed significantly decelerated end-to-end loop formation compared to poly(glycine-serine) and polyserine chains. This effect increases with increasing chain length. Studying the effect of solvent viscosity and of temperature on the reaction revealed that slow chain dynamics in polyglutamine chains can be mainly attributed to increased activation barriers.

The dynamics of loop formation are determined by the dimensions and by the internal motions of polypeptide chains. To obtain more information on these properties we performed time resolved fluorescence resonance energy transfer (FRET) experiments on a long poly(glycine-serine) chain. The results revealed formation of non-specific intramolecular hydrogen bonds in water, which reduces the end-to-end distance and the end-to-end diffusion coefficient compared to an ideal chain. Addition of high

concentrations of GdmCl, which represents a good solvent for polypeptide chains, increases the end-to-end distance and accelerates local end-to-end diffusion. This result suggests that non-specific intramolecular hydrogen bonds contribute to chain compaction in water, which is often observed early during protein folding and is usually attributed to hydrophobic interactions.

The second part of this thesis aimed at the detailed understanding of intermolecular diffusion processes during protein folding and assembly. We applied TTET to study site-specific formation of intermolecular contacts between unfolded model polypeptide chains and between unfolded model polypeptides and the small molecule naphthylacetic acid. In accordance with theory, the experiments showed that diffusion-controlled association is strongly influenced by the net charge of the interacting molecules. Results further revealed that the association reaction is slowed down by orientational constraints compared to the theoretical Smoluchowski diffusion limit. However, reorientation after initial contact formation can reduce the retarding effect of these geometric constraints.

In further studies the homotrimeric foldon domain from bacteriophage T4 fibrin served as a model to investigate the formation of site-specific protein-protein interactions during folding and assembly of an oligomeric protein. The foldon domain comprises the C-terminal 27 amino acids of each subunit of the fibrin protein, which consists in total of 486 amino acids per subunit. The function of this small globular domain is to assist assembly of the trimer into a rod-like coiled-coil structure. During folding/assembly from the GdmCl-unfolded state the foldon domain forms a monomeric intermediate in a very fast reaction on the submillisecond time scale. Assembly of this intermediate into a trimeric state occurs in two very fast consecutive bimolecular association reactions with rate constants of  $1.9 \cdot 10^6 \text{ M}^{-1} \text{ s}^{-1}$  and  $5.4 \cdot 10^6 \text{ M}^{-1} \text{ s}^{-1}$ , respectively. The resulting trimer is rapidly converted to the native state. These experiments were carried out at a residual GdmCl concentration of 0.58 M, which largely shields electrostatic interactions. Structural studies had further indicated an asymmetric charge distribution on the surface of the monomeric intermediate proposing an electrostatic steering of the monomer-monomer and the monomer-dimer interactions. We tested the effect of ionic strength and stabilizing co-solutes on thermodynamic stability and on the assembly kinetics of the foldon domain.

Increasing ionic strength stabilizes the native trimer. Kinetic experiments revealed that both association reactions are slowed down at low ionic strength, indicating unfavorable electrostatic interactions in the two assembly processes. The magnitude of the effects can be quantitatively explained by the net charge of the interacting molecules. As a result trimer formation is affected more strongly than dimer formation. Additionally, increasing the ionic strength strongly accelerates unfolding of the native state. These results indicate significant unfavourable electrostatic interactions in the transition state of trimerization and unfolding and argue against an electrostatic steering mechanism during recognition of the subunits.

It is an unsolved problem how fast two proteins can maximally encounter each other at specific sites during formation of protein-protein interactions. To study the maximum rate constant for site-specific formation of an encounter complex between monomeric intermediates of the foldon domain we applied TTET to the foldon E5R variant. In this variant an intermolecular salt bridge is replaced by a repulsive interaction, which results in a monomeric protein that structurally resembles the monomeric assembly intermediate. Our results showed that the speed-limit for formation of a site-specific encounter complex between two monomeric foldon intermediates at their interaction sites is  $3.7 \cdot 10^8 \text{ M}^{-1} \text{ s}^{-1}$ . This reaction is not affected by electrostatic contributions. Debye-Smoluchowski theory predicts a rate constant of  $4.6 \cdot 10^9 \text{ M}^{-1} \text{ s}^{-1}$  for non-specific encounter complex formation between two foldon monomers. This comparison shows that only 8% of all random encounters between the subunits lead to a site-specific encounter, indicating significant orientational constraints for this reaction. Comparison with the dynamics of formation of a dimeric assembly intermediate with  $k = 1.1 \cdot 10^6 \text{ M}^{-1} \text{ s}^{-1}$  at low ionic strength reveals that one out of 300 site-specific encounter complexes and one out of 4000 random complexes leads to formation of a stable dimer, indicating significant enthalpic and/or entropic barriers for dimerization.

At physiological protein concentrations assembly of wild type foldon is limited by bimolecular reactions, which lead to transient formation of a dimeric and a trimeric intermediate, respectively. We engineered two different monomeric foldon variants in which the elementary step of free diffusion of the subunits is eliminated and folding becomes an intramolecular process. In the single-chain foldon variant, three repeats of

the amino acid sequence of a monomeric foldon subunit are connected in N- to C-terminal direction to form a single continuous amino acid chain, resulting in a very high local monomer concentration. Folding of this variant is very slow and complex, indicating that formation of stable intermediates with non-native interactions cause slow folding reactions in this variant. This model is supported by the observation that folding is accelerated under destabilizing conditions. In a second approach, a foldon on-template variant was constructed with three monomeric foldon subunits attached via their N-terminus to a cyclic template. This orients the three subunits correctly relatively to each other, mimicking the situation in fibrin, where foldon is located at the C-terminal end of a long trimeric coiled-coil. Folding of the foldon on-template variant is very fast and does not show any evidence for formation of misfolded structures. Comparison with the results from single-chain foldon indicates that the correct relative orientation of the subunits is important for fast and efficient folding and that formation of stable intermediates early in folding prevents effective search for the native topology in the single-chain variant.

## 5. Acknowledgements

This work was carried out from November 2005 until April 2011 in the laboratory of Prof. Dr. Thomas Kiefhaber in the Department of Biophysical Chemistry at the Biozentrum of the University of Basel (November 2005 – April 2007) and in the Department of Chemistry of the Technische Universität München (May 2007 – April 2011).

First of all, I want to thank Prof. Dr. Thomas Kiefhaber for his excellent supervision. I am very grateful for his ideas and his motivating support. I very much enjoyed working on the interesting projects of peptide dynamics and protein folding.

Many thanks to all former and current members of the labs in Basel and in Munich: Alexander, Andreas M., Andreas R., Annett, Beat, Christian, Christophe, Daniel, Florian, Judith, Josef, Kerstin, Kristine, Manuela, Marilò, Matthias, Michael, Nathalie, Peter, Rita, Robert, Sabine K., Sabine R., Sarah, Stefan, Therese, Tobias A., Tobias S., Traudl, Ursula S. and Ursula Z. Many thanks for the support and the nice atmosphere in the lab.

Thanks to Prof. Dr. Wolfgang Haehnel from the University of Freiburg for the nice collaboration on the foldon-on-template project and to Yann v. Hansen from the physics department of the Technische Universität München for performing the calculations on the reactive surface on a foldon subunit. Many thanks to Florian Krieger who performed part of the experiments on the effect of solvent viscosity and temperature on intrachain dynamics. Special thanks also to Judith Habazettl for performing the NMR measurements.

Thanks to Michael Hösl, Sabine Rauch, Matthias Stecher, and Ursula Zinth for proofreading part of this thesis.

Many thanks to Annett for the joint exploration of the Munich nightlife and culture. She contributed much to fast acclimatization after we had moved from Basel to Munich. Many thanks also to my family and all my friends, especially to my parents and to Matthias, for their support not only during this thesis but during my whole life.

## 6. References

1. Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science* 181, 223-230.
2. Santoro, M. M. & Bolen, D. W. (1988). Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of the phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry* 27, 8063-8068.
3. 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.
4. Levinthal, C. (1968). Are there pathways for protein folding? *J. Chim. Phys.* 65, 44-45.
5. Levinthal, C. (1969). *Mössbauer Spectroscopy in Biological Systems*, Allerton House, Monticello, Ill.
6. Jackson, S. E. (1998). How do small single-domain proteins fold? *Folding & Design* 3, R81-R91.
7. Kim, P. S. & Baldwin, R. L. (1982). Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu. Rev. Biochem.* 51, 459-489.
8. Kim, P. S. & Baldwin, R. L. (1990). Intermediates in the folding reactions of small proteins. *Annu. Rev. Biochem.* 59, 631-660.
9. Brems, D. N., Plaisted, S. M., Dougherty, J. J., Jr. & Holzman, T. F. (1987). The kinetics of bovine growth hormone folding are consistent with a framework model. *J. Biol. Chem.* 262, 2590-2596.
10. Kuwajima, K. (1989). The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins: Struct. Funct. Gen.* 6, 87-103.
11. Fersht, A. R. (1993). Protein folding and stability: the pathway of folding of barnase. *FEBS Lett.* 325, 5-16.
12. Karplus, M. & Weaver, D. L. (1976). Protein folding dynamics. *Nature* 260, 404-406.

13. Karplus, M. & Weaver, D. L. (1994). Protein folding dynamics: the diffusion-collision model and experimental data. *Protein Sci.* 3, 650-668.
14. Baldwin, R. L. & Rose, G. D. (1999A). Is protein folding hierarchic? I. Local structure and protein folding. *TIPS* 24, 26-33.
15. Baldwin, R. L. & Rose, G. D. (1999B). Is protein folding hierarchic? II. Folding intermediates and transition states. *TIPS* 24, 77-83.
16. Wetlaufer, D. B. (1973). Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc. Natl. Acad. Sci. USA* 70, 697-701.
17. 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: Struct. Funct. Gen.* 21, 167-195.
18. 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.
19. Zwanzig, R., Szabo, A. & Bagchi, B. (1992). Levinthal's paradox. *Proc. Natl. Acad. Sci. USA* 89, 20-22.
20. Jackson, S. E. & Fersht, A. R. (1991). Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* 30, 10428-10435.
21. Jackson, S. E. & Fersht, A. R. (1991). Folding of chymotrypsin inhibitor 2. 2. Influence of proline isomerization on the folding kinetics and thermodynamic characterization of the transition state of folding. *Biochemistry* 30, 10436-10443.
22. Bachmann, A. & Kiefhaber, T. (2001). Apparent two-state tendamistat folding is a sequential process along a defined route. *J. Mol. Biol.* 306, 375-386.
23. Sanchez, I. E. & Kiefhaber, T. (2003). Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J. Mol. Biol.* 325, 367-376.
24. Eyring, H. (1935). The activated complex in chemical reactions. *J. Chem. Phys.* 3, 107-115.
25. Kramers, H. A. (1940). Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* 4, 284-304.

26. 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.
27. Pohl, F. M. (1976). Temperature-dependence of the folding kinetics of folding of chymotrypsinogen A. *FEBS Letters* 65, 293-296.
28. Schönbrunner, N., Pappenberger, G., Scharf, M., Engels, J. & Kiefhaber, T. (1997). Effect of preformed correct tertiary interactions on rapid two-state tendamistat folding: evidence for hairpins as initiation sites for  $\beta$ -sheet formation. *Biochemistry* 36, 9057-9065.
29. Schätzle, M., & Kiefhaber, T. (2006). Shape of the free energy barriers for protein folding probed by multiple perturbation analysis. *J. Mol. Biol.* 357, 655-664.
30. Leffler J. E. (1953). Parameters for the description of transition states. *Science* 117, 340-341.
31. Fersht, A. R. (1999). *Structure and Mechanism in Protein Science. A Guide to Enzyme Catalysis and Protein Folding*. Palgrave Macmillan, Basingstoke, United Kingdom.
32. Brandts, J. F. (1965). The nature of the complexities in the Ribonuclease conformational transition and the implications regarding clathrating. *J. Am. Chem. Soc.* 87, 2759-2760.
33. Brant, D. A., Miller, W. G. & Flory, P. J. (1967). Conformational energy estimates for statistically coiling polypeptide chains. *J. Mol. Biol.* 23, 47-65.
34. Schimmel, P. R. & Flory, P. J. (1967). Conformational energy and configurational statistics of poly-L-proline. *Proc. Natl. Acad. Sci. USA* 58, 52-59.
35. Miller, W. G., Brant, D. A. & Flory, P. J. (1967). Random coil configurations of polypeptide copolymers. *J. Mol. Biol.* 23, 67-80.
36. Ramachandran, G. N. & Sasisekharan, V. (1968). Conformation of polypeptides and proteins. *Adv. Protein Chem.* 23, 283-438.
37. Flory, P. J. (1969). *Statistical Mechanics of Chain Molecules*, Hanser Publishers, Munich.

