Novel weak alignment techniques for nuclear magnetic resonance spectroscopy and applications to biomolecular structure determination

Inauguraldissertation

zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Sebastian Meier
aus Deutschland

Basel, 2004

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Prof. Dr. Stephan Grzesiek
Prof. Dr. Frédéric Allain

Basel, den 6.7.2004

Prof. Dr. Marcel Tanner
Dekan

Acknowledgements

Research described in this thesis was carried out in the Biozentrum of the University of Basel. My sincere thanks go to my supervisor, Professor Stephan Grzesiek. I am deeply grateful for his support, guidance and encouragement and I was privileged to benefit from his ample knowledge over the years in Basel.

I wish to thank the members of the group headed by Prof. Stephan Grzesiek and the colleagues, past and present, with whom I had the pleasure to work: Bernhard, Daniel, Debbie, Elena, Eunjung, Florence, Giovanna, Jan, Jörg, Jürgen, Klara, Leo, Marco, Martin, Pernille, Sabine, Sarah, Sonja, Stefan, Suat and Victor. Among those, a sweeping “thank you” to my office mates Daniel and Florence for their help and for the scientific as well as social interactions.

I thank Prof. Jürgen Engel, Prof. Hans Peter Bächinger, Prof. Thomas Kiefhaber, Dr. Yu-Chin Li and Dr. Wolfgang Jahnke for fruitful collaborations and for their interest in our work.

I want to thank Prof. Frédéric Allain for kindly accepting to coreferee this thesis.

Thanks to those, who make/made Basel a more interesting and fun city, particularly Pernille, Familie Güthe, Zsazsa, Mason, Dan, The ecstatic secret society and Cindy a.k.a. Rod and Cindy, Flo, Tim, Sonja and Tom, Jan, Rainer, Jörg, Carmen, Janelle, Nicolas, Martin, Elsi and Gaia, Arne, Gio.

Greatest thanks to family and friends for their support.

Basel, May 2004 Sebastian Meier

Summary

Nuclear magnetic resonance spectroscopy has continuously been developing ever since its introduction as a structural method in bioscience. Recently established residual dipolar coupling techniques yield information on long-range order in weakly aligned samples as they define the orientation of vectors between nuclei in a common global reference frame. These data complement classical short-range information and have a unique potential especially for the characterization of non-globular states.

This thesis describes the development of novel methods for the weak alignment of biomacromolecules in charged gels and for the measurement of long-range residual dipolar couplings in perdeuterated proteins. These weak alignment techniques and other nuclear magnetic resonance methods were applied to study the structure and folding of various proteins such as the fibritin folding nucleus, the minicollagen cysteine rich domain and human protein tyrosine phosphatase 1B.

Results from this thesis have been published in the following peer-reviewed articles:

1. Meier S., Häussinger D., Grzesiek S.:
   Charged acrylamide copolymer gels as media for weak alignment.
   Journal of Biomolecular NMR 2002, **24**:351-356.

2. Meier S., Häussinger D., Jensen P., Rogowski M., Grzesiek S.:
   High-accuracy residual $^{1}H^{N}$-^{13}C and $^{1}H^{N}$-$^{1}H^{N}$ dipolar couplings in perdeuterated proteins.
   Journal of the American Chemical Society 2003, **125**:44-45.

3. Güthe S., Kapinos L., Möglich A., Meier S., Grzesiek S., Kiefhaber T.:
   Very fast folding and association of a trimerization domain from bacteriophage T4 fibritin.
   Journal of Molecular Biology 2004, **337**:905-915.

4. Meier S., Güthe S., Kiefhaber T., Grzesiek S.:
   Structure and dynamics in the fibritin folding nucleus containing a stable β-hairpin.
   2004, ***in preparation***

5. Meier S., Häussinger D., Pokidysheva E., Bächinger H.P., Grzesiek S.:
   Determination of a high-precision NMR structure of the minicollagen cysteine rich domain from Hydra and characterization of its disulfide bond formation.
   FEBS Letters 2004, ***in press***.

6. Meier S., Pokidysheva E., Milbradt A., Renner C., Häussinger D., Bächinger H.P., Moroder L., Grzesiek S., Holstein T., Özbek S., and Engel, J.:
   The structure of the Cys-rich terminal domain of Hydra minicollagen, which is involved in disulfide networks of the nematocyst wall.
   Journal of Biological Chemistry 2004, ***in press***.

7. Meier S., Li Y.C., Koehn J., Vlattas I., Wareing J., Jahnke W.G., Wennogle L.P., Grzesiek S.:
   Letter to the Editor: Backbone resonance assignment of the 298 amino acid catalytic domain of protein tyrosine phosphatase 1B (PTP1B).
   Journal of Biomolecular NMR 2002, **24**:165-166.

Contents

Abbreviations and symbols

$\{^1H\}$-^{15}N NOE	^{15}N steady state NOE upon ^1H saturation
A	diagonal alignment tensor with components A_{xx}, A_{yy}, A_{zz}
Å	0.1 nm
A-state	acidic state
A_{zz}	z component of the diagonalized alignment tensor, $D = \frac{A_{zz}}{2}\left\{(3\cos^2\theta - 1) + \eta\sin^2\theta\cos 2\varphi\right\}$
BMRB	BioMagResBank, www.bmrb.wisc.edu
COSY	correlation spectroscopy
CRD	cysteine rich domain
δ	chemical shift
δ_{ij}	Kronecker symbol
D	dipolar coupling
D_{max}	$-\frac{\mu_0}{4\pi^2}\frac{\gamma_I\gamma_J}{r_{IJ}^3}\frac{h}{2\pi}$, maximal solid state dipolar coupling
d_{NiNj}	$^1H^N$-$^1H^N$ NOE connectivity between residues i and j
DHPC	diheptanoyl-phosphatidylcholine
DMPC	dimyristoyl-phosphatidylcholine
E^0	Nernst potential
γ_X	gyromagnetic ratio of nucleus X
G	Gibbs free enthalpy
GSH	glutathione
η	$(A_{xx}-A_{yy})/A_{zz}$, asymmetry parameter
h	$6.6\cdot10^{-34}$ Js, Planck's constant
ΔH	enthalpy of transition
HMQC	heteronuclear multiple quantum coherence
HSQC	heteronuclear single quantum coherence
INEPT	insensitive nucleus enhancement by polarization transfer
I_z	*z* component of the spin operator
J	scalar coupling
μ_0	$4\pi\cdot10^{-7}$ V·s/(A·m), vacuum permeability

MCOL1	minicollagen 1
MCRD	minicollagen cysteine rich domain
MD	molecular dynamics
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser enhancement spectroscopy
NOWA	nematocyst outer wall antigen
$P_2(x)$	$1/2\cdot(3x^2-1)$, 2nd order Legendre polynomial
PDB	RCSB Protein Data Bank, www.rcsb.org/pdb
ppm	parts per million
PTK	protein tyrosine kinase
PTP	protein tyrosine phosphatase
R_1, R_2	longitudinal and transverse relaxation rate
r_{IJ}	internuclear distance between I and J
RDC	residual dipolar coupling
rmsd	root mean square deviation
ROE	rotating frame Overhauser effect
S	order parameter
S_{ij}	Saupe order matrix
ΔS	entropy of transition
θ,φ	polar angles
t_1, t_2, …	chemical shift evolution times
T_1, T_2	longitudinal and transverse relaxation time
TCEP	tris(carboxyethyl)phosphine hydrochloride
TEMED	N,N,N',N'-tetramethylethylenediamine
TOCSY	total correlation spectroscopy
TROSY	transverse relaxation optimized spectroscopy
U	uniformly labeled

Atoms and angles are referred to according to IUPAC nomenclature.

Chapter 1
Introduction

For almost five decades structural biology has formed the framework in which biological processes can be understood at the atomic level (Kendrew et al., 1958; Perutz et al., 1960). Today biomolecular structures are solved at ever-increasing rates, sizes and qualities, as the applied methods are continuously improving. Structural proteomics efforts are underway to explore all folds, which naturally occur in proteins. New techniques like protein design, structure prediction and molecular modeling have emerged from structural biology.

The various techniques for the determination of biomolecular structures mainly give complementary rather than redundant information. Data from cryo-electron microscopy and atomic force microscopy yield structures of extremely large systems or whole cells (Baumeister, 2002) at relatively low resolution, that can, however, often be combined with atomic resolution structures.

The most successful structural technique at atomic resolution is the diffraction of X-ray beams from three-dimensional crystals. However, the production of good crystals is a limiting factor. The use of synchroton radiation as well as cryo-crystallographic methods and advances in refinement techniques permit to study protein structures with molecular weights up to $1.8 \cdot 10^6$ Da (Ban et al., 2000) and with resolutions as good as 0.54 Å for small model proteins (Jelsch et al., 2000).

Nuclear magnetic resonance spectroscopy is the only alternative to X-ray crystallography for studying structures at atomic resolution and has the advantage of working in solution. The most severe drawback of the method is the size limitation, which results from increasing linewidths at increasing molecular weight. Due to the high number of small proteins in any genome, to the better solubility of smaller proteins and to the rapidity of preliminary analysis via NMR, NMR has been predicted to play an important role in the current structural proteomics efforts (Christendat et al., 2000). Domains of eucaryotic proteins have an average size of 153 ± 87 residues (Orengo et al., 1999). Most of these domains are thus amenable to state-

of-the-art heteronuclear NMR methods. With relaxation optimized spectroscopy, spectra of multimeric proteins have been obtained at molecular sizes up to 810 kDa in lysine decarboxylase (Tugarinov et al., 2004) and 900 kDa in the GroEL-GroES complex (Fiaux et al., 2002). Thus, NMR has the potential to contribute to the understanding of large modular proteins by resolving the biology of concerted domain actions. In a first example, resonance assignments and a characterization of interdomain and intradomain motions were performed on the 81.4 kDa multidomain enzyme malate synthase G (Tugarinov et al., 2002; Tugarinov and Kay, 2003; Korzhnev et al., 2004).

High-resolution solid state NMR (SSNMR) has allowed the characterization of proteins, which are not amenable to other structural techniques. SSNMR does not have an inherent molecular weight limit and allows studies of membrane proteins (de Groot, 2000), protein aggregates like amyloid fibrils (Petkova et al., 2002) and other proteins and nucleic acids, which are insoluble and non-crystallizeable. Local structural information, entire protein structures (Castellani et al., 2002) and *de novo* peptide structures (Luca et al., 2003) have been obtained with partially (Castellani et al., 2003) and uniformly labeled samples (Jaroniec et al., 2001).

NMR spectroscopy has been termed “the most versatile physicochemical method developed in the twentieth century” (Cohen et al., 1995). The method has a broad theoretical background and is understood from first principles (Abragam, 1961; Goldman, 1988). Since the discovery of the chemical shift (Proctor and Yu, 1950) NMR has been a tool for the analysis of molecular configuration in chemistry. NMR has evolved to a high-resolution method for atomic resolution structure determination of biomacromolecules in solution (Wüthrich, 1986; Ernst et al., 1987) and to a highly powerful clinical imaging technique (Lauterbur, 1973; Stehling et al., 1991).

NMR allows the concerted characterization of structure, dynamics, folding and various functional states of proteins, often with a single sample. It is an efficient screening technology for ligand and lead binding, even at low affinities, in bio- and pharmaceutical sciences (Hajduk et al., 1999). Hydrogen bonds in small molecules (Shenderovich et al., 1998), nucleic acids (Dingley and Grzesiek, 1998; Pervushin et al., 1998) and proteins (Wagner et al., 1983; Cordier and Grzesiek, 1999; Cordier et

al., 2003) have directly been studied using NMR. Furthermore, the method allows the characterization of only partially folded or unstructured states (Wright and Dyson, 1999). Flexible parts of proteins are often involved in binding and catalysis and are therefore best studied in solution by NMR. NMR data allow extracting time scales, amplitudes and anisotropies of motions in dynamic structures (Lipari and Szabo, 1982; Kay et al., 1989; Brüschweiler, 2003).

Structure determination by NMR in solution has traditionally relied on semi-quantitative local restraints, like the distance dependent nuclear Overhauser effect (NOE) between protons and dihedral angle information from coupling constants and chemical shifts. Errors in these local restraints can be cumulative and limit the accuracy of NMR-derived structures (Gronenborn, 2003). Especially structure determinations of non-globular states like partially unfolded proteins, multidomain proteins and nucleic acids with their sparse proton density can fail, when relying on traditional NMR parameters only.

The accuracy of NMR derived structures has been improved by the introduction of experimental restraints that characterize the long-range order in macromolecules (Tolman et al., 1995; Tjandra and Bax, 1997). A very weak degree of anisotropic orientation of the biomolecules is the prerequisite for obtaining these data. These anisotropic restraints contain information about internuclear vector orientations in a global common reference frame and therefore yield the orientation of internuclear vectors relative to each other, irrespective of their distance. Most prominent among these restraints is the dipolar coupling, that has been measured in solution NMR using liquid crystals and magnetic alignment (Saupe and Englert, 1963; Bothnerby et al., 1981). A routine use in solution NMR was, however, hampered by the complexity of the spectra even for small molecules, resulting from the high degree of ordering in original liquid crystalline samples. A major breakthrough has been achieved by the use of dilute liquid crystalline media (Tjandra and Bax, 1997) to scale down the alignment by a factor of ~1000 and to allow the convenient measurement of residual dipolar couplings (RDCs) in the weak coupling limit.

The size of the RDC between two nuclei depends on their magnetic moment, their distance, the angle of the internuclear vector with respect to the magnetic field and the order parameter of the internuclear vector. The RDC $\langle D \rangle$ is the time and ensemble averaged (indicated by brackets $\langle\ \rangle$) dipolar coupling between nuclei I and J given by

$$\langle D \rangle = -\frac{\mu_0}{4\pi^2}\frac{h}{2\pi}\gamma_I\gamma_J\left\langle\frac{3\cos^2\zeta - 1}{2\cdot r_{IJ}^{\ 3}}\right\rangle,$$

where ζ is the angle of the internuclear vector with the magnetic field and r_{IJ} is the internuclear distance. The factor $-\frac{\mu_0}{4\pi^2}\frac{h}{2\pi}\gamma_I\gamma_J$ only varies with the magnetic moment of the nuclei.

Due to heteronuclear isotope enrichment, pairwise RDCs between ^{1}H, ^{13}C and ^{15}N nuclei have become accessible, mostly by simple modifications of experiments normally used for resonance assignments. RDCs have been shown to improve NMR structures to qualities comparable to high-resolution X-ray structures and monitor local chemical exchange processes (Tolman et al., 1997) as well as global topologies in protein folding (Shortle and Ackerman, 2001). Domains and intermolecular complexes can be oriented relative to each other, often by the use of only very few RDC data (Tolman, 2001). Structure determinations have been carried out by the use of RDCs with a limited number of NOEs (Ramirez et al., 2000) or no NOE data at all (Delaglio et al., 2000; Hus et al., 2001). The fast identification of folds by RDCs can be used in structural proteomics efforts, even in cases, where the sequence similarity is too small to predict structural homologues (Annila et al., 1999). As the number of unique protein-folds is limited to roughly 1000 (Chothia, 1992), structural homologues of the investigated protein are often found, when searching the PDB with dipolar coupling input.

This thesis describes the development of charged gels as new media for weak alignment and introduces new methods for the measurement of long-range RDCs in perdeuterated proteins. Weak alignment is applied to high-resolution structural studies of native states, partially folded intermediates and unfolding. Furthermore, backbone assignments of a 35 kDa human enzyme were obtained with classical and relaxation optimized spectroscopy. The thesis is organized as follows:

In chapter 2 the development of charged copolymer gels as chemically robust synthetic media for the electrostatic alignment of biomolecules is described (Meier et al., 2002a). In addition, new experiments have been devised for the measurement of long-range residual dipolar couplings. In these experiments, the RDC is measured in perdeuterated proteins by magnetization transfer from $^{1}H^{N}$ spins to sequential and non-sequential $^{1}H^{N}$, $^{13}C^{aliphatic}$ and $^{13}C'$ atoms, including RDCs across hydrogen bonds (Meier et al., 2003). RDCs between spins up to nearly 8 Å apart could be detected.

In chapter 3 the solution structure of the trimeric T4 fibritin foldon domain is described (Güthe et al., 2004). Electrostatic contributions to the total stabilization energy were determined by NMR spectroscopy. At low pH values the foldon trimer disassembles and gives rise to an acidic state with a highly populated single hairpin. Stable natural hairpins have been difficult to obtain and are important models for the understanding of β-sheet formation. The thermodynamics and kinetics of foldon hairpin formation have been quantified with NMR spectroscopy (Meier et al., 2004a).

Chapter 4 describes the structure determination of a disulfide and proline rich minicollagen domain from *Cnidaria* (Meier et al., 2004c). Due to the detection of $^{1}D_{CH}$ and $^{1}D_{NH}$ dipolar couplings at natural abundance of ^{13}C and ^{15}N isotopes, the structure is of very high quality and reveals the disulfide pattern of the six cysteines and the hydrogen bond pattern in several turns of the small domain. The redox potential and the folding of these specialized collagen domains with their exceptional content of cysteines and prolines have been characterized (Meier et al., 2004b).

In chapter 5 the backbone assignment of the 298 amino acid catalytic domain of human protein tyrosine phosphatase 1B is described (Meier et al., 2002b). PTP1B is a promising target for the treatment of obesity and type II diabetes. The assignments are currently being used for lead screening by Novartis Pharmaceuticals Corporation.