38. Kuhn, W. (1934). Über die Gestalt fadenförmiger Moleküle in Lösungen. *Kolloid-Z.* 52, 269.
39. Kuhn, W. (1936). Beziehungen zwischen Molekülgröße, statistischer Molekülgestalt und elastischen Eigenschaften hochpolymerer Stoffe. *Kolloid-Z.* 76, 258.
40. Flory, P. J. (1953). *Principles of polymer chemistry*, Cornell University Press, Ithaca.
41. 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.* 97, 12565-12570.
42. Rouse, P. E. (1953). A theory of the linear viscoelastic properties of dilute solutions of coiling polymers. *J. Chem. Phys.* 21, 1272-1280.
43. Zimm, B. (1956). Dynamics of polymer molecules in dilute solutions: viscoelasticity, flow birefringence and dielectric loss. *J. Chem. Phys.* 24, 269-278.
44. Bixon, M. & Zwanzig, R. (1978). Optimized Rouse-Zimm theory for stiff polymers. *J. Chem. Phys.* 68, 1896-1902.
45. Szabo, A., Schulten, K. & Schulten, Z. (1980). First passage time approach to diffusion controlled reactions. *J. Chem. Phys.* 72, 4350-4357.
46. 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.
47. 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.
48. Fierz, B., Joder, K., Krieger, F. & Kiefhaber, T. (2006). Using triplet-triplet energy transfer to measure conformational dynamics in polypeptide chains, in *Methods in Molecular Biology vol. 350: Protein Folding Protocols*, Y. Bai, R. Nussinov, Editors, Humana Press.
49. Dexter, D. L. (1953). A theory of sensitized luminescence in solids. *J. Chem. Phys.* 21, 836-850.
50. Turro, N. (1991). *Modern Molecular Photochemistry*, University Science

Books, Sausalito, California.

51. 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.
52. Fierz, B., Satzger, H., Root, C., Gilch, P., Zinth, W. & Kiefhaber, T. (2007). Loop formation in unfolded polypeptide chains on the picoseconds to microseconds time scale. *Proc. Natl. Acad. Sci.* 104, 2163-2168.
53. 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.
54. Reiner, A., Henklein, P. & Kiefhaber, T. (2010). An unlocking/relocking barrier in conformational fluctuations of villin headpiece subdomain. *Proc. Natl. Acad. Sci. USA* 107, 4955-4960.
55. Krieger, F., Möglich, A. & Kiefhaber, T. (2004). Effect of proline and glycine residues on dynamics and barriers of loop formation in polypeptide chains. *J. Am. Chem. Soc.* 127, 3346-3352.
56. Huang, F., Hudgins, R. R. & Nau, W. N., (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.
57. 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.
58. Sahoo, H., Hennig, A. & Nau, W. N. (2006). Temperature-dependent loop formation kinetics in flexible peptides studied by time-resolved fluorescence spectroscopy. *Int. J. Photoenergy* DOI 10.1155/IJP/2006/89638, 1-9.
59. Roccatano, D., Sahoo, H., Zacharias, M. & Nau, W. N. (2007). Temperature dependence of looping rates in a short peptide. *J. Phys. Chem. B* 111, 2639-2646.

60. Bent, D. V. & Hayon, E. (1975). Excited state chemistry of aromatic amino acids and related peptides. III. Tryptophan. *J. Am. Chem. Soc.* 97, 2612-2619.
61. Gonelli, M. & Strambini, G. B. (1995). Phosphorescence lifetime of tryptophan in proteins. *Biochemistry* 34, 13847-13857.
62. Lapidus, L. J., Eaton, W. A. & Hofrichter, J. (2000). Measuring the rate of intramolecular contact formation in polypeptides. *Proc. Natl. Acad. Sci.* 97, 7220-7225.
63. Yeh, I.-C. & Hummer, G. (2002). Peptide loop-closure kinetics from microsecond molecular dynamics simulation in explicit solvent. *J. Am. Chem. Soc.* 124, 6563-6568.
64. Lapidus, L. J., Steinbach, P. J., Eaton, W. E, Szabo, A. & Hofrichter, J. (2002). Effects of chain stiffness on the dynamics of loop formation in polypeptides. Appendix: Testing a 1-dimensional diffusion model for peptide dynamics. *J. Phys. Chem. B.* 106, 11628-11640.
65. Buscaglia, M., Lapidus, L. J., Eaton, W. A. & Hofrichter, J. (2006). Effects of denaturants on the dynamic of loop formation in polypeptides. *Biophys. J.* 91, 276-288.
66. Neuweiler, H., Schulz, A., Böhmer, 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.
67. Vaiana, A. C., Neuweiler, H. & Sauer, M. (2003). Fluorescence quenching of dyes by tryptophan: interactions at atomic detail from combination of experiment and computer simulation. *J. Am. Chem. Soc.* 125, 14564-14572.
68. Doose, S., Neuweiler, H. & Sauer, M. (2005). A close look at fluorescence quenching of organic dyes by tryptophan. *Chem. Phys. Chem.* 6, 2277-2285.
69. Neuweiler, H., Löllmann, M., Doose, S. & Sauer, M. (2007). Dynamics of unfolded polypeptide chains in crowded environment studied by fluorescence correlation spectroscopy. *J. Mol. Biol.* 365, 856-869.
70. Bates, G. (2003). Huntingtin aggregation and toxicity in Huntington's disease. *Lancet* 361, 1642-1644.
71. Chen, S., Berthelie, W. Y. & Wetzel, R. (2001). Polyglutamine aggregation behaviour *in vitro* supports a recruitment mechanism of cytotoxicity. *J. Mol. Biol.* 311, 173-182.

72. Yang, W., Dunlap, J. R., Andrews, R. B. & Wetzel, R. (2002). Aggregated polyglutamine peptides delivered to nuclei are toxic to mammalian cells. *Hum. Mol. Genet.* 11, 2905-2917.
73. Jana, N. R., Tanaka, M., Wang, G. & Nukina, N. (2000). Polyglutamine length-dependent interaction of Hsp40 and Hsp70 family chaperones with truncated N-terminal huntingtin: their role in suppression of aggregation and cellular toxicity. *Hum. Mol. Gen.* 9, 2009-2018.
74. Satyal, S. H., Schmidt, E., Kitagawa, K., Sondheimer, N., Lindquist, S., Kramer, J. M. & Morimoto, R. I. (2000). Polyglutamine aggregates alter protein folding homeostasis in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* 97, 5750-5755.
75. Carmichael, J., Chatellier, J., Woolfson, A., Milstein, C., Fersht, A. R. & Rubinsztein, D. C. (2000). Bacterial and yeast chaperones reduce both aggregate formation and cell death in mammalian cell models of Huntington's disease. *Proc. Natl. Acad. Sci. USA* 97, 9701-9705.
76. Zhang, X., Smith, D. L., Meriin, A. B., Engemann, S., Russel, D. E., Roark, M., Washington, L. S., Maxwell, M. M., Marsh, J. L., Thompson, L. M., Wanker, E. E., Young, A. B., Housman, D. E., Bates, G. P., Sherman, M. Y. & Kazantsev, A. G. (2005). A potent small molecule inhibits polyglutamine aggregation in Huntington's disease neurons and suppresses neurodegeneration *in vivo*. *Proc. Natl. Acad. Sci. USA* 102, 892-897.
77. Tanaka, M., Machida, Y., Niu, S., Ikeda, T., Jana, N. R., Doi, H., Kurosawa, M., Nekooki, M. & Nukina, N. (2004). Trehalose alleviates polyglutamine-mediated pathology in a mouse model of Huntington disease. *Nat. Med.* 10, 148-154.
78. Sanchez, I., Mahlke, C. & Yuan, J. (2003). Pivotal role of oligomerization in expanded polyglutamine neurodegenerative disorders. *Nature* 421, 373-379.
79. Bodner, R. A., Outeiro, T. F., Altmann, S., Maxwell, M. M., Cho, S. H., Hyman, B. T., McLean, P. J., Young, A. B., Housman, D. E. & Kazantsev, A. G. (2006). Pharmacological promotion of inclusion formation: A therapeutic approach for Huntington's and Parkinson's diseases. *Proc. Natl. Acad. Sci. USA* 103, 4246-4251.

80. Chen, S., Bertheliet, V., Hamilton, J. B., O’Nuallain, B. & Wetzel, R. (2002). Amyloid-like features of polyglutamine aggregates and their assembly kinetics. *Biochemistry* 41, 7391-7399.
81. Bhattacharyya, A., Thakur, A. & Wetzel, R. (2005). Polyglutamine aggregation nucleation: Thermodynamics of a highly unfavorable protein folding reaction. *Proc. Natl. Acad. Sci. USA* 102, 15400-15405.
82. Wetzel, R. (2006). Kinetics and thermodynamics of amyloid fibril assembly. *Acc. Chem. Res.* 39, 671-679.
83. Crick, S. L., Jayaraman, M., Frieden, C., Wetzel, R. & Pappu, R. V. (2006). Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions. *Proc. Natl. Acad. Sci.* 103, 16764-16769.
84. Walters, R. H. & Murphy, R. M. (2009). Examining polyglutamine peptide length: a connection between collapsed conformations and increased aggregation. *J. Mol. Biol.* 393, 978-992.
85. Chen, S., Ferrone, F. & Wetzel, R. (2002). Huntington’s disease age-of-onset linked to polyglutamine aggregation nucleation. *Proc. Natl. Acad. Sci. USA* 99, 11884-11889.
86. Lee, C. C., Walters, R. H. & Murphy, R. M. (2007). Reconsidering the mechanism of polyglutamine peptide aggregation. *Biochemistry* 46, 12810-12820.
87. Singh, V. R. & Lapidus, L. J. (2008). The intrinsic stiffness of polyglutamine peptides. *J. Phys. Chem. B* 112, 13172-13176.
88. von Smoluchowski, M. V. (1917). Versuch einer mathematischen Theorie der Koagulationskinetik kolloider Lösungen. *Z. Phys. Chem.* 92, 129-168.
89. Bamford, C. H., Tipper, C. F. H. & Compton, R. G. (1985). Comprehensive chemical kinetics, Vol. 25: Diffusion-limited reactions. Elsevier Science Publishers.
90. Debye, P. (1942). Reaction rates in ionic solutions. *Transactions of the Electrochemical Society* 82, 265-272.
91. Debye, P. and Hückel, E. (1923). The theory of electrolytes. I. Lowering of freezing point and related phenomena. *Phys. Z.* 24, 185-206.

92. Northrup, S. H. & Erickson, H. P. (1992). Kinetics of protein-protein association explained by Brownian dynamics computer simulation. *Proc. Natl. Acad. Sci. USA* 89, 3338-3342.
93. Zhou, H.-X., Wong, K. Y. & Vijayakumar, M. (1997). Design of fast enzymes by optimizing interaction potential in active site. *Proc. Natl. Acad. Sci. USA* 94, 12372-12377.
94. Camacho, C. J., Kimura, S. R., DeLisi, C. & Vajda, S. (2000). Kinetics of desolvation-mediated protein-protein binding. *Biophys. J.* 78, 1094-1105.
95. Schlosshauer, M. & Baker, D. (2004). Realistic protein-protein association rates from a simple diffusional model neglecting long-range interactions, free energy barriers, and landscape ruggedness. *Protein Sci.* 13, 1660-1669.
96. O'Shea, E. K., Klemm, J. D., Kim, P. S. & Alber, T. (1991). X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science* 254, 539-544.
97. 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 (39), 12812-12819.
98. Zitzewitz, J. A., Ibarra-Molero, B., Fishel, D. R., Terry, K. L. & Matthews, C. R. (2000). Performed secondary structure drives the association reaction of GCN4-p1, a model coiled-coil system. *J. Mol. Biol.* 296, 1105-1116.
99. Bonvin, A. M., Vis, H., Breg, J. N., Burgering, M. J., Boelens, R. and Kaptein, R. (1994). Nuclear magnetic resonance solution structure of the Arc repressor using relaxation matrix calculations. *J. Mol. Biol.* 236, 328-341.
100. Breg, J. N., van Opheusden, J. H., Burgering, M. J., Boelens, R. & Kaptein, R. (1990). Structure of Arc repressor in solution: evidence for a family of  $\beta$ -sheet DNA-binding proteins. *Nature* 346, 586-589.
101. Milla, M. E. & Sauer, R. T. (1994). P22 Arc repressor: folding kinetics of a single-domain, dimeric protein. *Biochemistry* 33, 1125-1133.
102. Milla, M. E., Brown, B. M., Waldburger, C. D. & Sauer, R. T. (1995). P22 Arc repressor: transition state properties inferred from mutational effects on the rates of protein unfolding and refolding. *Biochemistry* 34, 13914-13919.

103. 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.
104. Gage, M. J., Zak, J. L. & Robinson, A. S. (2005). Three amino acids that are critical to formation and stability of the P22 tailspike trimer. *Protein Sci.* 14, 2333-2343.
105. Jaenicke, R. & Rudolf, R. (1980). *Protein Folding: 525-548*, Elsevier-North Holland, Amsterdam (R. Jaenicke, eds).
106. Schwarz, J. J. & Berget, P. B. (1989). Characterization of bacteriophage P22 tailspike mutant proteins with altered endorhamnosidase and capsid assembly activities. *J. Biol. Chem.* 264, 20112-20119.
107. Baechinger, H. P. & Engel, J. (2005). Thermodynamics and kinetics of collagen folding. *Handbook of Protein Folding, Part 1: 1059-1110*, Wiley-VCH (Buchner, J. & Kiefhaber, T. eds).
108. Tao, Y., Strelkov, S. V., Mesyanzhinov, V. V. & Rossmann, M. G. (1997). Structure of bacteriophage T4 fibrin: a segmented coiled coil and the role of the C-terminal domain. *Structure* 5, 789-798.
109. Letarov, A. V., Londer, Y. Y., Boudko, S. P. & Mesyanzhinov, V. V. (1999). The carboxy-terminal domain initiates trimerization of bacteriophage T4 fibrin. *Biochemistry (Mosc.)* 64, 817-823.
110. 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 fibrin, a segmented coiled-coil protein. *Eur. J. Biochem.* 269, 833-841.
111. 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 fibrin. *J. Mol. Biol.* 337, 905-915.
112. Habazettl, J., Reiner, A. & Kiefhaber, T. (2009). NMR structure of a monomeric intermediate on the evolutionarily optimized assembly pathway of a small trimerization domain. *J. Mol. Biol.* 389, 103-114.
113. Zhou, H.-X. (1993). Brownian dynamics study of the influences of electrostatic interaction and diffusion on protein-protein association kinetics. *Biophys. J.* 64, 1711-1726.

114. Zhou, H.-X. (1997). Enhancement of protein-protein association rate by interaction potential: accuracy of prediction based on local Boltzmann factor. *Biophys. J.* 73, 2441-2445.
115. Kiel, C., Selzer, T., Shaul, Y., Schreiber, G. & Herrmann, C. (2004). Electrostatically optimized Ras-binding Ral guanine dissociation stimulator mutants increase the rate of association by stabilizing the encounter complex. *Proc. Natl. Acad. Sci. USA* 101, 9223-9228.
116. Korennykh, A. V., Piccirilli, J. A. & Correll, C. C. (2006). The electrostatic character of the ribosomal surface enables extraordinarily rapid target location by ribotoxins. *Nat. Struct. Mol. Biol.* 13, 436-443.
117. Park, C. & Raines, R. T. (2001). Quantitative analysis of the effect of salt concentration on enzymatic catalysis. *J. Am. Chem. Soc.* 123, 11472-11479.
118. Wendt, H., Leder, L., Härmä, H., Jelesarov, I., Baici, A. & Bosshard, H. R. (1997). Very rapid, ionic strength-dependent association and folding of a heterodimeric leucine zipper. *Biochemistry* 36, 204-213.
119. Schreiber, G. & Fersht, A. R. (1996). Rapid, electrostatically assisted association of proteins. *Nat. Struct. Biol.* 3 (5), 427-431.
120. Selzer, T. & Schreiber, G. (1999). Predicting the rate enhancement of protein complex formation from the electrostatic energy of interaction. *J. Mol. Biol.* 287, 409-419.
121. Selzer, T., Albeck, S. & Schreiber, G. (2000). Rational design of faster associating and tighter binding protein complexes. *Nat. Struct. Biol.* 7, 537-541.
122. Solc, K. & Stockmayer, W. H. (1971). Kinetics of diffusion-controlled reaction between chemically asymmetric molecules. I. General theory. *J. Chem. Phys.* 54, 2981-2988.
123. Solc, K. & Stockmayer, W. H. (1973). Kinetics of diffusion-controlled reaction between chemically asymmetric molecules. II. Approximate steady-state solution. *Int. J. Chem. Kinet.* 5, 733-752.
124. Schmitz, K. S. & Schurr, J. M. (1972). Role of orientation constraints and rotational diffusion in bimolecular solution kinetics. *J. Phys. Chem.* 76, 534-545.

125. Schurr, J. M. & Schmitz, K. S. (1976). Orientation constraints and rotational diffusion in bimolecular solution kinetics. A simplification. *J. Phys. Chem* 80, 1934-1936.
126. Shoup, D., Lipari, G. Szabo, A. (1981). Diffusion-controlled bimolecular reaction rates. The effect of rotational diffusion and orientation constraints. *Biophys. J.* 36, 697-714.
127. Schlosshauer, M. & Baker, D. (2002). A general expression for bimolecular association rates with orientational constraints. *J. Phys. Chem. B* 106, 12079-12083.
128. Schreiber, G., Haran, G. & Zhou, H.-X. (2009). Fundamental aspects of protein-protein association kinetics. *Chem. Rev.* 109, 839-860.
129. Northrup, S. H., Allison, S. A. & McCammon, J. A. (1984). Brownian dynamics simulation of diffusion-influenced bimolecular reactions. *J. Chem. Phys.* 80, 1517-1524.
130. Zhou, H.-X. (1990). Kinetics of diffusion-influenced reactions studied by Brownian dynamics. *J. Phys. Chem.* 94, 8794-8800.
131. Zhou, H.-X. & Szabo, A. (1996). Theory and simulation of the time-dependent rate coefficients of diffusion-influenced reactions. *Biophys. J.* 71, 2440-2457.
132. Gabdouliline, R. R. & Wade, R. C. (1997). Simulation of the diffusional association of barnase and barstar. *Biophys J.* 72, 1917-1929.
133. Alsallaq, R. & Zhou, H.-X. (2008). Electrostatic rate enhancement and transient complex of protein-protein association. *Proteins* 71, 320-335.
134. Camacho, C. J., Weng, Z., Vajda, S. & DeLisi, C. (1999). Free energy landscapes of encounter complexes in protein-protein association. *Biophys. J.* 76, 1166-1178.
135. Zhou, H.-X. (2001). Disparate ionic-strength dependence of on and off rates in protein-protein association. *Biopolymers* 59, 427-433.
136. Huang, X., Dong, F. & Zhou, H.-X. (2005). Electrostatic recognition and induced fit in the  $\kappa$ -PVIIA toxin binding to Shaker potassium channel. *J. Am. Chem. Soc.* 127, 6836-6849.

137. Zhou, H.-X. (2010). From induced fit to conformational selection: a continuum of binding mechanism controlled by the timescale of conformational transitions. *Biophys. J.* 98, L15-L17.
138. Zhou, H.-X., Rivas, G. & Minton, A. P. (2008). Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys.* 37, 375-397.
139. Berg, H. C. (1983). *Random walks in biology*. Princeton University Press.
140. Krieger, F. (2004). *Dynamics in unfolded polypeptide chains as model for elementary steps in protein folding*. PhD thesis, Universität Basel.
141. Dix, J. A. & Verkman, A. S. (2008). Crowding effects on diffusion in solution and cells. *Annu. Rev. Biophys.* 37, 247-263.
142. Swaminathan, R., Hoang, C. P. & Verkman, A. S. (1997). Photobleaching recovery and anisotropy decay of green fluorescent protein GFP-S65T in solution and cells: cytoplasmic viscosity probed by green fluorescent protein translational and rotational diffusion. *Biophys. J.* 72, 1900-1907.
143. Haggie, P. M. & Verkman, A. S. (2002). Diffusion of tricarboxylic acid cycle enzymes in the mitochondrial matrix in vivo. Evidence for restricted mobility of a multienzyme complex. *J. Biol. Chem.* 277, 40782-40788.
144. Odijk, T. (2000). Depletion theory of protein transport in semi-dilute polymer solutions. *Biophys. J.* 79, 2314-2321.
145. Banks, D. S. & Fradin, C. (2005). Anomalous diffusion of proteins due to molecular crowding. *Biophys. J.* 89, 2960-2971.
146. Kuttner, Y. Y., Kozar, N., Segal, E., Schreiber, G. & Haran, G. (2005). Separating the contribution of translational and rotational diffusion to protein association. *J. Am. Chem. Soc.* 127, 15138-15144.
147. Minton A. P. (1981). Excluded volume as a determinant of macromolecular structure and reactivity. *Biopolymers* 20, 2093-2120.
148. Minton AP (2000). Effect of a concentrated “inert” macromolecular cosolute on the stability of a globular protein with respect to denaturation by heat and by chaotropes: a statistical thermodynamic model. *Biophys. J.* 78, 101-109.
149. Asakura, S. & Oosawa, F. (1954). On Interaction between two bodies immersed in a solution of macromolecules. *J. Chem. Phys.* 22, 1255-1256.

150. Vrij, A. (1976). Polymers at interfaces and the interactions in colloidal dispersions. *Pure and Appl. Chem.* 48, 471-483.
151. Parsegian, V. A., Rand, R. P. & Rau, D. C. (2000). Osmotic stress, crowding, preferential hydration, and binding: A comparison of perspectives. *Proc. Natl. Acad. Sci.* 97 (8), 3987-3992.
152. Zimmermann, S. & Minton A. (1993), Macromolecular crowding: biochemical, biophysical, and physiological consequences. *Annu. Rev. Biophys. Biomol. Struct.* 22, 27-65.
153. Rivas, G., Fernandez, J. A. & Minton, A. P (2001). Direct observation of the enhancement of noncooperative protein self-assembly by macromolecular crowding: Indefinite linear self-association of bacterial cell division protein FtsZ. *Proc. Natl. Acad. Sci. USA* 98, 3150-3155.
154. Rivas, G. & Minton, A. P. (2004). Non-ideal tracer sedimentation equilibrium: a powerful tool for the characterization of macromolecular interactions in crowded solutions. *J. Mol. Recognit.* 17, 362-367.
155. Mukherjee, S., Waagele, M. M., Chowdhury, P., Guo, L. & Gai, F. (2009). Effect of macromolecular crowding on protein folding dynamics at the secondary structure level. *J. Mol. Biol.* 393, 227-236.
156. Zhou, H.-X. (2008). Protein folding in confined and crowded environments. *Arch. Biochem. Biophys.* 469, 76–82.
157. Van den Berg, B., Wain, R., Dobson, C. M. & Ellis, R. J. (2000). Macromolecular crowding perturbs protein refolding kinetics: implications for folding inside the cell. *EMBO J.* 18, 6927–6933.
158. Qu, Y. X. & Bolen, D. W. (2002). Efficacy of macromolecular crowding in forcing proteins to fold. *Biophys. Chem.* 101, 155–165.
159. Tokuriki, N., Kinjo, M., Negi, S., Hoshino, M., Goto, Y., Urabe, I. & Yomo, T. (2004). Protein folding by the effects of macromolecular crowding. *Protein Sci.* 13, 113–125.
160. Ai, X. J., Zhou, Z., Bai, Y. W. & Choy, W.-Y. (2006). N-15 NMR spin relaxation dispersion study of the molecular crowding effects on protein folding under native condition. *J. Am. Chem. Soc.* 128, 3916–3917.
161. Yuan, J. M., Chyan, C.-L., Zhou, H.-X., Chung, T.-Y., Peng, H. B., Ping, G. H. & Yang, G. L. (2008). The effects of macromolecular crowding on the mechanical stability of protein molecules. *Protein Sci.* 17, 2156–2166.