Chapter 2
Media and methods for the measurement of residual dipolar couplings

2.1 Background

Several different NMR restraints can be derived for the structure determination of biomacromolecules in solution. Traditionally, the most important parameter for NMR structure determination has been the nuclear Overhauser effect (NOE) between protons, which are in close spatial proximity (Overhauser, 1953). The precision of NOE-derived distances, however, is compromised by spin diffusion and the imprecise knowledge of local dynamical processes of the interacting protons. NOEs therefore enter structure calculations in a rather qualitative manner. In addition, the unambiguous interpretation of NOE cross peaks gets increasingly difficult with increasing size of the protein (Zwahlen et al., 1998). The detection of proton-proton NOEs in high molecular weight proteins is additionally impaired by the usual (at least partial) deuteration of large proteins, which is necessary to lengthen transverse spin relaxation times. Other local restraints historically used in NMR comprise *J*-couplings (Karplus, 1959) or chemical shifts (Spera and Bax, 1991), which both give dihedral angle information. The use of only local information inherently limits the accuracy of NMR structures, especially for non-compact states (Gronenborn, 2003). Resulting NMR structures are of acceptable quality only due to the redundancy of the information (de Alba and Tjandra, 2002).

Residual dipolar couplings and other orientation-dependent data in weakly aligned samples alleviate these problems by complementing local NOE and dihedral angle information with long-range data.

The dipolar field of a magnetic nucleus with spin I adds to the local magnetic field and thus changes the resonance frequency of a nearby magnetic nucleus with spin J by a value depending on the orientation of the internuclear vector, the magnetic moment of the spins and their distance (Figure 2–1). For spin-1/2 nuclei, half of the

spins I are parallel to the magnetic field and half are antiparallel. Therefore, spins I will increase or decrease the frequency of J in 50 % of the cases each, thus giving rise to a symmetric splitting.

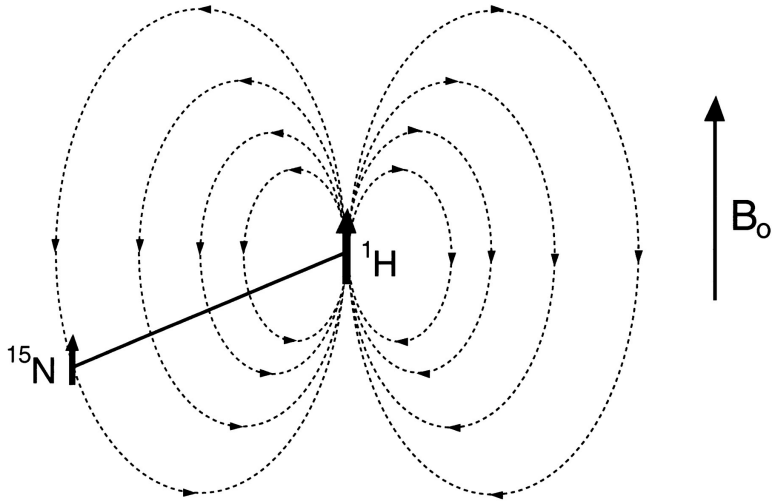


Figure 2–1. Dipolar coupling (taken from (Bax, 2003)). The dipolar field, in this case of the proton, influences the local field experienced by nearby nuclei, here the nitrogen. The magnitude of this change in the local field depends on the position of the nitrogen relative to the proton, whereas the parallel and the antiparallel spin state of the proton relative to the external magnetic field invert the sign of this influence.

In the heteronuclear case, the energy of the interaction between dipoles with identical orientation relative to the internuclear vector ***r*** is given from classical electrodynamics by

$$H_{DD} = -\frac{\mu_0}{4\pi^2}\frac{h}{2\pi}\gamma_I\gamma_J I_z J_z \frac{(3\cos^2\zeta - 1)}{2\cdot r_{IJ}^3},$$

where ζ is the angle of the internuclear vector relative to the external magnetic field, I_z and J_z are z components of the spin operators, r_{IJ} is the internuclear distance between spins I and J and γ_I, γ_J are gyromagnetic ratios.

Molecular tumbling yields a time and ensemble averaging, indicated by angle brackets, and gives

$$H_{DD} = D_{\max}\left\langle I_z J_z \frac{(3\cos^2\zeta - 1)}{2}\right\rangle,$$

where $D_{\max} = -\frac{\mu_0}{4\pi^2}\frac{h}{2\pi}\frac{\gamma_I\gamma_J}{r_{IJ}^3}$ is the maximal splitting in Hz and the internuclear distance r_{IJ} is assumed fixed.

In the heteronuclear case, scalar and dipolar couplings between the spins I and J are additive due to the identical form of the spin parts of the respective hamiltonians. The RDC is then extracted by subtraction of the measured scalar coupling, which is obtained in a reference experiment performed on an isotropic sample.

The residual dipolar coupling between two nuclei has the form

$$\langle D \rangle = D_{\max} \langle P_2(\cos\zeta) \rangle,$$

where $P_2(\cos\zeta) = (3\cos^2\zeta - 1)/2$. In isotropic solution the dipolar coupling averages to zero because the integral of $P_2(\cos\zeta)$ over the unit sphere vanishes. For typical alignment media used to detect RDCs in biomolecules, values of $\langle P_2(\cos\zeta) \rangle$ are adjusted to $\sim 10^{-3}$ such that the RDCs adopt values on the order of a few Hz.

It is often convenient to express the dipolar coupling in a molecular fixed frame. The molecular average orientation in the magnetic field has been described by the symmetric 3×3 Saupe order matrix S_{ij} given by (Saupe and Englert, 1963)

$$S_{ij} = \frac{1}{2} \langle 3\cos\zeta_i \cos\zeta_j - \delta_{ij} \rangle,$$

where $i,j = x,y,z$ are the axes of the molecular Cartesian coordinate system, ζ_i the angle of the axis with the magnetic field and δ_{ij} the Kronecker symbol.

The Saupe order matrix can then be used to obtain the orientation of an internuclear vector in the field, i.e. to derive $\langle P_2(\cos\zeta) \rangle$ from

$$\langle P_2(\cos\zeta) \rangle = \sum_{\substack{i=x,y,z \\ j=x,y,z}} c_i S_{ij} c_j,$$

where (c_x, c_y, c_z) are the coordinates of an internuclear vector in the Cartesian coordinate system fixed to the molecule.

As the Saupe order matrix is real, symmetric and traceless, it can be diagonalized by the rotation to the principal axis coordinate system. D then assumes the form

$$D = D_{\max} \langle P_2 \cos(\zeta) \rangle = D_{\max} \left[S'_{xx} c'^2_x + S'_{yy} c'^2_y + S'_{zz} c'^2_z \right],$$

where the primes denote quantities within the principal axis system.

The internuclear vector orientation in the principal axis frame is conveniently described by $\Omega=(\theta,\varphi)$, where θ and φ denote the polar angles of the vector. The dipolar coupling is then given as

$$D = D_{\max} S'_{zz} \left[\frac{(3\cos^2\theta - 1)}{2} + \frac{\eta}{2}\sin^2\theta\cos 2\varphi \right],$$

frequently also expressed as

$$D = \frac{A_{zz}}{2}\left\{(3\cos^2\theta - 1) + \eta\sin^2\theta\cos 2\varphi\right\},$$

where $A_{zz} = D_{max}\cdot S'_{zz}$ is the maximal RDC obtainable for a given alignment and η is the asymmetry parameter of the diagonalized alignment tensor $\mathbf{A}$, with $\eta = (A_{xx}-A_{yy})/A_{zz}$, Tr $\mathbf{A} = 0$ and elements $|A_{zz}|>|A_{yy}|>|A_{xx}|$. The geometric dependence of the RDC is exclusively orientational, if the distance r_{IJ} is known. If r_{IJ} is not known, both distance and orientation influence the RDC.

For a molecule with internal motions, the description of the RDC becomes more complicated, but is straightforward, if the internal motion does not affect the alignment process (Prestegard et al., 2000). In the limit that the internal motion is axially symmetric around the mean orientation, the dipolar coupling for the average orientation is simply scaled by an internal order parameter S ($0\le S\le 1$) and becomes

$$D = S\frac{A_{zz}}{2}\left\{(3\cos^2\theta - 1) + \eta\sin^2\theta\cos 2\varphi\right\}.$$

S corresponds to the order parameter used to describe spin relaxation (Lipari and Szabo, 1982), but senses motions up to the millisecond timescale. RDCs scale linearly with the order parameter, whereas spectral density functions and thus relaxation rates only contain S^2 terms. For structured residues in proteins relaxation derived order parameters are in the range $S^2 \approx 0.8$-0.95. Thus, local differences in the order parameter contribute only a few percent to the RDC in structured regions of a molecule and are often neglected.

Weak alignment does not only result in residual dipolar couplings but also in other orientation-dependent phenomena like residual chemical shift anisotropy, anisotropic

scalar couplings and quadrupolar couplings. Among those, the RDC is the most commonly used parameter, due to the inherent strength of the dipolar coupling ($D_{NH,max}$ = 22 kHz for $^{1}H^{15}N$ spin pairs with r_{NH} = 1.02 Å) and due to the convenient measurement of the splitting in only one experiment with high precision at weak alignment.

The two polar angles φ and θ cannot be determined unambiguously by the measurement of a single RDC value. The measured dipolar coupling limits the orientation of a bond vector to two symmetric lines on a sphere (Figure 2–2). However, this ambiguity of the bond vector orientation is diminished, if the protein is aligned in a second medium with different alignment tensor (Ramirez and Bax, 1998). By this means ambiguities of the internuclear vector orientation for peptide planes can be reduced (Figure 2–2) or in the case of non-planar molecular fragments even be removed (Al-Hashimi et al., 2000).

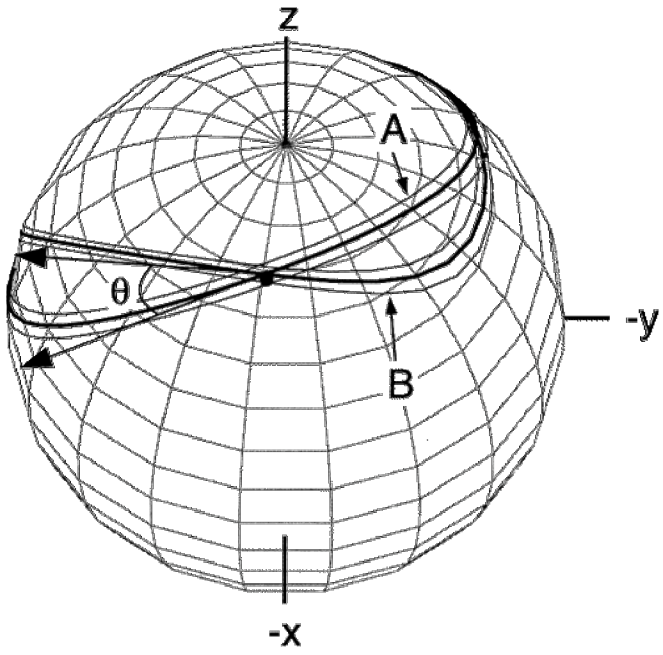


Figure 2–2. RDCs determine the orientations of internuclear vectors to a line with a relatively small error. The vector direction is ambiguous. By measuring the residual dipolar coupling in two alignment media, the orientation can be defined to the intersections of these lines (taken from (Ramirez and Bax, 1998)).

Weak alignment for the measurement of orientation-dependent NMR restraints is achieved by transient interactions of the solute with an anisotropic matrix. Residual dipolar couplings are measured at a net alignment of approximately 1 part out of

1000. In an ergodic system, this residual degree of alignment is identical to the time average that a molecule is interacting with the matrix, so that the molecule tumbles in solution for roughly 99.9 % of the time.

Tunable weak alignment has first been achieved by placing proteins into dilute aqueous liquid crystalline media, composed of dihexanoyl- (or diheptanoyl-) phosphatidylcholine and dimyristoyl-phosphatidylcholine (DHPC/DMPC) (Tjandra and Bax, 1997). These mixtures presumably form disc-like particles commonly termed bicelles (Sanders and Schwonek, 1992), which adopt oriented nematic liquid crystalline phases in the magnetic field due to their magnetic susceptibility. DMPC and DHPC are uncharged and interactions with the solute occur by steric collision. In this situation, simple obstruction models could predict the steric alignment tensor to high accuracy (Zweckstetter and Bax, 2000). The alignment tensor is modified, when bicelles are doped with charged amphiphiles, thus adding a net charge to the bicelles (Ramirez and Bax, 1998). This results in electrostatic contributions to the interaction of solute and alignment medium.

Other synthetic alignment media with varying steric and electrostatic contributions are lamellar liquid crystal phases, which can be obtained from cetylpyridinium halide/alcohol mixtures in bile (Helfrich phases (Prosser et al., 1998)) and from polyether/alcohol mixtures (Lα phases (Ruckert and Otting, 2000)). An alternative to the alignment in synthetic media is the use of negatively charged filamentous viruses (Clore et al., 1998; Hansen et al., 1998a) and purple membranes (Sass et al., 1999), which orient in the magnetic field due to their intrinsic magnetic susceptibility.

All of these alignment media have the drawback that they are stable only over a limited range of temperature, pH and ionic strength.

Mechanically strained covalent polyacrylamide hydrogels have been introduced as alignment media (Sass et al., 2000; Tycko et al., 2000), which are chemically inert and yield stable alignment over a wide range of temperature, pH and solvent composition. Anisotropy in the gels is generated by axial or radial squeezing, which results in anisotropic cavities in the gels. Weak alignment is thus imposed on the

solute macromolecule. As polyacrylamide is uncharged, the alignment results from steric interactions. The gels have been used for weak alignment in solutions of 8 M urea to characterize denatured states (Shortle and Ackerman, 2001) and in thermal unfolding up to 80 °C (Meier et al., 2004a).

In the sum, a large number of alignment media for the measurement of RDCs has been introduced over the last years, because gaining residual dipolar couplings under different orientations significantly increases the structural information available (Figure 2–2). Therefore, it was attractive to find a second chemically inert and stable orienting medium, which, in contrast to the steric alignment in polyacrylamide gels, yields electrostatic contributions to the weak alignment.

2.2 Charged copolymers as new synthetic alignment media

Magnetically pre-oriented purple membrane fragments (Sass et al., 1999; Sass et al., 2000) and Pf1 phages (Trempe et al., 2002) have previously been polymerized into polyacrylamide gels to introduce negative charges into such gels for the electrostatic alignment of solute biomolecules. We show that a simple copolymerization of negatively charged acrylic acid monomers into a copolymer with acrylamide and subsequent squeezing of the charged gel can also be used for electrostatic alignment, without relying on protein systems of limited stability.

Radical copolymerization of acrylamide and acrylate was performed in high ionic strength buffer at pH 8.2. The incorporation of charged monomers into the polymers with high efficiency was evident from the strong electroosmotic swelling of the gels upon washing in water. Electroosmotic swelling in charged polymers is induced by the counter cations of the acidic groups in the charged gels (Flory, 1953; Horkay et al., 2000). Swelling decreases with the pH around the pK_a of the acrylic acid in the copolymer ($pK_a \approx 4.8$), as the electroosmotic swelling is abolished by the protonation of acidic groups in the gel.

Anisotropy can be introduced by squeezing these polyacrylate/acrylamide copolymer gels like polyacrylamide gels or by drying them on a capillary after swelling to stretch the polymer. Alignment tensors with an $A_{zz,NH}$ of up to 60 Hz are obtained in 2 % (w/v) gels for the alignment of ubiquitin. Due to the electroosmotic swelling, the charged copolymer gels are mechanically stable and easy to handle at gel concentrations down to 0.1 % (w/v), whereas polyacrylamide gels are mechanically unstable below 3 % (w/v).

The alignment is tunable by the addition of salt and is thus electrostatic in nature as expected for a charged orientation medium. The electrostatic alignment is further evidenced by the pronounced correlation (correlation coefficient $r = 0.9$) of the alignment tensor in charged gels to tensors obtained in negatively charged purple membranes and Pf1 phages. Correlation coefficients to steric alignment tensors in

uncharged polyacrylamide gels (r = 0.5) and uncharged DMPC/DHPC bicelles (r = 0.2) are much smaller. Thus, the introduction of charged moieties into polyacrylamide gels yields additional independent structural information about the solute macromolecule.

The applicability of the medium has been tested on the partially folded 17 kDa antibiotic binding domain TipAS. The obtained RDCs are in good agreement ($Q = rms(D^{obs}-D^{calc})/rms(D^{obs}) = 0.39$) with the NMR structure (PDB code 1NY9) determined without use of these RDCs (Kahmann et al., 2003).

A detailed description of the results has been published in the following manuscript.

Original Publication

Meier, S., Häussinger, D. and Grzesiek, S.

Charged acrylamide copolymer gels as media for weak alignment

Journal of Biomolecular NMR **24**, 351-356 (2002).

Journal of Biomolecular NMR, **24:** 351–356, 2002.
KLUWER/ESCOM

Charged acrylamide copolymer gels as media for weak alignment

Sebastian Meier, Daniel Häussinger & Stephan Grzesiek*
Department of Structural Biology, Biozentrum, University of Basel, CH-4056 Basel, Switzerland

Received 15 October 2002; Accepted 6 November 2002

Key words: dipolar couplings, electroosmosis, electrostatic alignment, polyelectrolyte, TipAS, ubiquitin

Abstract

The use of mechanically strained acrylamide/acrylate copolymers is reported as a new alignment medium for biomacromolecules. Compared to uncharged, strained polyacrylamide gels, the negative charges of the acrylamide/acrylate copolymer strongly alter the alignment tensor and lead to pronounced electroosmotic swelling. The swelling itself can be used to achieve anisotropic, mechanical strain. The method is demonstrated for the alignment of TipAS, a 17 kDa antibiotic resistance protein, as well as for human ubiquitin, where alignment tensors with an $A_{ZZ,NH}$ of up to 60 Hz are achieved at a gel concentration of 2% (w/v). The alignment can be modulated by the variation of pH, ionic strength, and gel concentration. The high mechanical stability of the swollen gels makes it possible to obtain alignment at polymer concentrations of less than 1% (w/v).

Residual tensorial couplings have become standard parameters for probing structure and dynamics in biological macromolecules (Tolman et al., 1995; Tjandra and Bax, 1997). The prerequisite to obtain such orientation-dependent information in solution NMR is the weak alignment of biological macromolecules. Several liquid crystalline media have been introduced that achieve weak alignment of solute biomacromolecules by transient steric or electrostatic interactions. Recently, it has been shown that also mechanically strained polyacrylamide gels can be used as the carrier medium for such alignment experiments (Tycko et al., 2000; Sass et al., 2000; Chou et al., 2001). The highly inert gels allow for applications in hostile environments such as the conditions used for protein unfolding (Shortle and Ackerman, 2001). Alignment in the electrically neutral polyacrylamide is mainly caused by steric interactions and results in alignment tensors that are similar to uncharged lipid bicelles. Embedding of oriented, charged purple membranes (Sass et al., 2000) or filamentous phages (Trempe et al., 2002) into the acrylamide gels causes alignment by electrostatic interactions and leads to markedly different orientation tensors and therefore gain of additional information. In this communication we show that such electrostatic alignment can also be achieved in mechanically strained acrylamide/acrylate copolymers. The highly negatively charged gels exhibit strong electroosmotic swelling (Flory, 1953; Horkay et al., 2000) – an effect, which can be used advantageously to apply anisotropic mechanical strain to the gels. The swelling also results in increased mechanical stability. This makes it possible to achieve alignment by mechanical strain at gel concentrations of less than 1% (w/v), whereas uncharged acrylamide gels are difficult to handle at concentrations of less than 4% (w/v).

Preparation of acrylamide/acrylate copolymer gels was performed in analogy to polyacrylamide gels. A stock solution of acrylic acid/bisacrylamide was prepared as an aqueous solution of 29.2% w/v acrylic acid (MERCK) and 0.78% w/v N,N′-methylenebisacrylamide (BIS, SERVA) with the pH adjusted to the pK of acrylic acid (4.25) by the addition of one molar equivalent of NaOH. The commercially available premix of 29.2% w/v acrylamide and 0.78% w/v BIS (APPLICHEM) was used as a second stock solution for an acrylamide/bisacrylamide mixture. Aliquots of both stock solutions were mixed in specific ratios to obtain relative concentrations of acrylic acid in the range of 0–100% of the total acrylic

*To whom correspondence should be addressed. E-mail: Stephan.Grzesiek@unibas.ch

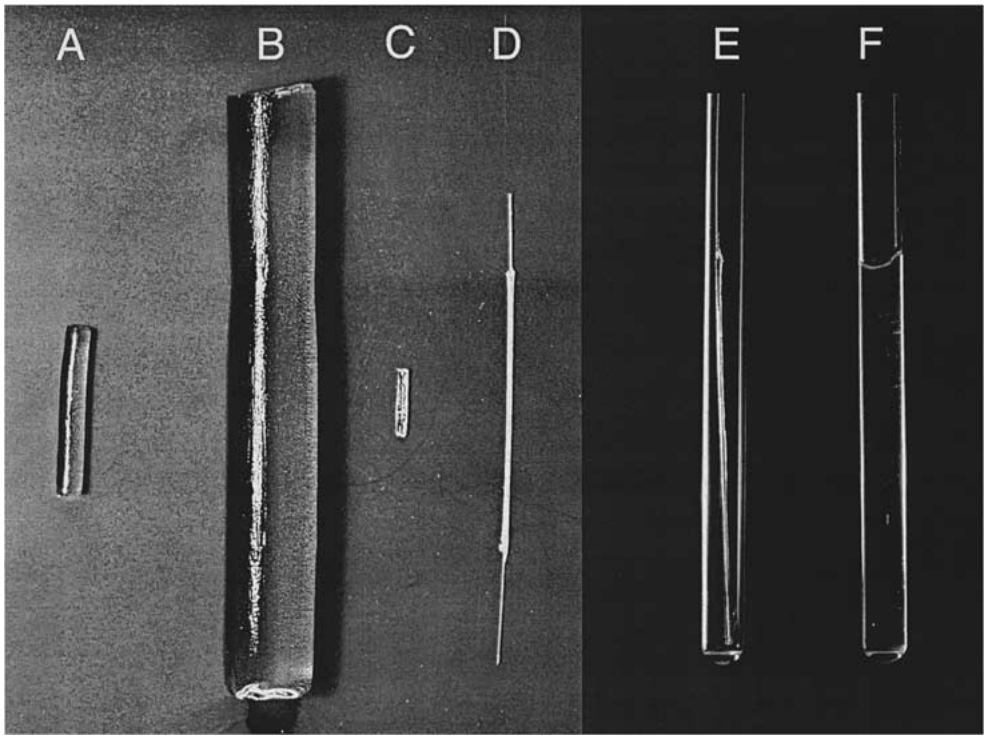


Figure 1. Swelling behavior of a 7% (w/v) 50% acrylate/50% acrylamide copolymer gel and preparation of an anisotropic NMR samples by stretch-drying. A: Gel after polymerization from a 300 μl solution in a 3.5-mm diameter plastic tube. B: Gel after washing in deionized MilliQ water. C: Gel dried on non-adhesive support. D: Gel dried on thin glass capillary stuck through center of the swollen gel. E: Dried gel on glass capillary in 5 mm NMR tube. F: Gel after reswelling with NMR sample solution and removal of the glass capillary.

acid/acrylamide monomers. Typically, these mixtures were diluted to a total monomer concentration of 7–10% w/v by a high ionic strength buffer (10×TBE, i.e., 0.9 M Tris, 0.9 M borate, 0.02 M EDTA, pH 8.2). Polymerization was then started at room temperature by the addition of 1% tetramethylethylenediamine (TEMED) and 0.15% ammonium peroxodisulfate. Polymerizations were either carried out on 300 μl reaction mixtures placed in a plastic tube of 3.5 mm inner diameter that was closed with parafilm on one end, or as 50 μl volumes placed in a teflon tube of 1.7 mm diameter. After the polymerization was completed, the gels were washed at least five times for several hours with a ∼100–1000-fold excess of deionized (MilliQ) water. During the washing, the polyelectrolyte gels undergo very pronounced electroosmotic swelling which leads to a volume increase of up to a hundredfold (see below and Figure 1A,B).

This effect was used to prepare NMR samples of biomacromolecules dissolved in an anisotropic acrylamide/acrylate gel matrix. To this end, a thin glass capillary (diameter ∼0.3 mm) drawn from a glass pipette was stuck through the center of the cylindrical swollen gels (Figure 1B). The gels were then dried in a drying oven at 50 °C on a non-adhesive support such as PVDC (polyvinylidene chloride) household wrapping foil. Drying of the swollen gel was complete within two days. Apparently, the drying and the adhesion to the glass capillary result in an asymmetric stretching of the dried gel in the direction of the glass capillary (Figure 1D). The glass capillary with the dried gel was then placed into a normal 5 mm NMR tube (Figure 1E) and the desired sample solution was added. After the gel had reswollen within the NMR tube, the glass capillary was pulled out (Figure 1F), and the sample was ready for NMR measurements.

As an alternative to this stretch-drying method for achieving anisotropy, the conventional compression method (Tycko et al., 2000; Sass et al., 2000; Chou et al., 2001) was also investigated. Due to lower mechanical stability at high acrylate concentrations (see below), this method was only practical for charged gels with a relative acrylate content of less than 40%. For this conventional method, the gels were dried on a non-adhesive support without the glass capillary. The dried cylindrical gels (Figure 1C) were placed into a

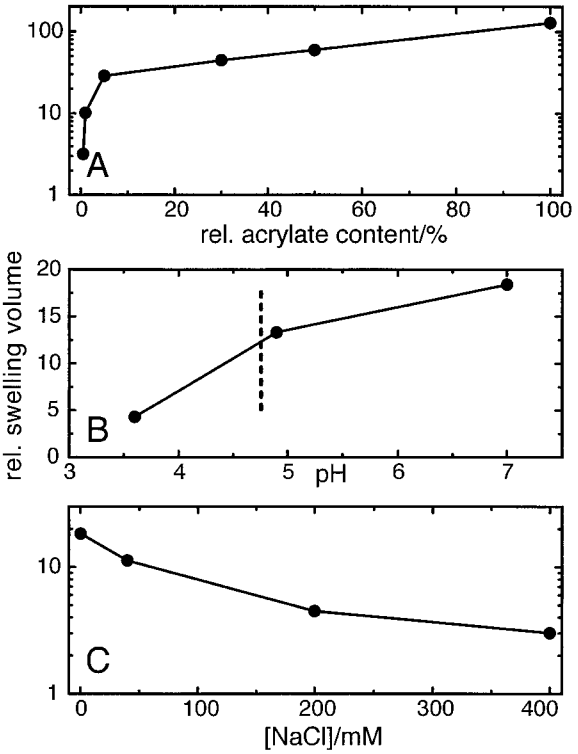


Figure 2. Swelling behavior of a 7% (w/v) acrylate/acrylamide copolymer gel. The relative swelling volume is calculated as the volume of the swollen gel divided by the original polymerization volume. A: Dependence on the relative acrylate content of the gel. Swelling was performed in deionized water. B: Dependence on pH. Swelling was performed in 10 mM sodium phosphate (pH 7.0 and pH 4.9) or ammonium acetate (pH 3.6) buffer solutions. The dashed line indicates the pK_a of polyacrylic acid. C: Dependence on salt content. Swelling was performed in 10 mM phosphate, pH 7.0 with varying amounts of NaCl.

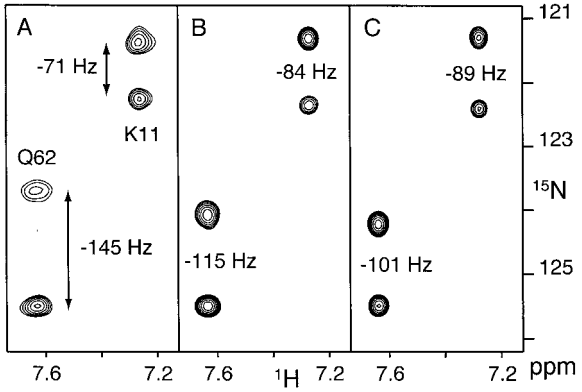


Figure 3. Ionic strength dependence of ubiquitin alignment in stretch-dried 50% acrylate/50% acrylamide copolymer gel. Sample conditions: 0.15 mM ^{15}N-labelled ubiquitin, pH 6.5, final gel concentration 2% (w/v), 298 K. NaCl concentrations are 0 (A), 60 (B) and 240 mM (C), respectively. Amplitudes A_{zz} and rhombicities η of the fitted alignment tensors were: -60.8, -24.1, -12.0 Hz and 0.16, 0.28, 0.39 for 0, 60, and 240 mM NaCl, respectively. The dipolar coupling D is calculated from A_{zz} and η as $D = A_{zz}\left(3\cos^2\theta - 1 + \eta\sin^2\theta\cos 2\phi\right)/2$ where θ and ϕ are the polar angles of the internuclear distance vector in the principal axis system of the orientation tensor.

5-mm NMR Shigemi tube and sample solution added. Compression was achieved by pressure applied via the plunger of the Shigemi tube as described (Tycko et al., 2000; Sass et al., 2000).

The swelling of the charged acrylate/acrylamide copolymers is caused by electroosmotic flow of water into the polymer matrix. Such swelling has been well described for various polyelectrolytes (Vermaas and Hermans, 1948; Katchalsky et al., 1951; Flory, 1953; Horkay et al., 2000). In brief, the concentration of mobile ions is always larger inside the gel than in an exterior solution because of the attraction by the fixed charges. In consequence, the osmotic pressure in the gel is larger than in the external solution, resulting in a net influx of solvent and swelling (Flory, 1953). Figure 2A depicts the relative volume increase from swelling in deionized water for an initial 7% (w/v) polyacrylate/acrylamide gel as a function of the relative acrylate content. For a 50% relative acrylate concentration, a more than 60-fold volume increase is observed. At this swelling ratio, corresponding to a final concentration of about 1 g gel per 1000 g water, the gels are still stable enough such that they can be pierced by the glass capillary or transferred onto another support while retaining their shape. At larger than 50% relative acrylate concentration, the volume further increases. However, the gels rapidly loose their mechanical stability. Presumably, this decrease in stability is caused by a lesser degree of polymerization or by a breaking of covalent bonds due to the increased osmotic pressure. Clearly, the swelling depends on the concentration of the fixed charges of the polyelectrolyte and therefore also on the ionization state (Flory, 1953) of the acrylic acid. A strong decrease of volume is observed for pH values lower than about 5, i.e., when the carboxylate of polyacrylate is titrated to neutrality ($pK_a \sim 4.8$, Figure 2B). Obviously, the ionic strength of the solvent also influences the electroosmotic swelling (Flory, 1953; Horkay et al., 2000; Horkay et al., 2001). For a dilute polymer and intermediate salt concentrations, the swelling ratio q is expected to be proportional to $S^{-3/5}$, where S is the ionic strength of the external solution (Flory, 1953). This behavior is observed approximately in Figure 2C

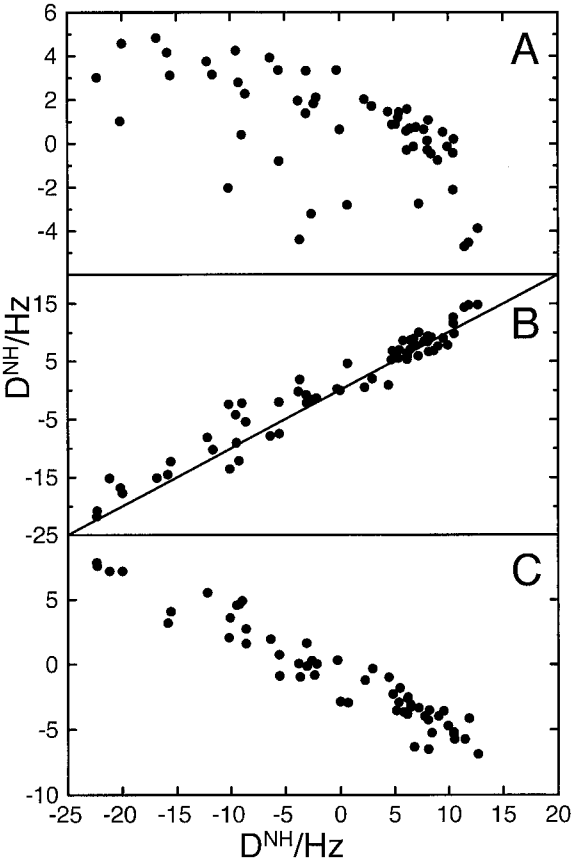


Figure 4. Correlation of RDCs determined for ubiquitin in different media. Horizontal axis: RDCs determined for 0.15 mM ubiquitin, 60 mM NaCl, pH 6.5, in stretched-dried 50% acrylate/50% acrylamide copolymer, final gel concentration 2%. A: Correlation to RDCs determined in compressed uncharged polyacrylamide gel (vertical axis, data are taken from Sass et al., 2000). Correlation coefficient $r = -0.59$. B: Correlation to RDCs (vertical axis) calculated from ubiquitin's crystal structure (1ubi) and a linear fit of the alignment tensor (Sass et al., 1999), $r = 0.97$. NMR quality factor $Q = 0.28$, where Q is defined as the ratio of the rmsd between observed and calculated couplings and the rms of the observed couplings (Cornilescu et al., 1998). C: Correlation to RDCs (vertical axis) measured for ubiquitin in a conventionally (longitudinal) compressed 35% acrylate/65% acrylamide copolymer, final gel concentration 7% (w/v), 0.15 mM ubiquitin, pH 6.5. $r = -0.96$.

for the swelling of a 50% acrylate/acrylamide gel (7% w/v, pH 7) as a function of external NaCl concentration. A fit to the data points (not shown) yields approximately $q \sim S^{-0.50}$. Clearly, at NaCl concentrations of several hundred millimolar, the electroosmotic swelling collapses and the behavior reverts to that of an uncharged gel. For divalent cations, this collapse is reported for concentrations as low as a few millimolar (Horkay et al., 2000, 2001). Therefore, the application of anisotropic strain by electroosmotic swelling is probably limited to cases where cations of higher valency are only required at low concentrations.