162. Charlton, L. M., Barnes, C. O., Li, C. G., Orans, J., Young, G. B. & Pielak, G. J. (2008). Residue-level interrogation of macromolecular crowding effects on protein stability. *J. Am. Chem. Soc.* 130, 6826–6830.
163. Homouz, D., Stagg, L., Wittung-Stafshede, P. & Cheung, M. S. (2009). Macromolecular crowding modulates folding mechanism of alpha/beta protein apoflavodoxin. *Biophys. J.* 96, 671–680.
164. Ladurner, A. G. & Fersht, A. R. (1999). Upper limit of the time scale for diffusion and chain collapse in chymotrypsin inhibitor 2. *Nat. Struct. Biol.* 6, 28-31.
165. Kozer, N. & Schreiber, G. (2004). Effect of crowding on protein-protein association rates: fundamental differences between low and high mass crowding agents. *J. Mol. Biol.* 336 (3), 763-774.
166. Schlarb-Ridley, B. G., Mi, H., Teale, W. D., Meyer, V. S., Howe, C. J. & Bendall, D. S. (2005). Implications of the effects of viscosity, macromolecular crowding, and temperature for the transient interaction between cytochrome f and plastocyanin from the cyanobacterium *Phormidium laminosum*. *Biochemistry* 44, 6232-6238.
167. Yael, P., Sherman, E., Haran, G. & Schreiber, G. (2009). Common crowding agents have only a small effect on protein-protein interactions. *Biophysical Journal* 97, 875-885.
168. Bullock, A. T., Cameron, G. G. & Smith, P. M. (1974). Electron spin resonance studies of spin-labelled polymers. *J. Chem. Soc., Faraday Trans. 2* 70, 1202-1210.
169. Mashimo, S., Iwasa, Y., Chiba, A. & Shinohara, K. (1975). Dielectric study on globule-coil transition in polymer solution. *J. Phys. Soc. Jpn.* 39, 1529-1537.
170. Mashimo, S. (1976). Dielectric study of chain motion of poly(*p*-chlorostyrene) in dilute solution. *Macromolecules* 9, 91-97.
171. Gavish, B. (1978). The role of geometry and elastic strains in dynamic states of proteins. *Biophys. Struct. Mech.* 4, 37-52.
172. Gavish, B. & Werber, M. M. (1979). Viscosity-dependent structural fluctuations in enzyme catalysis. *Biochemistry* 18, 1269-1275.
173. Krasilnikov O. V., Sabirov, R. Z., Ternovsky, V. I., Merzliak, P. G. and Muratkhodjaev, J. N. (1992). A simple method for the determination of the

- pore radius of ion channels in planar lipid bilayer membranes. *FEMS Microbiology Letters* 105, 93-100.
174. Barshtein, G., Almagor, A., Yedgar, S. & Gavish, B. (1995). Inhomogeneity of viscous aqueous solutions. *Phys. Rev.* 52, 555-557.
175. Lavalette, D. Tetreau, C., Tourbez, M. & Blouquit Y. (1999). Microscopic viscosity and rotational diffusion of proteins in a macromolecular environment. *Biophys. J.* 76, 2744-2751.
176. Yedgar, S., Tetreau, B., Gavish, B. & Lavalette, D. (1995). Viscosity dependence of O<sub>2</sub> escape from respiratory proteins as a function of cosolvent molecular weight. *Biophys. J.* 68, 665-670.
177. Miller, W. G., Brant, D. A. & Flory, P. J. (1967). Random coil configurations of polypeptide chains. *J. Mol. Biol.* 23, 67-80.
178. Kumar, V. D., Lee, L. & Edwards, B. F. (1990). Refined crystal structure of calcium-liganded carp parvalbumin 4.25 at 1.5-Å resolutions. *Biochemistry* 29, 1404-1412.
179. Wilton, D. J., Tunnicliffe, R. B., Kamatari, Y. Q., Akasaka, K. & Williamson, M. P. (2008). Pressure-induced changes in the solution structure of the GB1 domain of protein G. *Proteins* 71, 1432-1440.
180. Cordier, F., Hartmann, B., Rogowski, M., Affolter, M. & Grzesiek, S. (2006). Solution structure of the Brinker DNA binding domain in complex with the omb enhancer. *J. Mol. Biol.* 361, 659-672.
181. Richardson, J. S. (1981). The anatomy and taxonomy of protein structure. *Adv. in Prot. Chem.* 34, 167-339.
182. Rose, G. D., Gierasch, L. M. & Smith, J. A. (1985). Turns in peptides and proteins. *Adv. in Prot. Chem.* 37, 1-109.
183. Qui, L. & Hagen, S. J. (2004). A limiting speed for protein folding a low solvent viscosity. *J. Am. Chem. Soc.* 126, 3398-3399.
184. Pabit, S. A., Roder, H. & Hagen, S. J. (2004). Internal friction controls the speed of protein folding from a compact configuration. *Biochemistry* 43, 12532-12538.
185. Gunasekaran, K., Gomathi, L., Ramakrishnan, C., Chandrasekhar, J. & Balaram, P. (1998). Conformational interconversions in peptide  $\beta$ -turns:

- analysis of turns in proteins and computational estimates of barriers. *J. Mol. Biol.* 284, 1505-1516.
186. Hagen, S. J. & Eaton, W. A. (2000). Two-state expansion and collapse of a polypeptide. *J. Mol. Biol.* 301, 1019-1027.
  187. Cordier, F., Hartmann, B., Rogowski, M., Affolter, M. & Grzesiek, S. (2006). DNA recognition by the brinker repressor – an extreme case of coupling between binding and folding. *J. Mol. Biol.* 361, 659-672.
  188. Bachmann, A., Wildemann, D., Praetorius, F., Fischer, G. & Kiefhaber, T. (2011). Mapping backbone and side-chain interactions in the transition state of a coupled protein folding and binding reaction. *Proc. Natl. Acad. Sci.* 108, 3952-3957.
  189. 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.
  190. O’Nuallain, B., Thakur, A. K., Williams, A. D., Bhattacharyya, A. M., Chen, S., Thiagarajan, G. & Wetzel, R. (2006). Kinetics and thermodynamics of amyloid assembly using a high-performance liquid chromatography-based sedimentation assay. *Methods in Enzymol.* 413, 34-74.
  191. Dyson, H. J. & Wright, P. E. (2004). Unfolded proteins and protein folding studied by NMR. *Chem. Rev.* 104, 3607-3622.
  192. 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.
  193. Religa, T. L., Markson, J. S., Mayor, U., Freund, S. M. V. & Fersht, A. R. (2005). Solution structure of a protein denatured state and folding intermediate. *Nature* 437, 1053–1056.
  194. Neri, D., Billeter, M., Wider, G. & Wüthrich, K. (1992). NMR determination of residual structure in a urea-denatured protein, the 434-repressor. *Science* 257 (5076), 1559-1563.
  195. Logan, T. M., Theriault, Y. & Fesik, S. W. (1994). Structural characterization of the FK506 binding protein unfolded in urea and guanidine hydrochloride. *J. Mol. Biol.* 236, 637–648.

196. Yi, Q., Scalley-Kim, M. L., Alm, E. J. & Baker, D. (2000). NMR characterization of residual structure in the denatured state of protein L. *J. Mol. Biol.* 299, 1341–1351.
197. Kortemme, T., Kelly, M. J., Kay, L. E., Forman-Kay, J. D. & Serrano, L. (2000). Similarities between the spectrin SH3 domain denatured state and its folding transition state. *J. Mol. Biol.* 297, 1217–1229.
198. Wirmer, J., Schlörb, C. & Schwalbe, H. (2005). in *Protein Folding Handbook*, eds. Buchner, J. & Kiefhaber, T. (Wiley-VCH, Weinheim, Germany), Vol. 2, pp. 737–808.
199. Sanchez, 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.
200. Gong, H., Fleming, P. J. & Rose, G. D. (2005). Building native protein conformation from highly approximate backbone torsion angles. *Proc. Natl. Acad. Sci. USA* 102, 16227–16232.
201. Möglich, A., Joder, K. & Kiefhaber T. (2006). End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation. *Proc. Natl. Acad. Sci. USA* 103, 12394–12399.
202. Joder, K. (2005). *Dynamics of intermolecular interactions measured by triplet-triplet energy transfer experiments*. Master thesis, Universität Basel.
203. Perl, D., Jacob, M., Bano, M., Stupak, M., Antalik, M. & Schmid, F. X. (2002). Thermodynamics of a diffusional protein folding reaction. *Biophysical Chemistry* 96, 173–190.
204. Baxter, N. J., & Williamson, M. P. (1997). Temperature dependence of  $^1\text{H}$  chemical shifts in proteins. *J. Biomol. NMR* 9, 359–369.
205. Perkins, S. J. & Wüthrich, K. (1979). Ring current effects in the conformation dependent NMR chemical shifts of aliphatic protons in the basic pancreatic trypsin inhibitor. *Biochim. et Biophys. Acta* 576, 409–423.
206. Bundi, A. & Wüthrich, K. (1979).  $^1\text{H}$  NMR parameters of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH. *Biopolymers* 18, 285–297.

207. Gross, K.-H., & Kalbitzer, H. R. (1988). Distribution of chemical shifts in  $^1\text{H}$  nuclear magnetic resonance spectra of proteins. *J. Magn. Res.* 76, 87-99.
208. Meier, S., Güthe, S., Kiefhaber, T. & Grzesiek, S. (2004). Foldon, the natural trimerization domain of T4 fibrin, 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.
209. Fierz, B. & Kiefhaber, T., in *Protein Folding Handbook*, edited by J. Buchner and T. Kiefhaber (WILEY-VCH, Weinheim, 2005), 805-851.
210. von Smoluchowski, M. (1906). Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen. *Ann. d. Phys.* 21, 756-780.
211. von Smoluchowski, M. (1916). Drei Vorträge über Diffusion, Brownsche Molekularbewegung und Koagulation von Kolloidteilchen. *Phys. Z.* 17, 557-571.
212. Graham, R. & Lewis, J. R. (1978). Synthesis of 9-Oxoxanthen-2-carboxylic acids. *J. Chem. Soc., Perkin Trans. 1*, 876-881.
213. Luo, P. & Baldwin, R. L. (1997). Mechanism of helix induction by trifluoroethanol: A framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water. *Biochemistry* 36, 8413-8421.
214. Raiford, D. S., Fisk, C. L. & Becker, E. D. (1979). Calibration of methanol and ethylene glycol nuclear magnetic resonance thermometers. *Anal. Chem.* 51, 2050-2051.
215. Bax, A., & Davis, D. G. (1985A). Practical aspects of two-dimensional transverse NOE spectroscopy. *J. Magn. Reson.* 63, 207-213.
216. Piotto, M., Saudek, V., and Sklenar, V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* 2, 661-665.
217. Bax, A., & Davis, D. G. (1985). MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson.* 65, 355-360.
218. Goddard, T. D., & Kneller, D. G. SPARKY 3, University of California, San Francisco.

219. Johnson, K. A., Simpson, Z. B. & Blom T. (2009). Global Kinetic Explorer: A new computer program for dynamic simulation and fitting of kinetic data. *Analytical Biochemistry* 387, 20-29.
220. Johnson, K. A., Simpson, Z. B. & Blom T. (2009). FitSpace Explorer: An algorithm to evaluate multidimensional parameter space in fitting kinetic data. *Analytical Biochemistry* 387, 30-41.

## 7. Appendix

7.1 End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation.....	177
--	-----

# End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation

Andreas Möglich\*, Karin Joder, and Thomas Kiefhaber†

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

Edited by Robert L. Baldwin, Stanford University, Stanford, CA, and approved June 28, 2006 (received for review June 7, 2006)

**Characterization of the unfolded state is essential for the understanding of the protein folding reaction. We performed time-resolved FRET measurements to gain information on the dimensions and the internal dynamics of unfolded polypeptide chains. Using an approach based on global analysis of data obtained from two different donor–acceptor pairs allowed for the determination of distance distribution functions and diffusion constants between the chromophores. Results on a polypeptide chain consisting of 16 Gly-Ser repeats between the FRET chromophores reveal an increase in the average end-to-end distance from 18.9 to 39.2 Å between 0 and 8 M GdmCl. The increase in chain dimensions is accompanied by an increase in the end-to-end diffusion constant from  $(3.6 \pm 1.0) \times 10^{-7} \text{ cm}^2 \text{ s}^{-1}$  in water to  $(14.8 \pm 2.5) \times 10^{-7} \text{ cm}^2 \text{ s}^{-1}$  in 8 M GdmCl. This finding suggests that intrachain interactions in water exist even in very flexible chains lacking hydrophobic groups, which indicates intramolecular hydrogen bond formation. The interactions are broken upon denaturant binding, which leads to increased chain flexibility and longer average end-to-end distances. This finding implies that rapid collapse of polypeptide chains during refolding of denaturant-unfolded proteins is an intrinsic property of polypeptide chains and can, at least in part, be ascribed to nonspecific intramolecular hydrogen bonding. Despite decreased intrachain diffusion constants, the conformational search is accelerated in the collapsed state because of shorter diffusion distances. The measured distance distribution functions and diffusion constants in combination with Szabo–Schulten–Schulten theory were able to reproduce experimentally determined rate constants for end-to-end loop formation.**

chain collapse | chain dynamics | denaturant | FRET

Understanding the properties of the unfolded state is a major focus of protein folding studies. The rather harsh and non-physiological conditions required for populating unfolded states in equilibrium pose a major problem in the characterization of unfolded proteins under physiological conditions. For many proteins, rapid chain compaction in combination with major changes in spectroscopic properties was observed in refolding experiments starting from the denaturant-unfolded state (1–5). It has been discussed whether these rapidly formed compact structures represent the unfolded state under physiological conditions or whether the fast processes are due to the formation of specific folding intermediates separated from the unfolded state by energy barriers (3, 5, 6). Detailed structural information on unfolded proteins under physiological conditions recently became available from studies on natively unfolded proteins (7) and on engineered unstable protein variants (8, 9). For several proteins, compact unfolded states with both specific and nonspecific interactions were directly observed by NMR measurements (10–14). Analysis of the effect of mutations on solvent accessibility changes during folding also revealed the presence of residual structure in unfolded proteins that can be disrupted by side-chain mutations (15). These results from various independent experimental techniques indicate that significant parts of the native structure may already be present in unfolded proteins. This conclusion was supported by theoretical

work, which suggested the presence of native-like secondary structure and a large amount of intramolecular hydrogen bonds in unfolded proteins in water (16, 17). Although our understanding of the structural properties of unfolded proteins has recently increased, the contributions from specific interactions, typical for each individual protein sequence, versus contributions from non-specific side-chain and backbone interactions as a result of general polymer properties of polypeptide chains are still not well understood. Furthermore, only little is known about the dynamic properties of unfolded and partially folded proteins under different solvent conditions.

To gain insight into the structure and dynamics of unfolded proteins, we have recently studied the kinetics of intramolecular loop formation in model polypeptide chains with repetitive amino acid sequences and in sequences derived from natural proteins (18–21). Using intramolecular triplet–triplet energy transfer (TTET) between a donor and an acceptor group, we observed a decreasing rate constant for loop formation,  $k_c$ , with increasing denaturant concentration (22). As for protein folding reactions, a linear relationship between  $\ln k_c$  and the denaturant concentration was found for all investigated polypeptide chains. The effect of guanidinium chloride (GdmCl) and urea on  $k_c$  could be dissected into two components. Part of the effect is due to changes in solvent viscosity. The remaining decrease in  $k_c$  arises from binding of denaturant molecules to the polypeptide chain and could be quantitatively described by Schellman's weak-binding model (22–24). Because these studies revealed that the properties of unfolded polypeptide chains are drastically changing upon addition of denaturants, we wanted to obtain more detailed information on the structure and dynamics of unfolded chains in different solvents by performing time-resolved (tr) FRET experiments on a long poly(Gly-Ser) chain containing donor and acceptor groups near the ends. These measurements allow for the characterization of general properties of unfolded polypeptide chains in the absence of specific side-chain interactions. Global analysis of results obtained from two different donor–acceptor pairs enabled us to determine end-to-end distance distributions and intrachain diffusion constants between the FRET labels in water and to characterize the effect of denaturants.

## Results

**FRET Measurements on Flexible Polypeptide Chains.** We have previously used poly(Gly-Ser) chains of different length as a model to investigate the kinetics of loop formation in unfolded polypeptide

Conflict of interest statement: No conflicts declared.

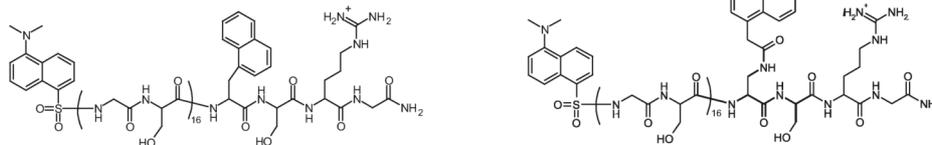
This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: SSS, Szabo–Schulten–Schulten; trFRET, time-resolved FRET; TTET, triplet–triplet energy transfer.

\*Present address: Department of Biochemistry and Molecular Biology, University of Chicago, Center for Integrative Science, W107A, 929 East 57th Street, Chicago, IL 60637.

†To whom correspondence should be addressed. E-mail: t.kiefhaber@unibas.ch.

© 2006 by The National Academy of Sciences of the USA



**Structure 1.** Structure of the (Gly-Ser)<sub>16</sub> peptide labeled with naphthalene/dansyl (Left) and pyrene/dansyl (Right).

chains (18, 19, 22). To obtain information on the dimensions and internal dynamics of these chains, we performed FRET measurements. The rate constant,  $k_T$ , for FRET between a donor (D) and an acceptor (A) group depends on the properties of the chromophores and on the distance,  $r$ , separating them according to (25)

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

with

$$R_0^6 = \frac{9,000 \ln 10 \kappa^2 \Phi_D}{128 \pi^5 N n^4} \int_0^\infty F_D(\lambda) \epsilon_A(\lambda) \lambda^4 d\lambda. \quad [1]$$

$\tau_D$  denotes the intrinsic fluorescence lifetime of the donor, and  $R_0$  is the characteristic Förster distance at which a given FRET pair shows 50% transfer efficiency.  $N$  is Avogadro's number,  $n$  is the refractive index of the medium, and  $\kappa^2$  is an orientational term assumed to equal 2/3 for rapidly reorienting fluorophores.  $\Phi_D$  is the fluorescence quantum yield of the donor,  $F_D(\lambda)$  is the normalized emission of the donor, and  $\epsilon_A(\lambda)$  is the extinction coefficient of the acceptor at wavelength  $\lambda$ . The basic FRET theory has to be extended when unfolded polypeptide chains are studied that represent rapidly interconverting ensembles of different conformations with a distribution of interchromophore distances,  $p(r)$ , which leads to a distribution of energy transfer efficiencies and to changes in the donor-acceptor distances of individual molecules during the donor lifetime. Haas *et al.* (26) developed a formalism that treats the effect of conformational heterogeneity and internal dynamics on the population of excited donor molecules in trFRET experiments.

$$\frac{\partial \bar{p}(r, t)}{\partial t} = -\frac{1}{\tau_D} \left\{ \left( 1 + \left( \frac{R_0}{r} \right)^6 \right) \bar{p}(r, t) \right\} + \frac{1}{p_0(r)} \frac{\partial}{\partial r} \left( p_0(r) D_{DA} \frac{\partial \bar{p}(r, t)}{\partial r} \right) \quad [2]$$

with

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

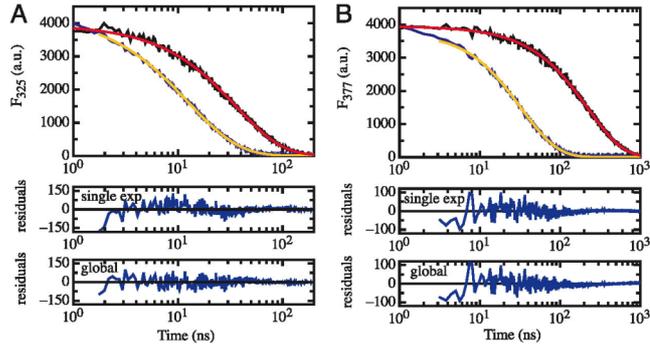
The first part on the right-hand side in Eq. 2 represents distance-dependent FRET where  $\bar{p}(r, t)$  is the time-dependent probability to find an excited donor molecule with an acceptor molecule at distance  $r$  and  $p_0$  is the equilibrium distribution of donor-acceptor distances. The second part on the right-hand side represents the effect of diffusion on FRET with  $D_{DA}$  denoting the diffusion constant for motions of the two fluorophores relative to each other.

Eq. 2 shows that FRET measurements on flexible chains with dynamics on the time scale of the donor lifetime provide the chance to obtain information on both the dimensions and the dynamics of a flexible polymer chain. However,  $p(r)$  and  $D_{DA}$  cannot be

determined from a single FRET experiment because both parameters contribute to the observed FRET efficiency. Several solutions have been proposed to determine both  $p(r)$  and  $D_{DA}$  in highly dynamic and flexible systems. Haas *et al.* (26) and Lakowicz *et al.* (27) measured FRET in solutions of different solvent viscosity to determine  $D_{DA}$  in peptides of different length. Lakowicz *et al.* (28) further performed FRET measurements in the presence of different amounts of fluorescence quenchers and globally analyzed the data to obtain  $p(r)$ . We took a different approach to determining both  $p(r)$  and  $D_{DA}$  in unfolded polypeptide chains by using two different donor-acceptor pairs with similar  $R_0$  values but with different donor lifetimes. This allows for a reliable determination of  $p(r)$  and  $D_{DA}$  in a global analysis of the data because the diffusional term has a stronger influence on the observed transfer efficiency with increasing donor lifetime.