The use of the anisotropic polyacrylate/acrylamide gel as an alignment medium for protein NMR was demonstrated on human ubiquitin as well as on TipAS, the 17 kDa antibiotic binding domain of the thiostrepton induced protein A from *streptomyces lividans* (Holmes et al., 1993). Figure 3 depicts small sections of a ^{1}H-coupled ^{1}H-^{15}N HSQC recorded on ubiquitin at pH 6.5 dissolved in a stretch-dried 50% acrylate/acrylamide gel of a final concentration of 2% (w/v). Without the addition of salt (Figure 3A), ^{15}N-splittings of -145 Hz and -71 Hz are observed for amino acids Q62 and K11, corresponding to a residual dipolar coupling (RDC) contribution of -52 and $+22$ Hz, respectively. Upon addition of salt (60 and 240 mM NaCl, Figure 3B,C), the dipolar contribution gradually diminishes to -8 and $+4$ Hz for both amino acids. This dependency confirms the electrostatic character of the residual alignment in the charged gels and proves that this method can still be used at relatively high ionic strengths. As expected, the observed dipolar couplings only show a weak correlation to couplings obtained from the alignment in mechanically strained, neutral polyacrylamide gels (Figure 4A), but correlate well (Figure 4B) with dipolar couplings predicted from ubiquitin's crystal structure (Vijay-Kumar et al., 1987). A negative correlation is observed between couplings determined in the stretch-dried charged gels and in the conventionally compressed charged gels (Figure 4C), thus indicating that the orientation tensor simply inverts when changing the direction of the applied mechanical strain. Table 1 quantitates the correlations between the charged gel alignment tensors for ubiquitin and other alignment tensors obtained in various media. The listed correlation coefficients present a measure for the linear dependence of the alignment tensors (Sass et al., 1999). It is evident that the orientation tensor for the stretch-dried charged gel in the absence of salt has a high correlation (-0.91) to the orientation tensor obtained with the negatively charged purple membranes (Sass et al., 1999), but has only weak correlation to uncharged DMPC/DHPC bicelles (0.21). Therefore, information obtained from uncharged bicelle and charged gel alignment is very complementary. An intermediate correlation (-0.55) is obtained to compressed polyacrylamide gels. Upon increasing the salt concentration, the correlation to the purple membrane orientation decreases, whereas the correlations to the uncharged acrylamide and the uncharged DMPC/DHPC bicelles increase. This clearly

Table 1. Correlation coefficients[a] between ubiquitin alignment tensors in different media

Sample	Charged gel, 0 mM NaCl[b]	Charged gel, 60 mM NaCl[b]	Charged gel, 240 mM NaCl[b]	Compressed charged gel[c]	Uncharged acrylamide gel[d]	DMPC/DHPC bicelles[e]	Purple membrane[e]
Charged gel, 0 mM NaCl	1.00	0.99	0.90	−0.98	−0.55	0.21	−0.91
Charged gel, 60 mM NaCl		1.00	0.95	−0.99	−0.58	0.25	−0.86
Charged gel, 240 mM NaCl			1.00	−0.96	−0.69	0.40	−0.80
Compressed charged gel				1.00	0.56	−0.22	0.87
Uncharged acrylamide gel					1.00	−0.92	0.63
DMPC/DHPC bicelles						1.00	−0.34
Purple membrane							1.00

[a]The correlation coefficient is calculated as the normalized scalar product of the irreducible components of two orientation tensors $<A^1|A^2>/(<A^1|A^1>^{1/2}<A^2|A^2>^{1/2})$ with $<A^1|A^2> := \Sigma_{m=-2,2} A^1_m A^{2*}_m$ (Sass et al., 1999).

[b]Further sample conditions are indicated in the legend to Figure 3.

[c]Anisotropy was achieved by conventional, longitudinal compression via the plunger of the Shigemi tube. Sample consisted of 35% acrylate/65% acrylamide copolymer, final gel concentration 7% (w/v), 0.15 mM ubiquitin, pH 6.5, 298 K. Amplitude A_{zz} and rhombicity η of the fitted alignment tensor were 9.7 Hz and 0.11, respectively.

[d]Data taken from (Sass et al., 2000).

[e]Data taken from (Sass et al., 1999).

indicates that the orientation mechanism changes gradually from mainly electrostatic to steric.

The rather strong alignment observed within the charged gels and their relatively high mechanical stability makes it possible to reduce the gel concentrations well below what is practical for orientation by strained uncharged polyacrylamide gels. The same method of stretch-drying on a glass capillary and reswelling in the sample tube could also be used on gels cast originally from a volume of 50 µl with a diameter of 1.7 mm. After reswelling in a 5-mm Shigemi tube, final gel concentrations of approximately 0.9% (w/v) were reached for 50% and 100% relative acrylate content. Under these conditions, alignment with maximal ^{1}H-^{15}N RDCs of about 10 Hz was still observed (data not shown) for ubiquitin at pH 6.5 with no additional salt. These couplings corresponded to an amplitude A_{zz} of the alignment tensor of −8.5 Hz (50% acrylate) and −11.4 Hz (100% acrylate).

The alignment by the stretched 50% acrylate/50% acrylamide copolymer was further tested on the 17 kDa antibiotic domain TipAS in its apo form. Figure 5A shows the ^{15}N downfield component of a doublet-separated sensitivity-enhanced HSQC (Cordier et al., 1999) for this protein in the copolymer at a final gel concentration of 3% (w/v). The quality of this spectrum is comparable to spectra of TipAS in normal aqueous solution. Note that the large number of relatively weak peaks in mostly random coil positions results from the fact that the first 56 amino acids of this 144 amino acid long protein do not adopt a stable conformation in the absence of antibiotic. For the remaining 86 non-proline residues 72 RDCs could be determined, which give an NMR quality factor Q (Cornilescu et al., 1998) of 0.39 relative to an NMR structure that was determined without the use of these RDCs (Figure 5B).

The spectral quality within the charged gel was further assessed by ^{15}N transverse relaxation times. Average T_2 values for the structured part of the negatively charged TipAS at pH 5.9 (pI = 4.6) were 74 ms both within the acrylate/acrylamide copolymer and in normal aqueous buffer. This indicates that within the error limits the rotational diffusion was not hindered by the interaction with the charged gel. For electrically neutral ubiquitin at pH 6.5 (pI = 6.5) and low ionic strength, a reduction in $^1H^N$ T_2s of about 20% was observed in the presence of 50% acrylate/50% acrylamide, final gel concentration 4% (w/v). Apparently rotational diffusion is already slightly hindered by the strong electrostatic interactions under these conditions. It should be noted that larger reductions in T_2 are observed for ubiquitin under identical conditions in the presence of negatively charged filamentous phages Pf1 (Hansen et al., 1998) or purple membranes (S.M., unpublished observations).

In summary, we have shown that negative charges can be introduced easily into mechanically strained acrylamide gels by copolymerization with acrylic acid. The resulting alignment is electrostatic in nature and deviates strongly from the alignment achieved by mechanically strained polyacrylamide gels. The electroosmotic swelling of the charged gels can be used advantageously to achieve mechanically induced

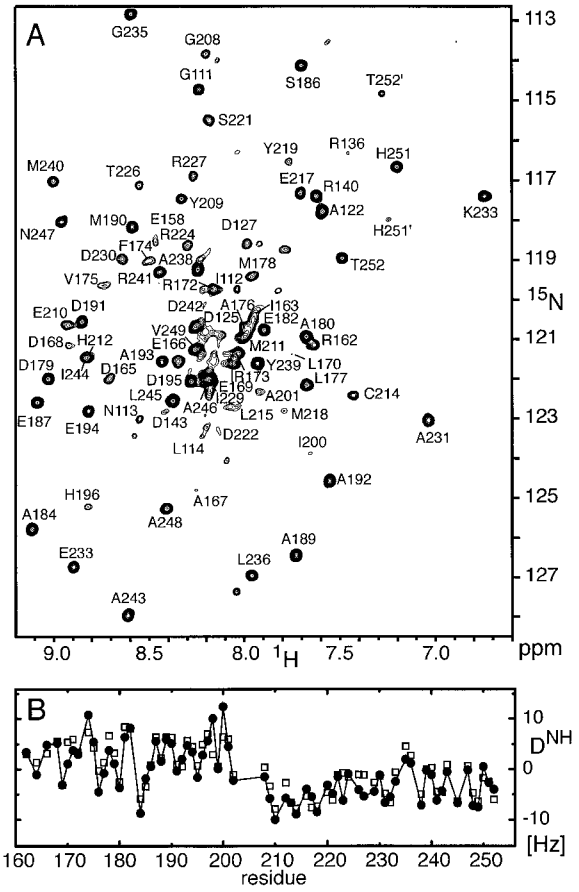


Figure 5. Use of anisotropic polyacrylate/acrylamide gel for determination of RDCs in TipAS. A: Downfield component of a ^{15}N-DSSE spectrum (18.7 T) recorded on 0.2 mM TipAS (0.2 mM), 10 mM phosphate, pH 5.9, in stretch-dried 50% acrylate/50% acrylamide copolymer, final gel concentration 3% (w/v), 298 K. B: Comparison of measured (filled circles) and theoretical (open squares) RDCs calculated from the NMR structure of the folded C-terminal part of TipAS. The NMR structure was determined without use of RDCs of the charged gel. The structure is all-helical leading to a characteristic periodicity in the dipolar couplings for many parts of the sequence.

anisotropy especially at low gel concentrations. Various possibilities are currently under investigation to copolymerize positively charged groups into the polyacrylamide matrix.

Acknowledgements

We are grateful to Martin Allan and Jan Kahmann for providing the TipAS protein and to Hans-Jürgen Sass for helpful discussions. This work was supported by SNF grant 31-61'757.00 to S.G.

References

Chou, J.J., Gaemers, S., Howder, B., Louis, J.M. and Bax, A. (2001) *J. Biomol. NMR*, **21**, 377–382.

Cordier, F., Dingley, A.J. and Grzesiek, S. (1999) *J. Biomol. NMR*, **13**, 175–180.

Cornilescu, G., Marquardt, J.L., Ottiger, M. and Bax, A. (1998) *J. Am. Chem. Soc.*, **120**, 6836–6837.

Flory, P.J. (1953) *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, NY.

Hansen, M.R., Mueller, L. and Pardi, A. (1998) *Nat. Struct. Biol.*, **5**, 1065–1074.

Holmes, D.J., Caso, J.L. and Thompson, C.J. (1993) *Embo J.*, **12**, 3183–3191.

Horkay, F., Tasaki, I. and Basser, P.J. (2000) *Biomacromolecules*, **1**, 84–90.

Horkay, F., Tasaki, I. and Basser, P.J. (2001) *Biomacromolecules*, **2**, 195–199.

Katchalsky, A., Lifson, S. and Eisenberg, H. (1951) *J. Polymer Sci.*, **7**, 571.

Sass, H.J., Musco, G., Stahl, S.J., Wingfield, P.T. and Grzesiek, S. (2000) *J. Biomol. NMR*, **18**, 303–309.

Sass, J., Cordier, F., Hoffmann, A., Cousin, A., Omichinski, J.G., Lowen, H. and Grzesiek, S. (1999) *J. Am. Chem. Soc.*, **121**, 2047–2055.

Shortle, D. and Ackerman, M.S. (2001) *Science*, **293**, 487–489.

Tjandra, N. and Bax, A. (1997) *Science*, **278**, 1111–1114.

Tolman, J.R., Flanagan, J.M., Kennedy, M.A. and Prestegard, J.H. (1995) *Proc. Natl. Acad. Sci. U.S.A.*, **92**, 9279–9283.

Trempe, J.F., Morin, F.G., Xia, Z., Marchessault, R.H. and Gehring, K. (2002) *J. Biomol. NMR*, **22**, 83–87.

Tycko, R., Blanco, F.J. and Ishii, Y. (2000) *J. Am. Chem. Soc.*, **122**, 9340–9341.

Vermaas, D. and Hermans, J.J. (1948) *Rec. trav. chim.*, **67**, 983.

Vijay-Kumar, S., Bugg, C.E. and Cook, W.J. (1987) *J. Mol. Biol.*, **194**, 531–544.

2.3 Long-range RDCs at high accuracy in perdeuterated proteins

Classical NMR spectra detect proton distances without giving direct information on heavy atom positions. X-ray structures of proteins, on the other hand, are derived from the scattering of X-ray beams by electrons. Thus, X-ray crystallography mainly detects heavy atom positions, but normally yields only poor information on hydrogen atoms. Neutron diffraction could yield information on proton and heavy atom positions in proteins (Schoenborn, 1969; Wlodawer et al., 1984), but suffers from experimental obstacles due to the need of large crystals to overcome the weak interaction of neutrons with matter and the low flux of neutron sources. Thus, the precise position of protons relative to heavy atoms has been difficult to assess for many years. However, the precise positioning of the amide proton is a prerequisite for DFT calculations of the still poorly understood $^{1}H^{N}$ chemical shift and its correlation with hydrogen bond strength and geometry (Barfield, 2002).

RDCs between ^{1}H and ^{13}C or ^{15}N nuclei yield novel high precision information on the relative position of protons and isotope labeled heavy atoms in biomacromolecules. Crystal structures have thus been refined by the use of RDC data (Cornilescu et al., 1998; Ulmer et al., 2003) to improve the proton coordinates. In this part of the thesis we have pursued the measurement of as many RDCs to the amide proton as possible to precisely define amide proton positions. This can either be achieved by measuring one or two different RDCs in many alignment media (Figure 2–2), or by measuring many different RDCs in few alignment media (Tolman et al., 2001). The measurement of many RDCs in few media is, however, more practicable, considering the fact that fewer samples are needed and that most proteins are not easily aligned in several alignment media.

In the following we describe the measurement of long-range $^{1}H^{N}$-$^{1}H^{N}$ and $^{1}H^{N}$-^{13}C RDCs. The results have been published in (Meier et al., 2003).

2.3.1 Basic schemes for the detection of RDCs

Dipolar and scalar couplings between nuclei can be measured by various approaches. In J-resolved spectroscopy, couplings are obtained as splittings in the directly or indirectly detected dimension of correlation spectra. The use of spin–state-selective filtering (Meissner et al., 1997) allows to separate α- and β- states into different sub-spectra, thus minimizing resonance overlap (Figure 2–3). One-bond RDCs were obtained with J-resolved spectra throughout this thesis by omitting decoupling during chemical shift detection, or by spin state selected IPAP (Ottiger et al., 1998) and DSSE (Cordier et al., 1999) spectra to separate the doublet components.

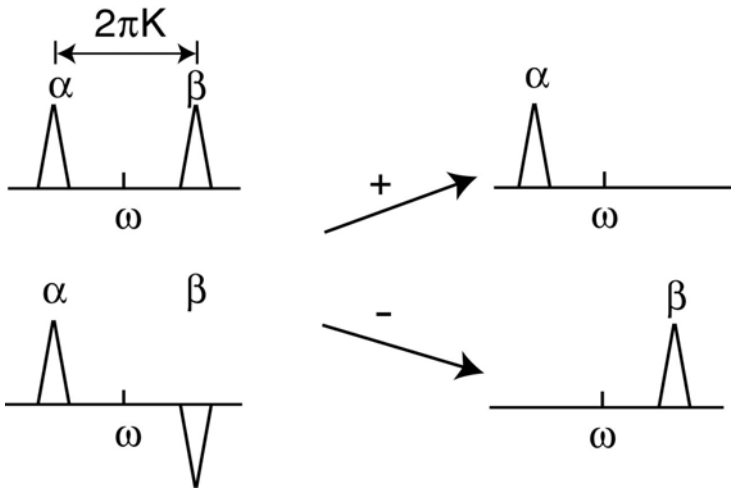


Figure 2–3. Spin state selective filters allow separating multiplet components into different sub-spectra, thus avoiding signal overlap.

Exclusive correlation spectroscopy (E-COSY) and double quantum/zero quantum spectra additionally allow the simultaneous measurement of couplings between different spin pairs from one multiplet (Griesinger et al., 1987; O'Neil-Cabello et al., 2004).

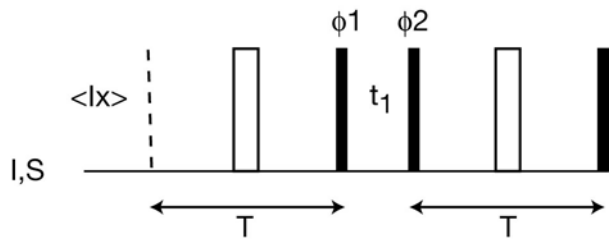


Figure 2–4. Quantitative J-correlation scheme based on the COSY. ϕ1 = x,-x; ϕ2 = x,x,-x,-x; receiver = 4x. The last 90° pulse acts as a purge pulse to convert antiphase terms to not detected multiple quantum terms.

Quantitative J-correlation experiments (Vuister and Bax, 1993) measure couplings by the ratio of signal intensities in a cross experiment and a diagonal or reference experiment. During a constant transfer time T, magnetization is dephased by scalar and dipolar coupling ($K = J+D$) between spins I and J (Figure 2–4). Spin operators develop as:

$$I_x \xrightarrow{T} I_x \cos(\pi K_{IS} T) + 2 I_y S_z \sin(\pi K_{IS} T) \xrightarrow{90\phi_1} I_x \cos(\pi K_{IS} T) - 2 I_z S_y (\pi K_{IS} T).$$

The frequencies ω_I and ω_S are then labeled by chemical shift evolution during t_1 in the homonuclear case. By the second 90° pulse ($90_{\phi 2}$) one of the resulting quadrature components is selected and antiphase S-magnetization is reconverted to antiphase I magnetization. The antiphase I magnetization refocuses to inphase I magnetization during the second interval T:

$$\begin{aligned} & I_x \cos(\pi K_{IS} T) - 2 I_z S_y \sin(\pi K_{IS} T) \\ & \xrightarrow{\omega \cdot t_1} I_x \cos(\pi K_{IS} T)\cos(\omega_I t_1) + I_y \cancel{\cos(\pi K_{IS} T)\sin(\omega_I t_1)} \\ & -2 I_z S_y \sin(\pi K_{IS} T)\cos(\omega_S t_1) + 2 I_z S_x \cancel{\sin(\pi K_{IS} T)\sin(\omega_S t_1)} \\ & \xrightarrow{90\phi_2} \xrightarrow{T} \xrightarrow{90_x} I_x \{\cos^2(\pi K_{IS} T)\cos(\omega_I t_1) - \sin^2(\pi K_{IS} T)\cos(\omega_S t_1)\} \quad . \end{aligned}$$

Non-detectable or suppressed components are crossed out in the above scheme. The magnetization transfer from I to S results in a signal, which is frequency-labeled with ω_S and has an intensity proportional to $-\sin^2(\pi K_{IS} T)$, whereas the diagonal signal frequency-labeled with ω_I is proportional to $\cos^2(\pi K_{IS} T)$.

If S is a heteronucleus, either the I_x term or the $2I_zS_y$ term is selected after an HMQC- or an HSQC-based transfer. The terms I_x and $2I_zS_y$ are thus separated into different reference and cross experiments, respectively.

In the case of coupling to several spins S_k, the additional nuclei act as passive nuclei, both in the cross and reference experiment, resulting in signal intensities

$$I_{Sk} = -A\sin^2(\pi K_{ISk}T)\prod_{l\neq k}\cos^2(\pi K_{ISl}T)$$

and

$$I_I = A\cos^2(\pi K_{ISk}T)\prod_{l\neq k}\cos^2(\pi K_{ISl}T),$$

where A is a constant incorporating factors like relaxation and data processing. Contributions of the passive couplings as well as other factors (A) thus cancel in the ratio of intensities

$$\frac{I_{Sk}}{I_I} = -\tan^2(\pi K_{ISk}T).$$

The evaluation of quantitative J experiments requires only the quantification of cross and diagonal intensities, resulting in a straightforward and simple determination of couplings. The lower limit of detectable couplings is given by the signal to noise ratio, which also determines the experimental error.

As the dipolar coupling is transferred through space, the detection of RDCs does not have to rely on the transfer pathways via scalar one- and two-bond couplings applied for the resonance assignment of proteins, but a direct long-range transfer is feasible (Hansen et al., 1998b; Tian et al., 2000). Whereas the identity of the coupling partner is immediately clear for the measurement of one- and two-bond dipolar couplings in the protein backbone, the measurement of long-range RDCs practically depends on a method, which obtains resonance frequencies and thus the identity of the dipolar coupled spin. Thus, quantitative J measurements are the method of choice for the measurement of long-range RDCs. In addition, residual dipolar couplings to several coupled spins can be resolved by quantitative J schemes. As mentioned above, the presence of several passive coupling partners does not impede the quantification of couplings in this approach.

2.3.2 Applications to perdeuterated proteins

The measurement of long-range RDCs in proteins is hindered by the fact that far-reaching transfers are truncated due to the high proton density in proteins and the high magnetic moment of protons. Thus, most of the magnetization is dephased by nearest neighbor ^{1}H spins. In the following we show that this problem can be overcome by using perdeuterated proteins with their peptide groups reprotonated in H_2O (LeMaster and Richards, 1988; Torchia et al., 1988). Alternatively, frequency-selective recoupling schemes of unwanted couplings have been used (Wu and Bax, 2002b). In analogy, the measurement of long-range dipolar couplings between carbons in the solid state (d = 2–7 Å) has been known to suffer from such "dipolar truncation effects" (Hodgkinson and Emsley, 1999) and has been alleviated by the dilution of ^{13}C spins (Castellani et al., 2002; Castellani et al., 2003) in isotope labeling schemes or by selective dipolar recoupling schemes (Jaroniec et al., 2001).

2.3.2.1 $^{1}H^{N}\leftrightarrow{}^{1}H^{N}$ RDCs

Residual dipolar couplings between amide protons ($^{1}H^{N}\leftrightarrow{}^{1}H^{N}$) were obtained for the perdeuterated, amide protonated B1 domain of streptococcal protein G (henceforward referred to as protein G) aligned in Pf1 phages and for perdeuterated, amide protonated human ubiquitin aligned in DMPC/DHPC bicelles. $^{1}H^{N}\leftrightarrow{}^{1}H^{N}$ dipolar couplings were measured with a COSY-type quantitative J scheme in an experiment normally used for the measurement of $^{3}J_{HNH\alpha}$ couplings (Vuister and Bax, 1993). The scalar coupling between these amides (five and more bonds) is negligible. Thus, the experiment demands only one measurement of couplings in the anisotropic phase. The RDC is obtained as

$$D_{HNiHNj} = \frac{\arctan\sqrt{-\left(\frac{I_{cross}}{I_{diag}}\right)}}{2\pi T},$$

where a de-/rephasing time of the transferred magnetization 2·T = 47 ms was applied.

Quantitative J-correlation does not yield the sign of the RDC. Ambiguities were resolved according to the value predicted from the refined protein G (1IGD) and ubiquitin (1D3Z) structures. In principle, the signs could also be obtained by more sophisticated experimental methods. However, this results in a reduced sensitivity (Wu and Bax, 2002a). Structure calculations based on RDCs with unknown sign, on the other hand, only lead to a minor deterioration of the structural quality (Tjandra et al., 2000; Wu and Bax, 2002a) and are thus well suited for incorporating RDCs measured with quantitative J schemes into structure determinations.

A total of 110 $^1H^N \leftrightarrow ^1H^N$ couplings for 3 mM ubiquitin and 79 $^1H^N \leftrightarrow ^1H^N$ couplings for 2.5 mM protein G (Meier et al., 2003) were obtained. In this experiment, RDCs between amide spins up to 7.7 Å apart (Figure 2–5) could be detected. As RDCs are not affected by spin diffusion, their structural interpretation is precise and straightforward.

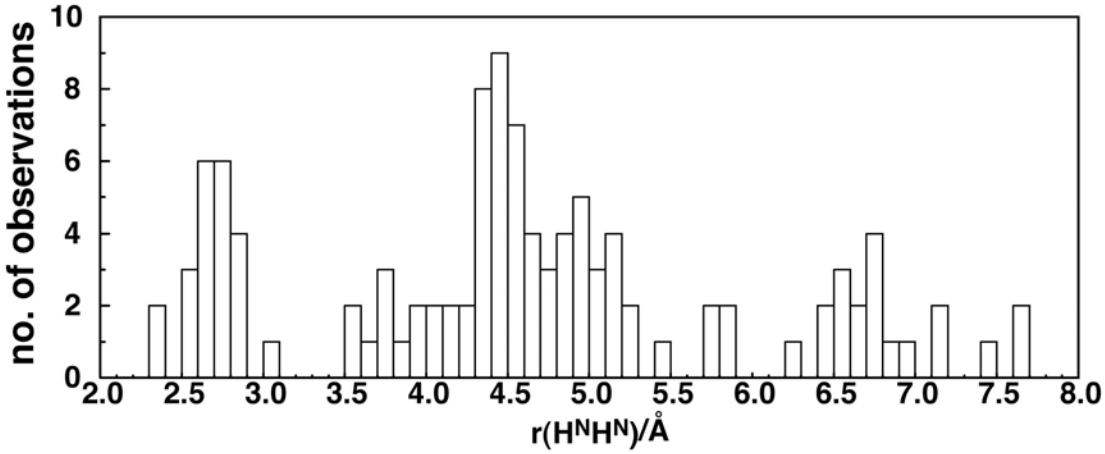


Figure 2–5. Number of $^1H^N \leftrightarrow ^1H^N$ dipolar couplings obtained for different distances between amide protons in ubiquitin. Distances around 2.8 Å are nearest neighbors in the α-helix, whereas distances around 4.5 Å are sequential amide protons in β-sheets and are difficult to detect via NOEs. Dipolar couplings above 4.7 Å are all couplings to non-sequential $^1H^N$ spins and correspond to the next "shell" of protons.

The measured RDCs are in excellent agreement with values expected from the alignment tensor, which was obtained by fitting $^1D_{NH}$ RDCs to the respective crystal structures assuming an N-H distance of 1.02 Å (Figure 2–6). The refined NMR structure of ubiquitin (1D3Z) obtained by the use of five one-bond and two-bond RDCs (Cornilescu et al., 1998) is in better agreement with the experimental long-range RDC data than the 1.8 Å crystal structure (1UBQ) (Vijay-Kumar et al., 1987) as

seen in Figure 2–6. The rmsd between measured and calculated RDCs is 1.92 Hz for the crystal structure and 0.98 Hz for the refined NMR structure.

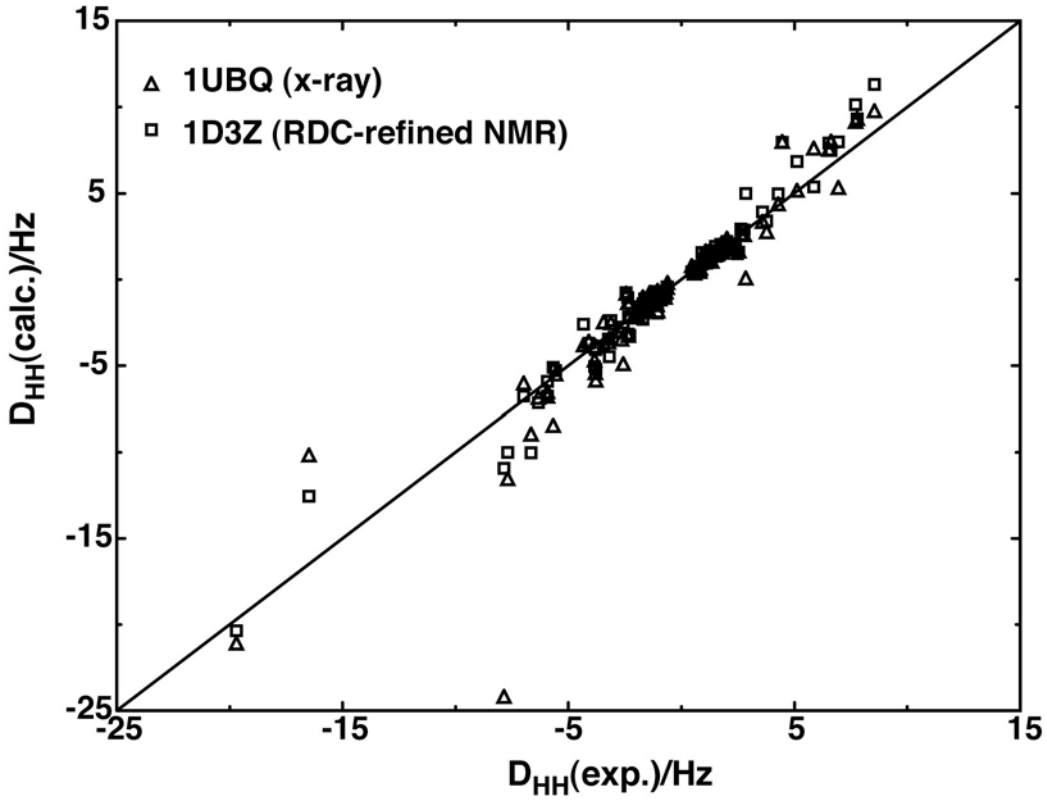


Figure 2–6. Correlation between measured and theoretical RDCs obtained for ubiquitin in DMPC/DHPC bicelles aligned to $A_{zz,NH}$ = 37.4 Hz. The agreement with theoretical values for the crystal structure 1UBQ (Vijay-Kumar et al., 1987) and the refined solution structure 1D3Z (Cornilescu et al., 1998) is shown. Theoretical values are calculated from an alignment tensor determined with $^{1}D_{NH}$ couplings as well as from H^N–H^N distances and orientations in the PDB-files.

Far-reaching RDCs (smaller values of D_{HH} in Figure 2–6) prove to exhibit smaller relative deviations from the expected values than short-range RDCs (Figure 2–6). This is due to the angular dependence of the RDC, where errors in the position of a nucleus result in smaller changes of the internuclear vector orientation for longer internuclear vectors (Meiler et al., 2001) and due to the fact that errors of atom distances result in smaller changes of the RDC for longer distances.

2.3.2.2 $^{1}H^{N}\leftrightarrow{}^{13}C$ RDCs

An improved determination of amide proton positions in the context of known heavy atom positions depends on relative structural information between protons and heavy atoms. Therefore, RDCs to heavy atoms are more efficient than $^{1}H^{N}\leftrightarrow{}^{1}H^{N}$ RDCs for the refinement of proton positions in crystal structures. In order to obtain this detailed structural information $^{1}H^{N}\leftrightarrow{}^{13}C$ dipolar couplings were measured with a quantitative J $^{1}H^{N}$-^{13}C HSQC scheme.

Measurements were performed on perdeuterated protein G aligned in Pf1 phages. The pulse sequence consists of two sequential HSQCs, a quantitative J $^{1}H^{N}$-^{13}C HSQC for the determination of $^{1}H^{N}\leftrightarrow{}^{13}C$ couplings and a second ^{1}H-^{15}N HSQC to label the amide group with the nitrogen chemical shift. Couplings to carbonyl and aliphatic ^{13}C nuclei were obtained in two separate experiments due to the limited excitation bandwidth of radiofrequency pulses. Separate cross and diagonal experiments were carried out, selecting for transferred and non-transferred signals, respectively, as described above. The signal intensities after refocusing the coupling are $I^{cross}\propto\sin^{2}(2\pi TK_{HC})$ and $I^{diag}\propto\cos^{2}(2\pi TK_{HC})$, respectively, where K_{HC} is the sum of the scalar and dipolar coupling and $2\cdot T$ is the transfer time during the INEPT step of the ^{1}H-^{13}C-HSQC. Measurements on an oriented sample yield $K_{HC}=J_{HC}+D_{HC}$, whereas measurements on an isotropic sample yield J_{HC} for the extraction of dipolar couplings D_{HC}.

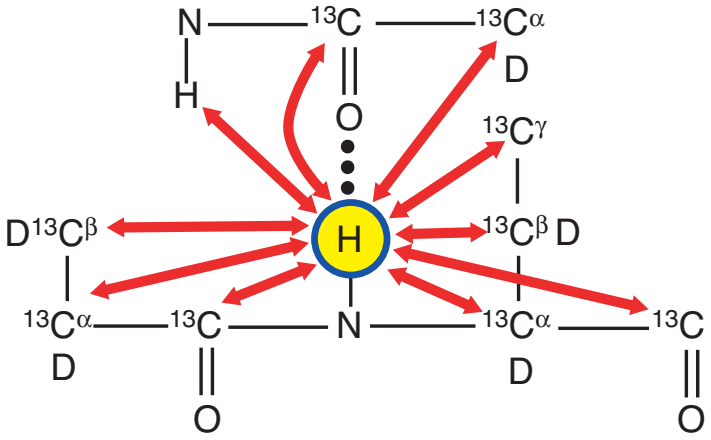


Figure 2–7. Residual dipolar couplings measured in perdeuterated protein G with quantitative J-schemes.

By this means RDCs from backbone amides for example to $^{13}C'$ and $^{13}C^{\alpha}$ nuclei across hydrogen bonds, nearby ^{13}C nuclei in turns and sidechain carbons were measured (Figure 2–7). In total, 99 $^{1}H^{N}\leftrightarrow{}^{13}C^{aliphatic}$, 85 $^{1}H^{N}\leftrightarrow{}^{13}C'$ and 79 $^{1}H^{N}\leftrightarrow{}^{1}H^{N}$ long-range dipolar couplings were obtained for protein G. The statistical error estimated from the reproducibility of the experiments (0.14 Hz for an $A_{zz,NH} = 25.4$ Hz) translates to errors in internuclear distance measurements of only ~0.03 Å for H-C distances of 2 Å, assuming the direction of the internuclear vector is known.

Together with easily obtainable $^{1}D_{NH}$, $^{1}D_{C\alpha H\alpha}$, $^{1}D_{NC'}$, $^{1}D_{C\alpha C'}$ and $^{1}D_{C\alpha C\beta}$ dipolar couplings, in total 555 RDCs were measured for the 66 residues (56 of which are ordered) of perdeuterated protein G from one alignment medium. Roughly half of these RDCs are obtained with the described quantitative J schemes. In concert with a similar number of RDCs measured upon steric alignment in lamellar Lα phases (Ruckert and Otting, 2000), this vast amount of dipolar coupling data will be used to study structure and dynamics with more than 16 residual dipolar couplings per residue.

2.3.2.3 RDCs and high-resolution X-ray structure

Fitting the measured RDCs to the 1.1 Å crystal structure of protein G (Derrick and Wigley, 1994) (1IGD) shows systematic deviations from the structure, clustering mainly into the loop regions (Figure 2–8). These deviations exceed by far the estimated experimental error and correlate well with crystallographic B factors in the crystal structure (Figure 2–8).

RDCs depend both on the molecular structure and on the order parameter of the internuclear vector in the structure. Thus, differences between RDC data and the X-ray structural model are either due to a false determination of atom positions in the flexible parts of the crystal structure conformer, or due to structural disorder, resulting in reduced order parameters and consequently in reduced RDCs in loops. Deviations for the most flexible residue in protein G, G46, (Barchi et al., 1994) in part result

from the stronger experimental than predicted $^1D_{NH}$, opposite to the trend expected for scaling down the coupling with a reduced order parameter. Thus, proton positions in the solution state structure are indeed not optimally defined by the crystal structure conformer.