**Spectroscopic Properties of the Fluorophore Systems.** To study the dimensions and dynamics in unfolded model polypeptide chains by global FRET analysis, we synthesized two different peptides, each containing 16 pairs of Gly-Ser between donor and acceptor. Both peptides have the dansyl group as acceptor at the N terminus and either naphthalene or pyrene as donor group introduced C-terminal from the (Gly-Ser)<sub>16</sub> stretch (Structure 1). Naphthalene-dansyl and pyrene-dansyl are two well suited donor-acceptor pairs for global FRET analysis because naphthalene and pyrene attached to a poly(Gly-Ser) chain have largely different fluorescence lifetimes in water of (36.9 ± 0.2) ns and (225.5 ± 1.0) ns, respectively (see Tables 1 and 2, which are published as supporting information on the PNAS web site). The  $R_0$  values of the two pairs in water are, however, very similar with (23.3 ± 0.4) Å and (20.5 ± 0.4) Å for naphthalene-dansyl and pyrene-dansyl, respectively, which makes the two pairs sensitive for the same range of donor-acceptor distances. Because we wanted to investigate the effect of denaturants on the chain properties, we additionally determined  $\tau_D$  and  $R_0$  of the FRET pairs between 0 and 8 M GdmCl. The results show that the two FRET pairs have similar  $R_0$  values but largely differ in their donor fluorescence lifetimes at all GdmCl concentrations (see Tables 1 and 2).

**Global Analysis of trFRET Measurements.** Fig. 1 shows fluorescence decay curves of naphthalene (Fig. 1A) and pyrene (Fig. 1B) attached to the ends of a (Gly-Ser)<sub>16</sub> sequence. For both donors, the fluorescence lifetime is decreased in the presence of the dansyl acceptor group indicating energy transfer. The decrease in lifetime is much more pronounced for pyrene, with a 6.8-fold decrease in fluorescence half-life versus a 2.7-fold decrease for naphthalene, although the pyrene-dansyl pair has a slightly smaller  $R_0$  value than the naphthalene-dansyl pair, and thus less energy transfer would be expected for the pyrene-dansyl pair. This observation indicates significant contributions from diffusional processes to the observed FRET efficiency (see Eq. 2) and is in agreement with our earlier results from TTET experiments that revealed end-to-end contact formation on the 50-ns time scale in a (Gly-Ser)<sub>16</sub> peptide labeled with xanthone and naphthalene at the ends (19, 22), which is on a time scale similar to the donor lifetimes.



**Fig. 1.** Fluorescence decay curves for the different donor-only and donor-acceptor peptides in water. (A) Naphthalene fluorescence decay curves in a naphthalene-only peptide (black) and the (Gly-Ser)<sub>16</sub> peptide labeled with naphthalene and dansyl (blue). (B) Fluorescence decay curves of the pyrene-only (black) and pyrene-dansyl (Gly-Ser)<sub>16</sub> (blue) peptide. The fluorescence decay curves of the donor-only peptides can be fitted with single-exponential functions with lifetimes ( $\tau_D$ ) of (36.9 ± 0.2) ns (naphthalene) and (225.5 ± 1.0) ns (pyrene). The decay curves of the donors in the presence of the acceptors were fitted globally (red curves) to Eqs. 2 and 3. The residuals show the results from single-exponential and global fits to the donor fluorescence decay curves in the presence of acceptor.

Fluorescence decay curves for both donors in the presence of acceptor are not single-exponential (Fig. 1). A fit of the individual FRET data according to Eq. 2 did not allow the determination of  $p(r)$  and  $D_{DA}$  because the observed fluorescence decay curve can be fitted by either close donor-acceptor distances or by rapid diffusion. Only global analysis of the fluorescence decay curves was able to resolve this ambiguity. The analysis used Eq. 2 and global values for  $p(r)$  and  $D_{DA}$  in the two peptides, which seems justified because the chromophores are small compared with the chain dimensions. A skewed Gaussian distribution (Eq. 3) was assumed for  $p(r)$  (29, 26).

$$p(r) = c\pi r^2 \exp(-a(r-b)^2) \quad [3]$$

The parameter  $a$  determines the width of the distribution, and  $c = f(a, b)$  is a normalization constant. The distribution is offset from zero by  $b$ . TTET experiments on end-to-end loop formation in the same peptide have previously shown kinetics in the Gaussian chain limit both in water and in the presence of high GdmCl concentrations (19), indicating that this approximation is valid. The global fit yielded the donor-acceptor distribution function shown in Fig. 2A and a  $D_{DA}$  value of  $(3.6 \pm 1.0) \times 10^{-7} \text{ cm}^2 \text{ s}^{-1}$  in water.

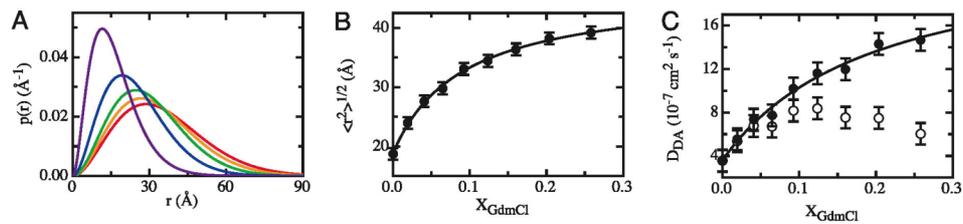
**Effect of GdmCl on Chain Dimensions and Dynamics.** To elucidate the effect of denaturants on the chain properties, we performed FRET measurements in the presence of various GdmCl concentrations. At all GdmCl concentrations, the donor fluorescence showed complex decay curves in the presence of acceptor. With increasing denaturant concentration, the apparent FRET efficiency decreases as indicated by a decreasing effect of the presence of acceptor on the donor lifetime (see Tables 1 and 2). Global analysis of data from the two donor-acceptor pairs revealed that increasing GdmCl concen-

trations lead to an increase in the average donor-acceptor distance and to a broadening of the distribution function (Fig. 2A). To quantify this behavior, we calculated the rms distance between donor and acceptor,  $\langle r^2 \rangle^{1/2}$ , from the  $p(r)$  functions (Fig. 2B). In the absence of denaturant, a  $\langle r^2 \rangle^{1/2}$  value of  $(18.7 \pm 1.0) \text{ \AA}$  is obtained.  $\langle r^2 \rangle^{1/2}$  increases hyperbolically with denaturant concentration and reaches a value of  $39.2 \text{ \AA}$  at 8 M GdmCl. The effect of [GdmCl] on  $\langle r^2 \rangle^{1/2}$  can be quantitatively described by a binding isotherm employing the Schellman model for weak interactions (23, 24), which was also able to describe the effect of denaturants on the rate constant of loop formation (22).

$$\langle r^2 \rangle^{1/2} = \langle r_0^2 \rangle^{1/2} \times \left( 1 + \gamma \frac{(K_{ex} - 1)X_D}{(K_{ex} - 1)X_D + 1} \right) \quad [4]$$

Here,  $\langle r_0^2 \rangle^{1/2}$  denotes  $\langle r^2 \rangle^{1/2}$  in the absence of denaturant and  $K_{ex}$  is an exchange constant on the mole fraction scale ( $X_D$ ) for replacing water bound to the peptide by GdmCl.  $\gamma$  is a factor that reflects the sensitivity of  $\langle r^2 \rangle^{1/2}$  toward denaturant binding. Fitting the experimentally determined  $\langle r^2 \rangle^{1/2}$  values (Fig. 2B) to Eq. 4 yields  $\langle r_0^2 \rangle^{1/2} = (18.9 \pm 0.9) \text{ \AA}$ ,  $K_{ex} = 12.0 \pm 2.7$ , and  $\gamma = 1.46 \pm 0.13$ . This  $K_{ex}$  value is slightly smaller than the value obtained for the effect of GdmCl on the kinetics of loop formation in the same polypeptide chains ( $K_{ex} = 27.4 \pm 1.5$ ) (22).

A more complex behavior is observed for the effect of GdmCl on  $D_{DA}$  (Fig. 2C).  $D_{DA}$  increases from  $(3.6 \pm 1.0) \times 10^{-7} \text{ cm}^2 \text{ s}^{-1}$  at 0 M GdmCl up to a maximum value of  $8.4 \times 10^{-7} \text{ cm}^2 \text{ s}^{-1}$  at 4 M GdmCl ( $X_D = 0.12$ ). At higher GdmCl concentrations, a slight decrease in  $D_{DA}$  is observed. In our previous studies, we showed that the rate constant for end-to-end loop formation in the (Gly-Ser)<sub>16</sub> peptide is inversely proportional to solvent viscosity,  $\eta$  (18, 22). Accordingly, we corrected the values for  $D_{DA}$  at the different



**Fig. 2.** Effect of GdmCl on the dimensions and dynamics of the (Gly-Ser)<sub>16</sub> chain. (A) Effect of GdmCl on the donor-acceptor distance probability distribution function  $p(r)$  in the (Gly-Ser)<sub>16</sub> peptide. Results for GdmCl concentrations of 0 M (violet), 2.0 M (blue), 4.0 M (green), 6.0 M (orange), and 8.0 M (red) are displayed. (B) rms end-to-end distances,  $\langle r^2 \rangle^{1/2}$ , calculated from  $p(r)$  curves at GdmCl concentrations between 0 and 8 M. (C) Effect of GdmCl on the donor-to-acceptor diffusion constants ( $D_{DA}$ ) obtained from the global analysis (open circles) and on the viscosity-corrected values (filled circles) according to Eq. 5. The solid lines in B and C describe fits to the weak-binding model (Eqs. 4 and 5, respectively).

GdmCl concentrations using Eq. 5. The viscosity-corrected values ( $D'_{DA}$ ) increase hyperbolically with denaturant concentration and reach a value of  $(14.7 \pm 1.0) \times 10^{-7} \text{ cm}^2 \text{ s}^{-1}$  at 8 M GdmCl. The effect of GdmCl can also be described by the weak binding model as follows:

$$D'_{DA} = D_{DA}^0 \times \left( 1 + \gamma \frac{(K_{ex} - 1)X_D}{(K_{ex} - 1)X_D + 1} \right) \quad [5]$$

with

$$D'_{DA} = \left( \frac{\eta}{\eta_0} \right) \times D_{DA},$$

where  $D_{DA}^0$  represents the value of  $D_{DA}$  in the absence of GdmCl. The fit of the data with Eq. 5 yields  $D_{DA}^0 = (3.7 \pm 0.9) \times 10^{-7} \text{ cm}^2 \text{ s}^{-1}$ ,  $K_{ex} = 5.9 \pm 4.0$ , and  $\gamma = 5.6 \pm 2.3$ , similar to the exchange constant obtained of the effect of GdmCl on chain dimension.

## Discussion

**Using Multiple Donor–Acceptor Pairs for FRET Measurements in Flexible Systems.** FRET data from a single donor–acceptor pair do not allow a reliable determination of donor–acceptor distances in unfolded or partially folded polypeptide chains due to contributions from intrachain diffusion. Our results show that this ambiguity can be resolved by performing global analysis of trFRET data obtained with two different donor–acceptor pairs. The global analysis allows for the determination of the donor–acceptor distance distribution function  $p(r)$  and additionally yields the intrachain diffusion constant between the donor and acceptor ( $D_{DA}$ ). The prerequisites for this method are (i) the use of two FRET pairs with similar  $R_0$  values but significantly different donor lifetimes and (ii) chain dynamics on the same time scale as the donor lifetimes. Our earlier results on the kinetics of intrachain loop formation in unfolded polypeptide chains showed that major changes in the donor–acceptor distances occur on the 5- to 50-ns time scale, depending on amino acid sequence and loop length (19, 20). This finding indicates that intrachain diffusion has major effects on the observed FRET efficiency if the donor lifetimes are slower than 1–5 ns.

The effect of chain dynamics on FRET efficiency resolves a paradox reported for unfolded BBL protein, for which an apparent chain compaction was observed in FRET experiments upon increasing the temperature from 280 to 360 K (30). Because the viscosity of water decreases from 1.43 to 0.33 cP (1 P = 0.1 Pa·sec) in this temperature range (31) and intrachain diffusion is inversely dependent on solvent viscosity (18, 22), the contributions from intrachain diffusion to FRET increase significantly with increasing temperature. In the FRET studies on BBL protein, naphthylalanine, which has a fluorescence lifetime on the same time scale as chain dynamics (see Fig. 1A), was used as the donor, which leads to increased transfer efficiencies and thus to apparent chain compaction if the dynamic component is neglected.