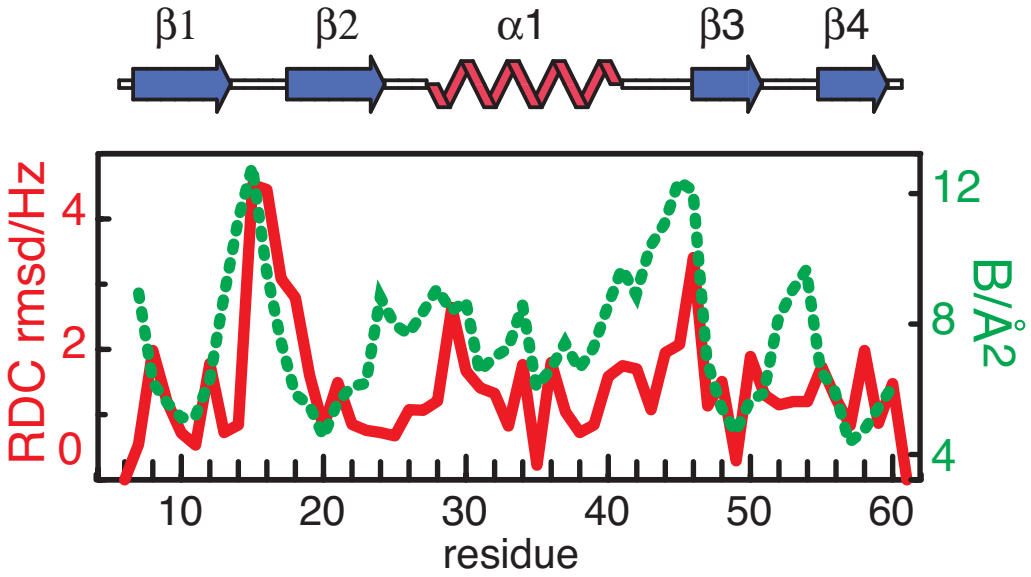


Figure 2–8. Correlation between rmsd values of residual dipolar couplings (dotted line) fitted to the high-resolution crystal structure of protein G (1IGD) and crystallographic B-factors (solid line). RDCs show that at least two turns do not adopt the conformation of the static X-ray structure.

A recent study on protein G has shown that the refinement of 1IGD by C^α–C′ and N–C′ dipolar couplings increases the agreement between observed and predicted N–H and C^α–H^α dipolar couplings (Ulmer et al., 2003). Likewise, neutron diffraction data have pointed out that the coordinates of protons in disordered protein regions often deviate by more than 0.2 Å from the positions predicted by high-resolution X-ray crystallography (Engler et al., 2003). This validates our finding that the agreement between experimental and predicted dipolar couplings is not limited by the dynamics of the protein, but can be improved by adjusting details of the static structure. Remaining discrepancies between the adjusted structure and the measured RDCs can be removed by applying simple motional models (Ulmer et al., 2003; Clore and Schwieters, 2004).

Original Publication

Meier, S., Häussinger, D., Jensen, P., Rogowski, M., and Grzesiek, S.

High-accuracy residual $^{1}H^{N}$-^{13}C and $^{1}H^{N}$-$^{1}H^{N}$ dipolar couplings in perdeuterated proteins

Journal of the American Chemical Society **125**, 44-45 (2003).

J|A|C|S
COMMUNICATIONS

Published on Web 12/10/2002

High-Accuracy Residual $^1H^N-^{13}C$ and $^1H^N-^1H^N$ Dipolar Couplings in Perdeuterated Proteins

Sebastian Meier, Daniel Häussinger, Pernille Jensen, Marco Rogowski, and Stephan Grzesiek*

Biozentrum, University of Basel, Basel 4056, Switzerland

Received September 29, 2002; E-mail: stephan.grzesiek@unibas.ch

Residual tensorial couplings of weakly aligned molecules have become an invaluable source of structural and dynamical information in high-resolution NMR.[1] Most commonly, this information is extracted from one-bond residual dipolar couplings (RDCs). The size of RDCs depends on the direction of the internuclear distance vector, its length r, its order parameter S, the magnetic moments of the nuclei, and the degree of alignment. Although in principle, long-range RDCs could define the molecular geometry more efficiently, their use has not been widespread, because the $1/r^3$ dependence of dipolar couplings reduces the signal size and therefore sensitivity and introduces ambiguities in cases where the internuclear distance is not determined by the covalent geometry. However, for small to medium-sized biomacromolecules, a significant part of the sensitivity problem is not caused by incoherent signal losses from relaxation during the transfer delays, but by coherent broadening from multiple residual dipolar couplings to protons.

In this communication we show that perdeuteration followed by reprotonation of labile hydrogen positions[2,3] greatly alleviates this problem for the observation of long-range RDCs from amide protons ($^1H^N$) to surrounding ^{13}C as well as $^1H^N$ nuclei. Very recently, Wu and Bax[4] have proposed a similar approach for the enhancement of $^1H^N \leftrightarrow {}^1H^N$ RDCs. For small perdeuterated proteins, a large number (up to 10 in protein G) of such RDCs to ^{13}C and $^1H^N$ can be observed from individual amide protons with high accuracy, thus defining individual internuclear distances to within few picometers.

In contrast to experiments for the determination of one- and two-bond $^1H^N \leftrightarrow {}^{13}C$ RDCs,[5] the long-range couplings were measured by a quantitative J-correlation experiment (Figure 1) that detects frequencies of both coupling partners. The three-dimensional (3D) experiment consists of two concatenated $^1H-{}^{13}C/{}^1H-{}^{15}N$ HSQCs. In brief, proton magnetization H^N_y present at point *a* is transferred by dipolar and scalar couplings ($K_{HC} = J_{HC} + D_{HC}$) into carbon antiphase magnetization $H^N_zC_y$ at point *b*. After a frequency labeling period t_1, the carbon antiphase magnetization is transferred into the usual $H^N_zN_y$ magnetization for the $^1H-{}^{13}N$ HSQC at point *c*. The transfer scheme gives rise to signals at frequency positions (ω_C, ω_N, ω_H) with intensities proportional to $\sin^2(2\pi K_{HC}T)$. A second, 2D reference experiment is carried out where the phase cycle for $\phi_{1,2,3}$ is changed for the selection of in-phase H^N_z magnetization at point *b*. The t_1-period is not sampled, and the experiments detects signals at frequency positions (ω_N, ω_H) with intensities proportional to $\cos^2(2\pi\ K_{HC}T)$. The coupling constant K_{HC} can then be extracted from the ratio of cross and reference peak intensities after a suitable correction for the 2- and 3D acquisition schemes has been applied. This procedure is conceptually similar to the 2D quantitative $^1H-{}^{13}C$-HMQC,[6] and further details are given in the Supporting Information.

The high accuracy of the method was demonstrated on weakly aligned, perdeuterated protein G for which a 1.1 Å crystal structure

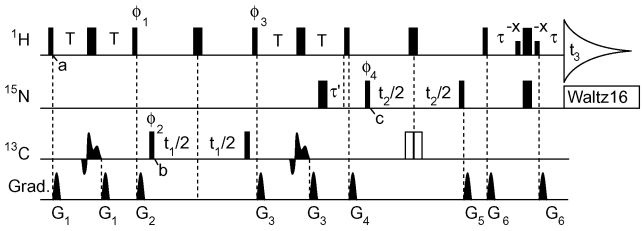


Figure 1. Pulse sequence of the quantitative HCN-HSQC. Narrow and wide pulses denote 90° and 180° flip angles, respectively, and unless indicated the phase is *x*. Further details are described in the Supporting Information.

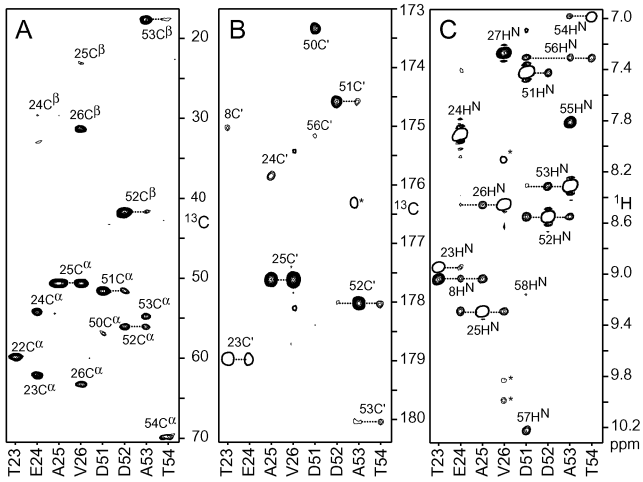


Figure 2. Cross sections extracted from 800 MHz 3D experiments for the detection of $^1H^N$ RDCs to aliphatic ^{13}C (A), carbonyl ^{13}C (B), and $^1H^N$ (C) nuclei. Negative intensities are shown as single thick contour lines. Asterisks indicate cross-peaks resulting from overlap of NH resonances or incomplete side chain deuteration. Experimental conditions: 2.2 mM $^{15}N/{}^{13}C/{}^2H$ protein G, pH 5.6, 98% H_2O/2% D_2O, 30 mg/mL Pf1,[9] 25 °C. (A) aliphatic HCN-HSQC. T = 10 ms; data matrixes consisted of 110* × 20* × 768* points; $t_{1,2,3,max}$ = 8.8, 24, 64 ms; t_{exp} = 32 h. (B) carbonyl HCN-HSQC. T = 10 ms; 50* × 20* × 768* data points; $t_{1,2,3,max}$ = 30, 24, 64 ms; t_{exp} = 14 h. (C) ^{15}N-edited $^1H-{}^1H$ COSY-HMQC. $^1H^N \leftrightarrow {}^1H^N$ transfer time = 47 ms; 72* (^1H) × 72* (^{15}N) × 768* ($^1H^N$) data points; $t_{1,2,3,max}$ = 23, 50, 83 ms; t_{exp} = 38 h.

(1IGD)[7] exists. Due to the limited ^{13}C RF strength, two HCN-HSQC experiments were carried out, optimized separately (Supporting Information) for the detection of $^1H^N$ RDCs to aliphatic and carbonyl ^{13}C nuclei. Figure 2A,B shows cross sections from these 3D experiments extracted at the $^1H^N-{}^{15}N$ frequency positions of residues T23−V26 and D51−T54. Clearly visible are a large number of correlations to $^{13}C^\alpha$, $^{13}C^\beta$, and $^{13}C'$ nuclei of the same and the preceding amino acid; in some cases, the correlations even extend across the hydrogen bond (T23$H^N \leftrightarrow$ Y8C′, D51$H^N \leftrightarrow$ T56C′), to the second next residue (T54$H^N \leftrightarrow$ D52C′), or to $^{13}C^\gamma$ resonances (data not shown). In addition to these long-range $^1H^N \leftrightarrow {}^{13}C$ RDCs, a large number of $^1H^N \leftrightarrow H^N$ RDCs (Figure 2C)

10.1021/ja028740q CCC: $25.00

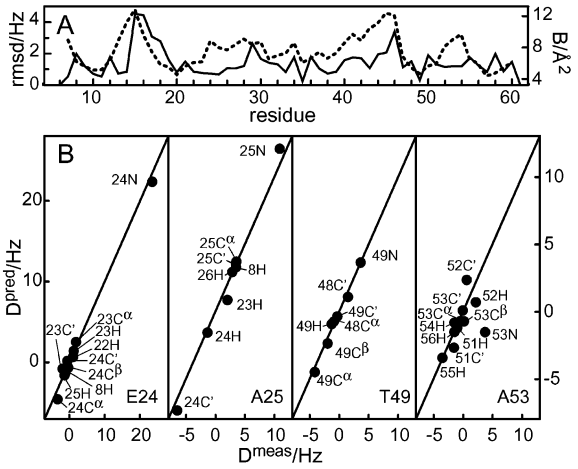


Figure 3. Accuracy of RDCs to $^1H^N$ in protein G. (A) rms deviations between measured and predicted RDCs per single residue (solid line) and crystallographic *B*-factor (dotted line). (B) comparison of measured and predicted RDCs for residues E24, A25, T49, A53.

could be detected by a water flip-back 3D ^{15}N-edited ^{1}H−^{1}H COSY-HMQC,[8] which is conventionally used for the quantification of scalar $^3J_{HNH\alpha}$ couplings. The observed $^1H^N \leftrightarrow {}^1H^N$ correlations span distances of up to 7.2 Å.

To separate dipolar and scalar parts of the coupling constants, a second set of quantitative HCN-HSQCs was carried out on an isotropic sample of protein G. The dipolar couplings D_{HC} were then calculated as the difference of the two measurements. This procedure is not necessary for K_{HNHN}, since J_{HNHN} is negligible. Quantitative J-correlation is insensitive to the sign of the couplings constants. Therefore the method yields two possibilities for the value of D when $|J| < |D|$. In principle, more sophisticated (and less sensitive) techniques[4] could be used to resolve some of these ambiguities. In the present case, the ambiguities were resolved according to the value predicted from the crystal structure and an alignment tensor that was determined independently from one-bond $^1H^N \leftrightarrow {}^{15}N$ RDCs, assuming an N−H distance of 1.02 Å.

The number and accuracy of the determined dipolar couplings is very high. For a total of 99 $^1H^N \leftrightarrow {}^{13}C^{aliphatic}$, 85 $^1H^N \leftrightarrow {}^{13}C'$, and 75 $^1H^N \leftrightarrow {}^1H^N$ RDCs the rms deviations between predicted and measured values are 0.8, 1.1, 1.1 Hz, respectively (Supporting Information). Including the 55 $^1H^N \leftrightarrow {}^{15}N$ RDCs, the total number of couplings per amide proton is 5.8 for residues T7−E61. The deviations are not equally distributed along the primary sequence (Figure 3A), but are strongest for the loop/turn regions around residues T16 and G46 located between strands β1 and β2 and between helix α1 and β3. In other parts of the sequence, the deviations are very small, for example, 0.7, 0.7, and 0.3 Hz for E24, A25, and T49, respectively (Figure 3B).

Systematic errors in D_{HC} arise from imperfections of RF pulses. These were estimated smaller than 3% on the basis of a determination of $^1J_{CH}$ on a sample of ^{13}C-acetate using the first part of the HCN-HSQC sequence. Similar estimates can be made for D_{HH}.[8] Additional systematic errors of less than 2% for D_{HC} (D_{HH}) arise from incomplete ^{13}C-labeling (amide protonation). On the basis of these estimates, the total systematic error should be smaller than 4%, whereas the statistical error was determined as smaller than 0.14 Hz from a repetition of the experiments. For the present alignment tensor ($D_{NH,max}$ = 25.4 Hz), such errors indicate that a single H−C distance of 2 Å can be determined to a precision that is higher than 0.03 Å.

Many of the deviations observed in Figure 3A are larger than these experimental errors and must be the result of an inaccurate description of the solution state by the crystal coordinates. A large part of these deviations result from short-range RDCs (e.g., Figure 3B, A53, and Supporting Information) where slight variations in structure lead to large changes in expected RDC values. A good correlation is observed between the deviations of the RDC values and the crystalline *B*-factor (Figure 3A). This indicates that disorder may be a main cause for the deviations in the loop regions and that the description by a single structure (with uniform order parameters) may not be justified.

In contrast to NOEs, which can also be obtained over large distances in perdeuterated proteins, RDCs do not suffer from noncoherent transfer processes such as spin diffusion or other nonlinear effects.[10,4] Therefore, their interpretation in terms of structure and dynamics is straightforward. Clearly, the present analysis depended on the availability of a high-resolution structure to determine the orientation tensor and to test the accuracy of the measured RDCs. In principle, the large number of observable long- and short-range RDCs should overdetermine the problem of an a priori structure calculation. This overdetermination can be further augmented by upper limits in cases where no long-range RDCs are detected and by the use of additional alignment media. Efforts are underway to explore this potential for the fast and accurate determination of small protein structures. Further possibilities exist to adapt the proposed experiments to somewhat larger proteins by TROSY techniques. Finally, for cases where initial structural information is available, the current example shows that the large number of accurate RDCs to single protons can reveal additional, unknown details of proton position and dynamics.

Acknowledgment. This work was supported by SNF Grant 31-61'757.00.

Supporting Information Available: Details of the HCN-HSQC experiment and Figures showing correlations between measured and predicted $H^N \leftrightarrow {}^{13}C$, $^1H^N \leftrightarrow {}^1H^N$, and $^1H^N \leftrightarrow {}^{15}N$ RDCs (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Tolman, J. R. *Curr. Opin. Struct. Biol.* **2001**, *11*, 532−539.
(2) LeMaster, D. M.; Richards, F. M. *Biochemistry* **1988**, *27*, 142−150.
(3) Torchia, D. A.; Sparks, S. W.; Bax, A. *J. Am. Chem. Soc.* **1988**, *110*, 2320−2321.
(4) Wu, Z.; Bax, A. *J. Am. Chem. Soc.* **2002**, *124*, 9672−9673.
(5) Yang, D.; Venters, R. A.; Mueller, G.; Choy, W. Y.; Kay, L. E. *J. Biomol. NMR* **1999**, *14*, 333−343.
(6) Zhu, G.; Bax, A. *J. Magn. Reson. Ser. A* **1993**, *104*, 353−357.
(7) Derrick, J. P.; Wigley, D. B. *J. Mol. Biol.* **1994**, *243*, 906−918.
(8) Kuboniwa, H.; Grzesiek, S.; Delaglio, F.; Bax, A. *J. Biomol. NMR* **1994**, *4*, 871−878.
(9) Hansen, M. R.; Mueller, L.; Pardi, A. *Nat. Struct. Biol.* **1998**, *5*, 1065−1074.
(10) Tian, F.; Fowler, C. A.; Zartler, E. R.; Jenney, F. A., Jr.; Adams, M. W.; Prestegard, J. H. *J. Biomol. NMR* **2000**, *18*, 23−31.

JA028740Q

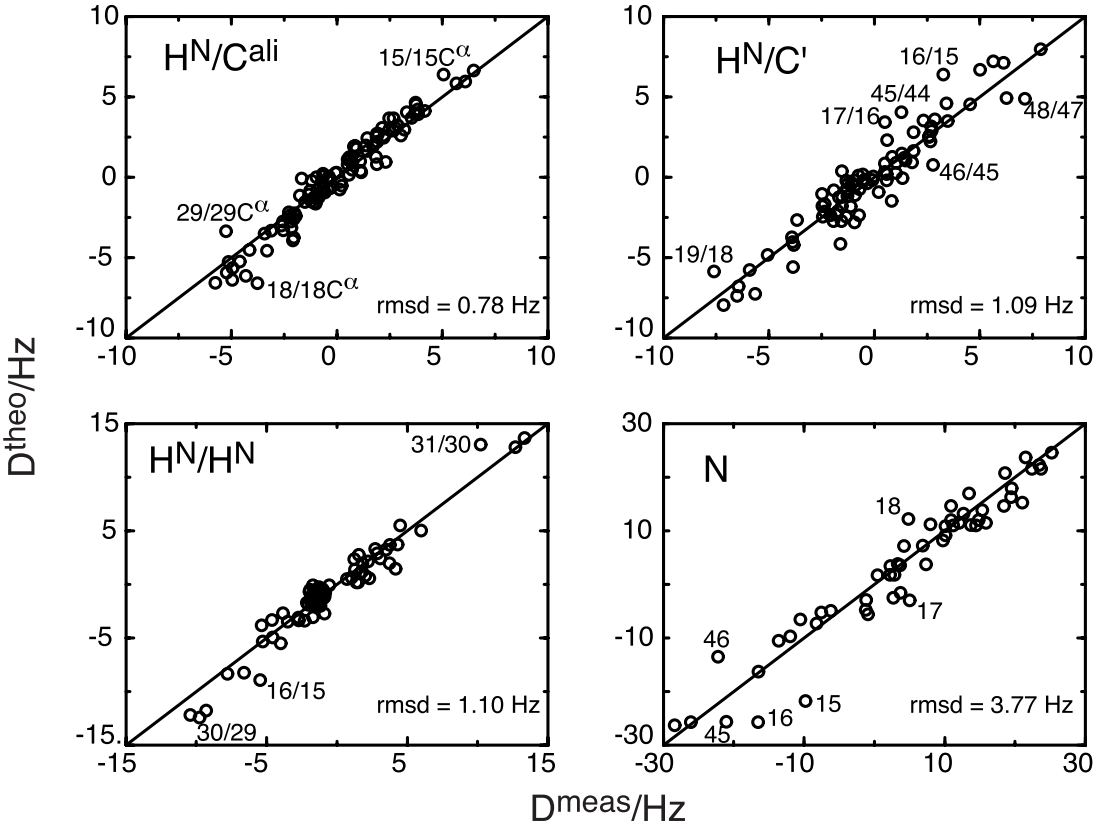
Experimental details of quantitative HCN-HSQC

Rectangular high (low) power ^{1}H pulses (carrier at the $^{1}H_2O$ resonance) are applied using $\gamma_H B_1$ = 28 (0.2) kHz. For detection of RDCs between $^{1}H^{N}$ and aliphatic (carbonyl) ^{13}C nuclei, the carrier frequency for solid shape ^{13}C-pulses is set to 43 (177) ppm. ^{13}C 90°-pulses have an RF strength of 27 kHz, whereas the shaped ^{13}C 180° G3-pulses (Emsley, L.; Bodenhausen, G. Chem. Phys. Lett. 1990, 165, 469-476) have a duration of 290 μs (aliphatic ^{13}C detection) or 745 μs (carbonyl ^{13}C detection) such that they effectively invert the ^{13}C chemical shift range from 18 to 68 or from 165 to 185 ppm, respectively. For the $^{1}H^{N}$ to carbonyl ^{13}C RDC-experiment, an additional 180° decoupling pulse (G3-profile) is applied on the $^{13}C^{\alpha}$ resonances in the center of the t_1-evolution period. Open rectangles correspond to two ^{13}C 180°-pulses ($\gamma_c B_3$ = 6.28 kHz) applied in consecution at 177 and 56 ppm. Delay durations: τ = 2.25 ms; τ' = 2.7 ms; T = 10 ms. Phase cycling: ϕ_1 = y; ϕ_2 = x,-x; ϕ_3 = y; ϕ_4 = 2x,2(-x); receiver = x,2(-x),x. Quadrature detection in the t_1- and t_2-dimensions is obtained by altering ϕ_2 and ϕ_4 in the States-TPPI manner, respectively. Gradients (sine bell shaped; 25 G/cm at center): $G_{1,2,3,4,5,6}$ = 1 (x), 1.14 (z), 1 (y), 3 (z), 2 (y), and 0.4 (x) ms (direction). For the reference experiment, no t_1-incrementation is carried out and the phase cycling is: ϕ_1 = x,-x; ϕ_2 = 8x,8(-x); ϕ_3 = 4x,4(-x); ϕ_4 = 2x,2(-x); receiver = x,2(-x), x, -x,2x, -x.

Details of the data analysis for quantitative HCN-HSQC

The coupling constant K_{HC} was extracted from the ratio of cross and reference peak intensities of the HCN-HSQC in the following way: peaks of the cross experiment were fitted by the time domain fitting routine nlinLS contained in the NMRPipe package (Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. J. Biomol. NMR 1995, 6, 277-293) such that the amplitudes (A_c) of the t_1-FIDs were obtained. The amplitude of the reference signal (A_R) was obtained from an analogous fit. The amplitudes were corrected for the different numbers of acquired scans in both experiments. $|K_{HC}|$ was then calculated from the ratio A_C/A_R of both amplitudes as $|K_{HC}| = \mathrm{atan}(|A_C/A_R|^{1/2})/(2\pi T)$.

Correlations between all measured and predicted $H^N \leftrightarrow {}^{13}C$, ${}^1H^N \leftrightarrow {}^1H^N$, and ${}^1H^N \leftrightarrow {}^{15}N$ RDCs



Supporting Information Figure 1. Correlations between the measured and calculated values for ${}^1H^N \leftrightarrow {}^{13}C^{aliphatic}$, ${}^1H^N \leftrightarrow {}^{13}C'$, ${}^1H^N \leftrightarrow {}^1H^N$, and ${}^1H^N \leftrightarrow {}^{15}N$ RDCs. One-bond ${}^1H^N \leftrightarrow {}^{15}N$ RDCs were obtained from doublet separated HSQC-experiments. The calculated values were obtained from the coordinates of the 1.1 Å crystal structure (1IGD) and an alignment tensor derived from all measured couplings in the regular secondary structure elements of protein G using an H-N distance of 1.02 Å. Outliers are labeled by their 1IGD sequence number and group into three distinct regions: a β-hairpin (residue 15-19), the beginning of the helix (29-31), and a loop (44-48).

The observed long-range correlations span distances of maximally 3.2 Å and 7.2 Å for ${}^1H^N \leftrightarrow {}^{13}C$ and ${}^1H^N \leftrightarrow {}^1H^N$, respectively. Note that due to the angular dependency of the dipolar couplings [$D = A_{zz}((3\cos^2\theta - 1)/2 + \eta/2 \sin^2\theta \cos 2\phi)/r^3$], not all nuclear pairs that fall within these limits lead to detectable cross-peaks in the dipolar experiments.

Chapter 3
Local structure in a folding nucleus from T4 fibritin

The trimeric fibritin protein is a rodlike structural protein of 52 kDa per monomer and is attached to the "neck" of the bacteriophage T4. The C-terminal domain of the fibritin protein consists of 27 amino acids per monomer and is necessary for the correct folding and trimerization of the 486 amino acid holoprotein *in vivo* and *in vitro* (Letarov et al., 1999). Due to its role as a folding nucleus, the domain has been termed foldon. Foldon is an effective trimerization tool in protein engineering and has been used to induce and enhance trimerization of collagen and HIV glycoproteins (Frank et al., 2001; Yang et al., 2002). In the following, the sequence of the foldon domain is renumbered to 1-27, corresponding to residues 457-483 in fibritin.

In chapter 3.1 we report the solution structure of the foldon domain. Optical spectroscopy has been used by collaborators in the group of Thomas Kiefhaber (Biozentrum, Basel) to determine the overall stability and the folding and association kinetics of the trimer (Güthe et al., 2004). The detailed folding kinetics show that the foldon domain is formed via a collapsed monomeric intermediate, which concomitantly trimerizes to the native structure.
In chapter 3.2 we quantify the strength of electrostatic inter-subunit interactions in the foldon trimer by NMR spectroscopy. We show that the monomeric folding intermediate detected by optical methods can be stabilized via the weakening of an inter-subunit ion bond at low pH. The partially folded monomer is thus amenable to a detailed NMR analysis. The kinetics and thermodynamics of secondary structure formation in the single β-hairpin of the acidic state are reported (Meier et al., 2004a).

3.1 Solution structure and stability of the foldon trimer

The solution structure of the homotrimeric foldon domain in isolation was determined to a backbone rmsd of 0.27 Å (Figure 3–1 A) using weak alignment techniques for the measurement of $^1D_{NH}$, $^1D_{NC'}$, $^1D_{C\alpha C\beta}$ and $^1D_{CH}$ RDCs in U–^{13}C, U–^{15}N labeled protein. Ambiguous NOEs were iteratively identified as intra- or intermolecular contacts from structures calculated with unambiguous NOEs and with RDCs. The central structural element of each subunit in the foldon trimer is a short twisted β-hairpin (Figure 3–1), in which β-strands are connected by a type I' tight β–turn (in the notation of (Richardson, 1981)). Hydrophobic sidechains of the hairpin residues are solvent-shielded by an N-terminal polyproline II helix, by a C-terminal 3_{10} helix as well as by hydrophobic contacts along the trimer interface.

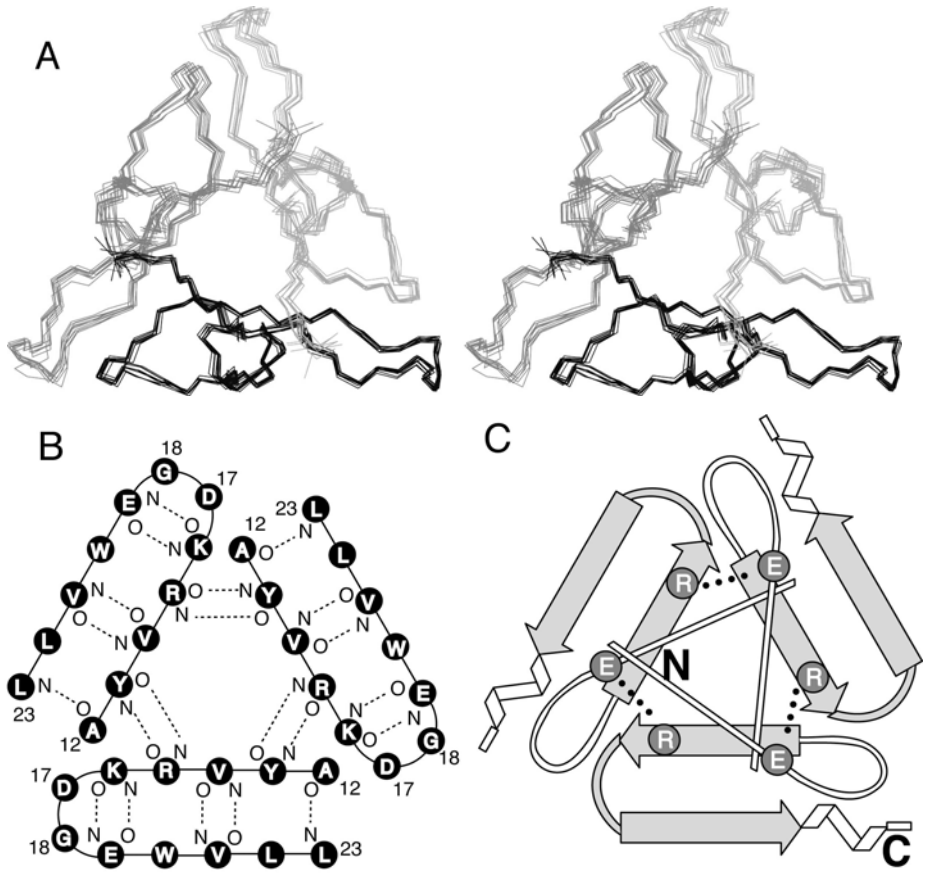


Figure 3–1. Structure and topology of the foldon trimer. (A) Bundle of the ten lowest energy structures out of 100 calculated with CNS (Brunger et al., 1998), shown as a stereo representation of the backbone. (B,C) Schematic representations of the topology showing inter-subunit hydrogen bonds R15-Y13 (B) and inter-subunit salt bridges E5-R15 (C). Strands in the β-hairpin are short and are connected by a type I' β–turn. The N- and the C-terminus are indicated for the lower monomer in (C).

The foldon domain adopts a trimeric β-propeller structure in solution, like in crystal structures of larger C-terminal fibritin constructs (Tao et al., 1997; Strelkov et al., 1998). Model-free spectral density fits of relaxation data (Lipari and Szabo, 1982; Dosset et al., 2000) reveal that the whole foldon domain is highly rigid at room temperature. Thermal unfolding monitored with NMR spectra shows that the trimer is highly stable and unfolds with high cooperativity at a transition temperature around 75 °C at pH 7 and 300 μM total peptide concentration.

The structural characterization of the foldon trimer in solution and a kinetic analysis of folding and association have been published in (Güthe et al., 2004). A bundle of the 10 lowest energy structures out of 100 calculated has been deposited in the PDB under accession number 1RFO.

Original Publication

Güthe, S., Kapinos, L., Möglich, A., Meier, S., Grzesiek, S., and Kiefhaber, T.

Very fast folding and association of a trimerization domain from bacteriophage T4 fibritin

Journal of Molecular Biology **337**, 905-15 (2004).

doi:10.1016/j.jmb.2004.02.020 J. Mol. Biol. (2004) 337, 905–915

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Very Fast Folding and Association of a Trimerization Domain from Bacteriophage T4 Fibritin

Sarah Güthe[1]†, Larisa Kapinos[1]†, Andreas Möglich[1]† Sebastian Meier[2], Stephan Grzesiek[2] and Thomas Kiefhaber[1]*

[1]*Division of Biophysical Chemistry, Biozentrum der Universität Basel Klingelbergstrasse 70, CH-4056 Basel, Switzerland*

[2]*Division of Structural Biology Biozentrum der Universität Basel, Klingelbergstrasse 70 CH-4056 Basel, Switzerland*

**Corresponding author*

The foldon domain constitutes the C-terminal 30 amino acid residues of the trimeric protein fibritin from bacteriophage T4. Its function is to promote folding and trimerization of fibritin. We investigated structure, stability and folding mechanism of the isolated foldon domain. The domain folds into the same trimeric β-propeller structure as in fibritin and undergoes a two-state unfolding transition from folded trimer to unfolded monomers. The folding kinetics involve several consecutive reactions. Structure formation in the region of the single β-hairpin of each monomer occurs on the submillisecond timescale. This reaction is followed by two consecutive association steps with rate constants of $1.9(\pm 0.5) \times 10^6\ M^{-1}\ s^{-1}$ and $5.4(\pm 0.3) \times 10^6\ M^{-1}\ s^{-1}$ at 0.58 M GdmCl, respectively. This is similar to the fastest reported bimolecular association reactions for folding of dimeric proteins. At low concentrations of protein, folding shows apparent third-order kinetics. At high concentrations of protein, the reaction becomes almost independent of protein concentrations with a half-time of about 3 ms, indicating that a first-order folding step from a partially folded trimer to the native protein ($k = 210(\pm 20)\ s^{-1}$) becomes rate-limiting. Our results suggest that all steps on the folding/trimerization pathway of the foldon domain are evolutionarily optimized for rapid and specific initiation of trimer formation during fibritin assembly. The results further show that β-hairpins allow efficient and rapid protein–protein interactions during folding.

Keywords: protein folding; protein association; trimeric proteins; prolyl isomerization; fast folding

Introduction

Fibritin is a rod-like structural protein of bacteriophage T4, which is attached to the neck of the virion *via* its N-terminal domain to form the collar structures ("whiskers"). Fibritin consists of an N-terminal anchor domain (residues 1–46), a large central coiled-coil part (residues 47–456) and a small C-terminal globular domain (residues 457–486).[1] The 30 amino acid residue C-terminal domain was termed foldon, since it was shown to be essential for fibritin trimerization and folding *in vivo* and *in vitro*.[1–3] Each subunit of the foldon domain consists of a single β-hairpin, which assemble into a β-propeller-like structure in the trimer.[1] The trimer is stabilized by hydrophobic interactions involving Trp476 of each subunit, intermolecular salt-bridges between Glu461 and Arg471, and intermolecular backbone hydrogen bonds between Tyr469 and Arg471 (Figure 1). Expression of the isolated foldon domain (residues 457–483) yields a stable trimer, which shows a cooperative two-state thermal unfolding transition.[4] Residues 484–486 were omitted from this study, since this region is unordered in the X-ray structure of fibritin.[1]

The foldon domain was proposed to be an evolutionarily optimized trimerization/folding

† S.Gü., L.K. and A.M. contributed equally to this work.

Abbreviations used: F-moc, *N*-(9-fluorenyl)-methoxycarbonyl; GdmCl, guanidinium chloride; RDC, residual dipolar coupling.