**Effect of GdmCl on Polypeptide Chain Properties.** Global analysis of FRET data obtained from the two donor–acceptor pairs attached to the ends of a (Gly-Ser)<sub>16</sub> chain allowed us to determine both  $p(r)$  and  $D_{DA}$  as a function of GdmCl concentration. The results show that GdmCl has major effects on the chain dimensions and dynamics. In the presence of 8 M GdmCl, the  $\langle r^2 \rangle^{1/2}$  value is more than twice as large as it is in the absence of denaturant (Fig. 2B). At the same time, the flexibility of the chain increases as indicated by a 4-fold increase in the diffusion constant between 0 and 8 M GdmCl (Fig. 2C). The effect of GdmCl on  $\langle r^2 \rangle^{1/2}$  and  $D_{DA}$  can be described by Schellman's weak-binding model. The observed effects are not caused by changes in peptide secondary structure as indicated by virtually identical CD spectra in the far-UV region at all denaturant concentrations (see Fig. 4, which is published as supporting information on the PNAS web site). In addition, NMR measurements

did not detect any interactions involving the naphthalene moiety (data not shown).

**Chain Dimensions at High Denaturant Concentrations.** We compared the average chain dimensions in the (Gly-Ser)<sub>16</sub> peptide with data determined for GdmCl-unfolded proteins. A study by Damaschun *et al.* (32) measured and compiled small-angle x-ray scattering (SAXS) data on 12 proteins unfolded in 6 M GdmCl. They found that the observed radii of gyration,  $R_g$ , extrapolated to zero protein concentration scale with the number of residues,  $N$ , according to

$$R_g = R_g^0 \times N^\nu, \quad [6]$$

with values of  $R_g^0 = (4.4 \pm 0.3) \text{ \AA}$  and  $\nu = 0.5 \pm 0.02$ . In a recent study, Kohn *et al.* (33) evaluated SAXS data from 28 proteins unfolded at various protein concentrations and different concentrations of GdmCl. They found values of  $R_g^0 = (2.08 \pm 0.19) \text{ \AA}$  and  $\nu = 0.598 \pm 0.028$ . Assuming a Gaussian chain,  $R_g$  can be converted to  $\langle r^2 \rangle$  with

$$R_g^2 = 1/6 \times \langle r^2 \rangle. \quad [7]$$

With the parameters reported by Damaschun *et al.* (32), this results in  $\langle r^2 \rangle^{1/2} = 61.9 \text{ \AA}$  for a GdmCl-unfolded 33-residue protein, which corresponds to separation of donor and acceptor in our model peptides. This value is significantly larger than the  $\langle r^2 \rangle^{1/2}$  value of  $(36.4 \pm 1.0) \text{ \AA}$  found for the (Gly-Ser)<sub>16</sub> peptide in the presence of 6 M GdmCl. Shorter average end-to-end distances in poly(Gly-Ser) chains are expected because of the high content of glycine residues, which leads to shorter chains compared with natural protein sequences (34, 35). From the data of Kohn *et al.* (33), a  $\langle r^2 \rangle^{1/2}$  value of 41.2 Å is obtained for a 33-residue protein, which seems to be too low when considering the significantly smaller fraction of glycine residues in natural proteins.

We further compared our experimental data with predictions from polymer theory. The use of parameters for the dimensions of polypeptide chains reported by Flory and coworkers (35, 36) yields  $\langle r^2 \rangle^{1/2} = 37.6 \text{ \AA}$  for a (Gly-Ser)<sub>16</sub> chain when the properties of serine are approximated by alanine (36, 37). Although the validity of the underlying “isolated-pair hypothesis” has recently been shown to be oversimplified (16), this number still provides a useful estimate, which should, however, only apply to  $\Theta$  conditions where a real polymer chain behaves like an ideal chain and repulsive steric interactions (excluded volume) are exactly compensated by intramolecular attractive forces. To assess the influence of the repulsive steric interactions in the absence of intrachain or solvent interactions, we performed all-atom Monte Carlo simulations of sterically allowed peptide conformations as described in ref. 21. For the (Gly-Ser)<sub>16</sub> peptide, the simulations gave a value of  $(43.9 \pm 1.0) \text{ \AA}$  for  $\langle r^2 \rangle^{1/2}$  between the labels, which is in close agreement with the limiting value of  $(46.4 \pm 1.5) \text{ \AA}$  of the GdmCl-binding isotherm (Fig. 2B).

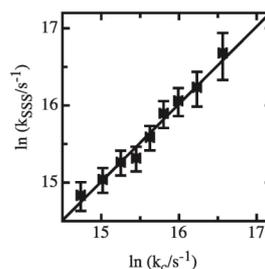
**Properties of Polypeptide Chains in Water.** The above comparisons show that the dimensions of the (Gly-Ser)<sub>16</sub> chain in the presence of high concentrations of denaturant are similar to those predicted for an ideal chain in the absence of intramolecular interactions and to those found for proteins unfolded at high denaturant concentrations. The dimensions of the (Gly-Ser)<sub>16</sub> chain are drastically reduced in water because of the formation of intramolecular interactions, which leads to decreased chain flexibility as indicated by the reduced  $D_{DA}$  value. The (Gly-Ser)<sub>16</sub> peptide does not contain any hydrophobic side chains except for the fluorescence labels, which were shown to have a broad distance distribution (Fig. 2A). NMR experiments did not detect any interactions between donor and acceptor. In addition, fluorescence anisotropy measurements revealed freely rotating chromophores (see *Materials and Methods*), and TTET experiments did not show any fast components in the

kinetics of loop formation in a (Gly-Ser)<sub>16</sub> loop with similar chromophores. Thus, the only possible intrachain interactions in our model chains are hydrogen bonds between backbone amide and backbone carbonyl groups or between backbone carbonyl groups and the serine side chain. Our results indicate that these intramolecular hydrogen bonds are more favorable than hydrogen bonds between the peptide chain and water, which may be due to a favorable solvation free energy of solvent-exposed intramolecular hydrogen bonds, which was recently found to be the major determinant for the stability of alanine-based  $\alpha$ -helices (38–40). The presence of intramolecular hydrogen bonds in unfolded proteins has major consequences for the contributions of hydrogen bonds to protein stability. The hydrogen bond inventory discussed by Fersht (41) assumes that each peptide unit in an unfolded chain forms two hydrogen bonds to water, which are broken upon folding, and a single intramolecular hydrogen bond is formed in the native state. Our results suggest that intramolecular hydrogen bonds are formed both in the unfolded and the native state. Folding thus leads to a breakage of nonspecific intramolecular hydrogen bonds, the formation of specific intramolecular hydrogen bonds, and a burial of the majority of hydrogen bonds, which should result in an unfavorable solvation free energy (38).

Our results reveal that formation of intramolecular hydrogen bonds slows down internal chain dynamics and decreases the chain dimensions. This picture is compatible with results from the effect of chain length on the kinetics of loop formation. In the presence of 6 M GdmCl, a persistence length compatible with the predicted value for an ideal chain (35, 36) is observed (19). In water, however, chain stiffness is significantly increased and Gaussian chain behavior was only observed for longer chains with  $N > 20$  (19). The results from our FRET measurements imply that the increased chain stiffness in water is not due to a local chain stiffness but rather reflects an overall loss in flexibility and chain stiffening due to the formation of a network of intramolecular hydrogen bonds. A comparison of the observed chain properties in water with the dynamics of loop formation (19, 22) allows an estimate of the time scale for structural rearrangement of the intramolecular hydrogen bonding network. Because local chain motions and chain stiffness are affected by the intramolecular interactions, their breakage must be slower than local chain motions, which are on the 10-ps time scale for bond rotations. On the other hand, single-exponential kinetics are observed for loop formation, which indicates rapid conformational equilibration of the ensemble of unfolded states (18, 19, 37, 42, 43). These considerations set an upper limit of a few hundred picoseconds on the interconversion of the peptide conformations.

The intramolecular hydrogen bonds are broken upon binding of GdmCl to the peptide chain, which competes for the hydrogen bonding donor and acceptor groups. However, the scaling laws for loop formation derived for Gaussian chains still apply to the experimentally observed kinetics of loop formation in water. This observation indicates that chain compaction due to hydrogen bond formation does not dramatically alter the polymer behavior of an unfolded chain and that the chain can still be approximated as a Gaussian chain, which is in agreement with conclusions from Fitzkee Rose (44) based on simulations of random coil conformations including intramolecular interactions.

**Comparison of the Results with Predictions from Szabo–Schulten–Schulten Theory.** Szabo, Schulten, and Schulten (42) developed a theory for treating the dynamics of intramolecular end-to-end contact formation in Gaussian chains (SSS theory). For polymers with only a small fraction of loop conformations in equilibrium and with fast dynamics between different conformations, a single-exponential time constant,  $\tau$ , for loop formation (42) was predicted, consistent with the kinetics of loop formation observed in TTET experiments (18, 19, 22). SSS theory gives a relationship between  $\tau$ ,  $p(r)$ , and the diffusion constant ( $D$ ) between the ends of the polymer.



**Fig. 3.** Comparison of rate constants for contact formation determined by TTET,  $k_c$ , with corresponding rate constants calculated according to SSS theory ( $k_{SSS}$ ) using the  $p(r)$  and  $D_{DA}$  values determined by FRET experiments.  $k_{SSS}$  was calculated according to Eq. 8 with a value of 4.4 Å for the reactive boundary,  $r_b$ . The solid line shows the linear correlation with a slope of  $1.00 \pm 0.14$  ( $r = 0.99$ ) and an intercept of  $\ln(k_{SSS}/s^{-1}) = [-0.1 \pm 2.2]$ .

$$\tau = \frac{1}{D} \int_{r_b}^{\infty} \frac{1}{p(x)} \left( \int_x^{\infty} p(y) dy \right)^2 dx \Big/ \int_{r_b}^{\infty} p(x) dx \quad [8]$$

To test SSS theory, we used Eq. 8 with the values for  $D_{DA}$  and  $p(r)$  determined in our FRET experiments to calculate rate constants for end-to-end contact formation in the (Gly-Ser)<sub>16</sub> peptide ( $k_{SSS} = 1/\tau$ ). The results were compared with rate constants for loop formation,  $k_c$ , measured by TTET (19, 22). The only remaining unknown quantity in Eq. 8 is the value of the reactive boundary,  $r_b$ , which designates the distance between two groups required for loop formation. For TTET from xanthone to naphthalene, a bimolecular quenching rate constant of  $k_q = (4.0 \pm 0.2) \times 10^9 \text{ M}^{-1} \text{ s}^{-1}$  was determined (20). Comparing this value with diffusion constants for naphthalene and xanthone of  $(6.5 \pm 1.0) \times 10^{-6} \text{ cm}^2 \text{ s}^{-1}$  and  $(5.6 \pm 1.0) \times 10^{-6} \text{ cm}^2 \text{ s}^{-1}$  (45), respectively, yields a reactive boundary of  $(4.4 \pm 0.5) \text{ \AA}$  for TTET (46), which was used to calculate  $k_{SSS}$  at different GdmCl concentrations. The results showed a linear correlation between the calculated  $k_{SSS}$  values and the experimental  $k_c$  values (Fig. 3) with a slope of the correlation plot of  $1.00 \pm 0.14$  ( $r = 0.99$ ) indicating that SSS theory is able to describe the dynamics of end-to-end loop formation in long unfolded polypeptide chains.

**Implications for the Early Steps in Protein Folding.** The poly(Gly-Ser) chain investigated in our studies serves as a model for an unfolded polypeptide chain that lacks specific side-chain interactions. Comparison of the chain properties in water and in the presence of high concentrations of GdmCl shows drastically reduced chain dimensions in water that are most likely caused by the formation of intramolecular hydrogen bonds. The decrease in chain dimensions accelerates loop formation despite a decreased end-to-end diffusion constant. These results suggest that nonspecific chain compaction can speed up protein folding by accelerating the conformational search on the free-energy landscape for productive side-chain interactions. Because the observed effects of denaturants on chain dimensions and dynamics were obtained in a model chain devoid of side-chain interactions, they likely represent general properties of unfolded polypeptide chains. The results therefore imply that intramolecular hydrogen bond formation has a major contribution to chain compaction early in folding, which is commonly attributed to “hydrophobic collapse.” The results further suggest that rapid structural changes during protein folding do not necessarily indicate the formation of partially folded intermediates. Thus, other methods are required to distinguish between the formation of the defined intermediates and nonspecific chain compaction early in refolding (6, 47).

## Materials and Methods

**Peptide Synthesis and Purification.** Peptides were synthesized with an amidated C terminus using standard Fmoc chemistry on an Applied Biosystems 433A peptide synthesizer as described in ref. 19, using TentaGel S RAM resin (Rapp Polymere, Tübingen, Germany). Resin-bound peptides were N-terminally labeled with dansyl chloride (Fluka, Buchs, Switzerland). The resin was incubated for 30 min in a solution of 10% (vol/vol) diisopropylethylamine in dimethylformamide containing a 10-fold excess of dansyl chloride. Naphthalene was introduced via Fmoc-1-naphthylalanine (NALa; Bachem, Bubendorf, Switzerland). Pyrene was attached to the side chain of diamino propionic acid (NovaBiochem, Läufelfingen, Switzerland). The amine side chain of diamino propionic acid was protected by a methyltrityl group that is selectively cleaved in 2% (vol/vol) TFA in dichloromethane. 1-Pyreneacetic acid (Aldrich, Buchs, Switzerland) was coupled to diamino propionic acid by using standard Fmoc chemistry. Chemicals were of peptide synthesis or higher grade and were purchased from Fluka or Acros Organics (Basel, Switzerland).

Donor- and acceptor-only peptides consisted of eight Gly-Ser repeats, and the dansyl group was replaced by an acetyl moiety or 1-naphthylalanine was replaced by alanine, respectively. Peptides were purified on an RP-8 HPLC column (Merck, Darmstadt, Germany) (19). Peptides labeled with two fluorophores required a second purification on an RP-8 HPLC column with a mixture of acetonitrile and 50 mM ammonium acetate buffer (pH 6.0). Greater than 95% purity of the peptides was confirmed by analytical HPLC and mass spectrometry.

**Spectroscopic Measurements.** Steady-state fluorescence measurements were carried out on an Aminco Bowman series 2 fluorimeter (SLM Aminco).  $R_0$  values were calculated according to Eq. 1. *N*-acetyl tryptophan amide was used as a standard for determining the quantum yield ( $\Phi_F = 0.144$ ) (48). The donor-only peptides showed low steady-state fluorescence anisotropies of  $0.005 \pm 0.006$  and  $0.001 \pm 0.014$  for naphthalene and pyrene, respectively, which gives rotational correlation times smaller than 0.5 ns. This is much

faster than the average fluorescence decay observed in the trFRET measurements and justifies the use of  $\kappa^2 = 2/3$  in Eq. 1.

Time-resolved fluorescence measurements were conducted with a time-correlated single-photon counting fluorimeter (FLS900; Edinburgh Instruments, Edinburgh, U.K.) with irradiation at 284 nm by using a H<sub>2</sub> flash lamp with a 1.5-ns pulse width operating at 40 kHz. Fluorescence decay curves were monitored at 325 and 377 nm for naphthalene and pyrene, respectively. Peptide concentrations were 8  $\mu$ M in 10 mM potassium phosphate buffer (pH 7.0)/0.5% DMSO. GdmCl concentrations were determined by refractive index (49). Samples were degassed and equilibrated at 22.5°C. A mixture of 4  $\mu$ M acceptor-only and 4  $\mu$ M donor-only peptide showed no energy transfer, ruling out intermolecular FRET processes.

CD measurements were performed on an Aviv DS62 spectropolarimeter with peptide concentrations between 15 and 150  $\mu$ M.

**Data Analysis.** Data were analyzed with the programs MATLAB (MathWorks, Natick, MA) and pro Fit (QuantumSoft, Zürich, Switzerland). Fluorescence decay curves from the two FRET pairs were fitted globally to the numerical solution of Eqs. 2 and 3 with the global parameters *a*, *b*, and  $D_{DA}$ , because even for simple  $p(r)$ , Eq. 2 does not have an analytical solution. The lifetimes of the two donors in the absence of acceptor were fitted separately at each GdmCl concentration with single-exponential functions (see Tables 1 and 2), and the results were used in the global analysis.

**Monte Carlo Simulation of Donor-Acceptor Distance Distribution.** Random conformations for the (Gly-Ser)<sub>16</sub> peptide without steric overlap were generated with all-atom hard-sphere model as described in ref. 21. For simulations, xanthone and naphthalene, which have similar size as the FRET pairs used in these studies, were used as donor and acceptor. In total, >60,000 valid conformations without steric clashes were generated.

We thank Buzz Baldwin, Beat Fierz, and Elisha Haas for discussion; Annett Bachmann and George Rose for comments on the manuscript; and Thomas Aust for mass spectroscopy.

1. Kuwajima, K. (1989) *Protein Struct. Funct. Genet.* **6**, 87–103.
2. Pitsyn, O. B., Pain, R. H., Semisotnov, G. V., Zerovnik, E. & Razglyayev, O. I. (1990) *FEBS Lett.* **262**, 20–24.
3. Eaton, W. A., Thompson, P. A., Chan, C. K., Hagen, S. J. & Hofrichter, J. (1996) *Structure (London)* **4**, 1133–1139.
4. Segel, D., Bachmann, A., Hofrichter, J., Hodgson, K., Doniach, S. & Kiefhaber, T. (1999) *J. Mol. Biol.* **288**, 489–500.
5. Roder, H., Maki, K., Latypov, R. F., Cheng, H. & Shastry, M. C. R. (2005) in *Protein Folding Handbook*, eds. Buchner, J. & Kiefhaber, T. (Wiley-VCH, Weinheim, Germany), Vol. 1, pp. 491–535.
6. Bachmann, A. & Kiefhaber, T. (2002) *Biophys. Chem.* **96**, 141–151.
7. Dyson, H. J. & Wright, P. E. (2004) *Chem. Rev.* **104**, 3607–3622.
8. 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) *Science* **295**, 1719–1722.
9. Religa, T. L., Markson, J. S., Mayor, U., Freund, S. M. V. & Fersht, A. R. (2005) *Nature* **437**, 1053–1056.
10. Neri, D., Billeter, M., Wider, G. & Wuthrich, K. (1992) *Science* **257**, 1559–1563.
11. Logan, T. M., Theriault, Y., Fesik, S. W. (1994) *J. Mol. Biol.* **236**, 637–648.
12. Yi, Q., Scalley-Kim, M. L., Alm, E. J. & Baker, D. (2000) *J. Mol. Biol.* **299**, 1341–1351.
13. Kortemme, T., Kelly, M. J., Kay, L. E., Forman-Kay, J. D. & Serrano, L. (2000) *J. Mol. Biol.* **297**, 1217–1229.
14. Wirmer, J., Schlörb, C. & Schwalbe, H. (2005) in *Protein Folding Handbook*, eds. Buchner, J. & Kiefhaber, T. (Wiley-VCH, Weinheim, Germany), Vol. 2, pp. 737–808.
15. Sanchez, I. E. & Kiefhaber, T. (2003) *J. Mol. Biol.* **327**, 867–884.
16. Pappu, R. V., Srinivasan, R. & Rose, G. D. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 12565–12570.
17. Gong, H., Fleming, P. J. & Rose, G. D. (2005) *Proc. Natl. Acad. Sci. USA* **102**, 16227–16232.
18. Bieri, O., Wirz, J., Hellrung, B., Schutkowski, M., Drewello, M. & Kiefhaber, T. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 9597–9601.
19. Krieger, F., Fierz, B., Bieri, O., Drewello, M. & Kiefhaber, T. (2003) *J. Mol. Biol.* **332**, 265–274.
20. Krieger, F., Fierz, B., Axthelm, F., Joder, K., Meyer, D. & Kiefhaber, T. (2004) *Chem. Phys.* **307**, 209–215.
21. Krieger, F., Möglich, A. & Kiefhaber, T. (2005) *J. Am. Chem. Soc.* **127**, 3346–3352.
22. Möglich, A., Krieger, F. & Kiefhaber, T. (2005) *J. Mol. Biol.* **345**, 153–162.
23. Schellman, J. A. (1987) *Biopolymers* **26**, 549–559.
24. Schellman, J. A. (2002) *Biophys. Chem.* **96**, 91–101.
25. Förster, T. (1948) *Ann. Phys.* **2**, 55–75.
26. Haas, E., Katchalski-Katzir, E. & Steinberg, I. Z. (1978) *Biopolymers* **17**, 11–31.
27. Lakowicz, J. R., Kusba, J., Wicz, W., Gryczynski, I. & Johnson, M. L. (1990) *Chem. Phys. Lett.* **173**, 319–326.
28. Lakowicz, J. R., Kusba, J., Gryczynski, I., Wicz, W., Szmajcinski, H. & Johnson, M. L. (1991) *J. Phys. Chem.* **95**, 9654–9660.
29. Haas, E., Wilchek, M., Katchalski-Katzir, E. & Steinberg, I. Z. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 1807–1811.
30. Sadqi, M., Lapidus, L. J. & Munoz, V. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 12117–12122.
31. Weast, R. C., ed. (1972) *Handbook of Chemistry and Physics* (CRC, Cleveland).
32. Damaschun, G., Damaschun, H., Gast, K. & Zirwer, D. (1998) *Biochemistry (Moscow)* **63**, 259–275.
33. Kohn, J. E., Millett, I. S., Jacob, J., Zagrovic, B., Dillon, T. M., Cingel, N., Dothager, R. S., Seifert, S., Thiagarajan, P., Sosnick, T. R., et al. (2004) *Proc. Natl. Acad. Sci. USA* **101**, 12491–12496.
34. Schimmel, P. R. & Flory, P. J. (1967) *Proc. Natl. Acad. Sci. USA* **58**, 52–59.
35. Miller, W. G., Brant, D. A. & Flory, P. J. (1967) *J. Mol. Biol.* **23**, 67–80.
36. Flory, P. J. (1969) *Statistical Mechanics of Chain Molecules* (Hanser, Munich).
37. Fierz, B. & Kiefhaber, T. (2005) in *Protein Folding Handbook*, eds. Buchner, J. & Kiefhaber, T. (Wiley-VCH, Weinheim, Germany), Vol. 2, pp. 805–851.
38. Baldwin, R. L. (2003) *J. Biol. Chem.* **278**, 17581–17588.
39. Lopez, M., Chin, D.-H., Baldwin, R. L. & Makhatadze, G. I. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 1298–1302.
40. Richardson, J. M., Lopez, M. & Makhatadze, G. I. (2005) *Proc. Natl. Acad. Sci. USA* **102**, 1413–1418.
41. Fersht, A. R. (1987) *Trends Biochem. Sci.* **12**, 301–304.
42. Szabo, A., Schulten, K. & Schulten, Z. (1980) *J. Chem. Phys.* **72**, 4350–4357.
43. Bieri, O. & Kiefhaber, T. (2000) in *Protein Folding: Frontiers in Molecular Biology*, ed. Pain, R. (Oxford Univ. Press, Oxford), pp. 34–64.
44. Fitzkee, N. C. & Rose, G. D. (2004) *Proc. Natl. Acad. Sci. USA* **101**, 12497–12502.
45. Terazima, M., Okamoto, K. & Hirota, N. (1995) *J. Chem. Phys.* **102**, 2506–2515.
46. von Smolouchowski, M. V. (1917) *Z. Phys. Chem.* **129**–168.
47. Kay, M. S. & Baldwin, R. L. (1996) *Nat. Struct. Biol.* **3**, 439–445.
48. Navon, A., Ittah, V., Landsman, P., Scheraga, H. A. & Haas, E. (2001) *Biochemistry* **40**, 105–118.
49. Pace, C. N. (1986) *Methods Enzymol.* **131**, 266–280.