E-mail address of the corresponding author: t.kiefhaber@unibas.ch

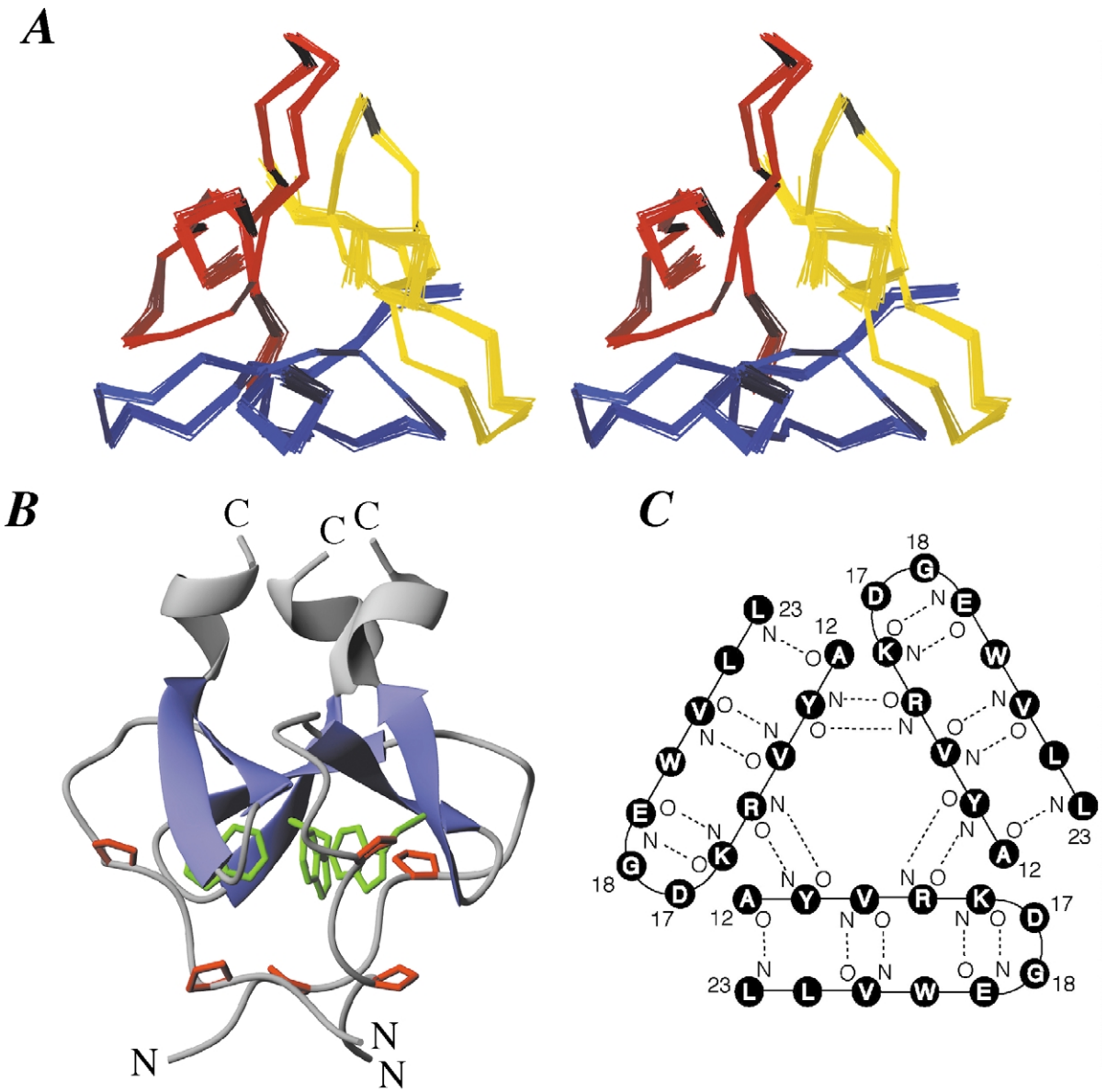


Figure 1. A, Stereo view of a bundle of the 20 lowest-energy structures of the trimeric foldon domain determined by NMR spectroscopy. Each subunit is displayed in a different color. B, Side view of the foldon structure with the single Trp residues at position 20 of each chain highlighted in green and the two prolyl residues at positions 4 and 7 highlighted in red. C, Topology of the interactions of the three β-hairpins in the native foldon domain. The figures in A and B were prepared using the program MOLMOL[37] and rendered with PovRay.

motif, as its only known function is to promote folding of fibritin.[3] The small size of its structured part (27 amino acid residues) and its simple fold make the foldon domain a perfect system for a detailed study on the mechanism of a folding reaction linked to intermolecular association steps. All previous folding studies on trimeric proteins investigated large filamentous proteins, which show extremely slow and complex folding kinetics, usually accompanied by irreversible aggregation reactions.[5]

We expressed the foldon domain in *Escherichia coli* and synthesized it by solid-phase *N*-(9-fluorenyl)methoxycarbonyl (F-moc) chemistry to investigate its structure, stability and folding mechanism. For clarity, we are numbering the foldon sequence from residues 1 to 27

```
1                  11                  21
GYIPEAPRDG QAYVRKDGEW VLLSTFL
```

corresponding to residues 457–483 in fibritin. All kinetic and stability data presented here were obtained using the chemically synthesized foldon domain, whereas the recombinant *E. coli* product was used for structural analysis. The *E. coli* product and the synthetic foldon domain showed identical stability and folding behavior. Further, the additional C-terminal amino acid residues Ser-Pro-Ala, which are present in the wild-type fibritin sequence, do not affect any thermodynamic or kinetic properties of the foldon domain.

Results and Discussion

Structure and stability of the foldon domain

To test whether the 27 amino acid residue foldon domain adopts the same fold as in fibritin, we solved its solution structure to a backbone rmsd of 0.31 Å with 28 experimental restraints per residue (Figure 1 and Table 1). The structure largely resembles the crystal structure in constructs carrying the 75 and 120 C-terminal amino acid residues of fibritin, in which the foldon domain constitutes only a minor part of the total construct.[1,6] Only at the immediate N terminus (residues 1–3) the isolated foldon domain assumes a slightly different and presumably more relaxed structure compared to fibritin. The trimer consists of an N-terminal hydrophobic stretch in left-handed polyproline II helix conformation between Pro4 and Pro7, which is connected to a β-hairpin (residues 12–23) and forms a hydrophobic cap of the hairpin on the N-terminal side (Figure 1). The hairpin terminates in a 3_{10} helix at the C terminus with homophilic interactions of hydrophobic residues (Tyr2, Ile3, Val14, Leu23, Leu27) between the monomers along the symmetry axis. Large-scale nanosecond dynamics as evidenced by ^{15}N relaxation occur only at the most N-terminal residue, Tyr2. All other residues in the highly rigid foldon domain exhibit order parameters $S^2 > 0.77$ at 25 °C as determined by the program TENSOR.[7]

Table 1. Statistics of the foldon NMR structure

rmsd from experimental distance constraints (Å)	
All (607)[a]	0.046 ± 0.002
rmsd from NMR data	
NMR quality factor Q[b]	0.199 ± 0.0062
rmsd (Hz) between measured and calculated dipolar couplings (81)[c]	1.91 ± 0.07
Experimental dihedral constraints (deg.)[d] (43)	1.73 ± 0.29
$^3J_{HNHA}$ coupling constants (Hz) (22)	0.89 ± 0.05
Total number of restraints per monomer	753
Deviation from the idealized covalent geometry	
Bonds (Å)	0.0079 ± 0.0004
Angles (deg.)	0.92 ± 0.03
Impropers[e] (deg.)	0.75 ± 0.06
Coordinate precision[f] (Å)	
Backbone non-hydrogen atoms	0.273
All non-hydrogen atoms	0.636
Non-Gly, non-Pro residues in Ramachandran regions[g]	
Most favored (%)	91.7
Allowed (%)	8.3
Generously allowed (%)	0.0
Disallowed (%)	0.0

The statistics were obtained from a subset of the 40 best energy structures out of 100 following a standard simulated annealing protocol with dipolar restraints incorporated. Individual simulated annealing structures are fitted to each other using residues 2–27 of all subunits. The number of the various constraints per monomer is given in parentheses.

[a] Distance restraints comprise: 111 intraresidual NOEs; 139 sequential NOEs ($|i-j|=1$); 73 short range NOEs ($1<|i-j|\leq 5$); 128 long-range NOEs ($|i-j|\leq 5$); 156 intermolecular NOEs; 11 H-bonds (eight intramolecular, three intermolecular). For each backbone hydrogen bond constraint, there are two distance restraints: r_{NH-O}, 1.7–2.5 Å, r_{N-O}, 2.3–3.5 Å.

[b] The NMR quality factor Q is defined as the ratio of the rmsd between observed and calculated couplings and the rmsd of the observed couplings.[35]

[c] The 81 RDCs comprise 22 $^1D_{HN}$, 20 $^1D_{H\alpha C\alpha}$, 12 $^1D_{C\alpha C\beta}$, 13 $^1D_{NC'}$ (0.231), 14 $^1D_{CH3}$. Ramping the force constant for RDCs in the structure calculation from 0.001 kcal mol^{-1} Hz^{-2} to 0.5 kcal mol^{-1} Hz^{-2} was determined as optimal.

[d] The dihedral angle constraints comprise 69 ϕ and 60 ψ angles.

[e] The improper torsion restraints serve to maintain planarity and chirality.

[f] The coordinate precision is defined as the average rms difference between the individual simulated annealing structures and the mean coordinates. Values are reported for residues 2–27.

[g] These values are calculated with the program PROCHECK-NMR.[36] Values are reported for all residues.

The equilibrium unfolding properties of the foldon domain were measured by guanidinium chloride (GdmCl)-induced unfolding transitions at various concentrations of protein. Figure 2 shows unfolding curves at monomer concentrations of 5 μM and 30 μM. The coincidence of fluorescence and CD-monitored transition curves demonstrates that the trimer unfolds in a cooperative two-state transition at both concentrations of protein. Two-state unfolding is observed for all concentrations of protein between 2 μM and 100 μM. The sensitivity of the unfolding transitions to changes in protein concentration is expected for unfolding of a native trimer (N) to unfolded monomers (U):

$$N \rightleftharpoons 3U \qquad (1)$$

The transitions at 5 μM and 30 μM can be fit globally to equation (1) (continuous line in Figure 2) by using:

$$K_{eq} = \frac{3f_U^3[M]_0^2}{1-f_U} \qquad (2)$$

where $[M]_0$ indicates the total monomer concentration ($[M]_0 = [U] + 3[N]$), f_u is the fraction of

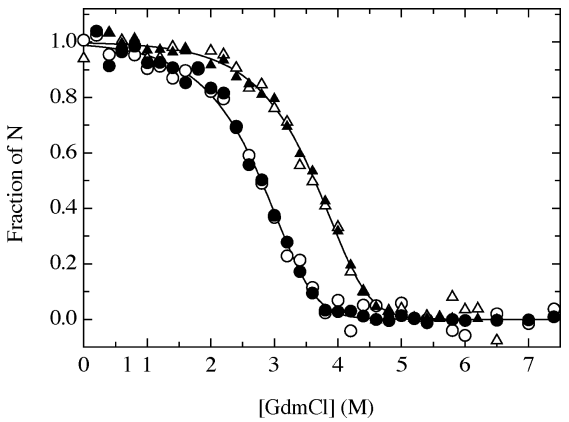


Figure 2. GdmCl-induced unfolding transition of the foldon domain at pH 7.1, 20 °C. Transitions at 5 μM (○,●) and 30 μM (▲,△) total monomer concentration ($[M]_0$) were measured by changes in Trp fluorescence (●,▲) and in far-UV CD at 228 nm (○,△). The data were normalized to fraction of native molecules using the result of a global fit of all data according to equations (14a) and (14b) (continuous lines).

unfolded monomer ($f_U = [U]/[M_0]$) and K_{eq} is the equilibrium constant (for details, see Materials and Methods). The global fit yields a free energy of unfolding of $\Delta G^0(H_2O) = 89.2(\pm 0.6)$ kJ/mol, which is unusually high compared to stabilities of small, single-domain proteins of similar size. However, this value applies to standard conditions of 1 M total monomer concentration. At typical physiological protein concentrations around 5 μM this corresponds to $\Delta G = 29.7$ kJ, which is similar to the stabilities found for small monomeric proteins. The change in free energy with GdmCl ($m_{eq} = \partial\Delta G^0/\partial[\text{GdmCl}]$) is $-10.4(\pm 0.2)$ (kJ/mol)/M, which is the value expected for a monomeric globular protein of the size of the folded trimer.[8] This shows that native foldon has properties comparable to those of small monomeric proteins with a compact hydrophobic core and a cooperative two-state unfolding transition.

Burst phase fluorescence changes

To investigate the folding kinetics of the foldon domain we performed stopped-flow refolding experiments starting from GdmCl-unfolded protein. Figure 3A shows a refolding trace at a residual denaturant concentration of 0.58 M and $[M]_0$ of 5 μM. The kinetics were monitored by the change in intrinsic tryptophan fluorescence above 320 nm. Within the first millisecond of refolding, a major burst phase reaction occurs, which leads to a significant increase in fluorescence intensity above the signals of both the unfolded and the native protein. This indicates very rapid structural changes in the dead-time of stopped-flow mixing (about 1 ms). The fluorescence intensity decreases slowly and reaches the value of the native protein after about 300 seconds. The burst phase increase in fluorescence is observed for all measured concentrations of protein (0.5 μM to 200 μM), indicating that the reaction occurs within the monomer. Even the fastest, diffusion-controlled association reaction to a partially folded dimer could not be complete within 1 ms at a monomer concentration of 1 μM and below, if we assume a maximum second-order rate constant[9] of about $1 \times 10^9\ M^{-1}\ s^{-1}$. To further investigate the structural changes occurring in the burst phase, we monitored the folding kinetics at single wavelengths between 290 nm and 430 nm. Extrapolating the kinetic traces at the individual wavelengths to time zero allows the determination of the fluorescence spectrum of the burst phase intermediate (Figure 3B). Comparison of the fluorescence spectra of native and unfolded protein with the zero timepoint spectrum shows that the burst phase intermediate has a fluorescence emission maximum around 330 nm, which is between the emission maximum of the native protein ($\lambda_{max} = 317$ nm) and the unfolded state ($\lambda_{max} = 345$ nm). The significantly blue-shifted fluorescence maximum and the largely increased fluorescence intensity in the intermediate relative

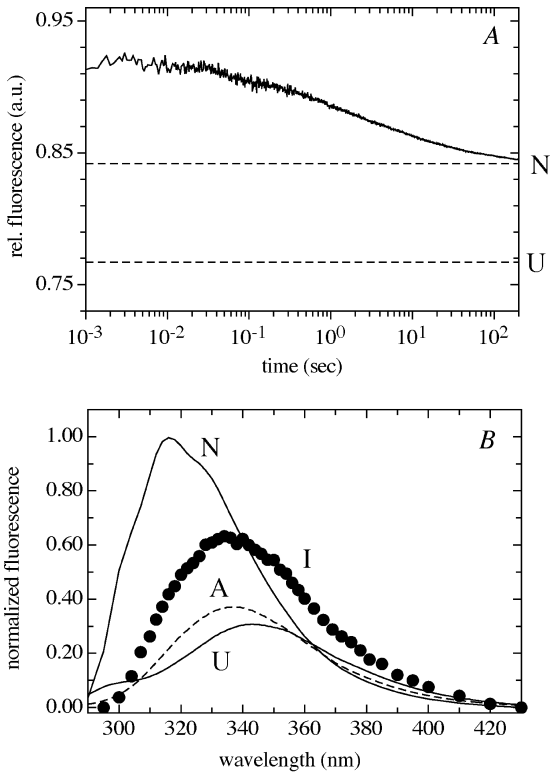


Figure 3. A, Refolding of the foldon domain in 0.58 M GdmCl, pH 7.1 ($[M]_0$ 5 μM) measured by the change in Trp fluorescence using a 320 nm emission cut-off filter. The broken lines represent the signals of the native and of the unfolded state at 0.58 M GdmCl, as indicated. The signal of the unfolded state is extrapolated from the unfolded baseline at high concentrations of GdmCl to 0.58 M GdmCl (see Materials and Methods). B, Comparison of the fluorescence spectrum of the kinetic burst phase intermediate (I) with the spectra of native (N) protein in 0.58 M GdmCl, the unfolded protein (U) in 8.2 M GdmCl and the monomeric A-state formed at pH 2 (A). The spectrum of I was determined in single-wavelength detection stopped-flow experiments. The fluorescence intensity extrapolated to $t = 0$ is shown. $[M]_0$ was 5 μM for all spectra.

to the unfolded state suggest that the burst phase intermediate has a significantly more hydrophobic environment around the single tryptophan residue at position 20 in each β-hairpin (see Figure 1). The absence of a tyrosine fluorescence band at 303 nm in the burst phase intermediate further indicates significant chain compaction, which allows efficient energy transfer from the two tyrosine residues at positions 2 and 13 to Trp20. Similar fluorescence properties are observed for an acid-induced monomeric state (A-state) of the foldon domain, which shows virtually the same fluorescence emission spectrum as the burst phase intermediate but with reduced fluorescence intensity (Figure 3).

Fast and slow steps during association of the foldon domain

To determine the nature of the rate-limiting steps

during folding and association of the foldon domain, we analyzed the concentration-dependence of the refolding kinetics at final concentrations of protein between 0.5 μM and 200 μM. Figure 4A shows that the kinetics are strongly concentration-dependent. A very slow reaction on the hundreds of seconds timescale is observed at low concentrations of protein. At higher concentrations of protein ($[M]_0 > 10$ μM) a concentration-dependent faster reaction with a half-time of about 4 ms at $[M]_0 = 102$ μM and a concentration-independent slow reaction ($\tau = 50$ s) are observed. The foldon domain contains two prolyl residues (Pro4 and Pro7) per subunit, which are in the *trans* conformation in the native state. This makes prolyl isomerization reactions a possible source for the slow, concentration-independent kinetics. Figure 5 shows that the slow reaction is catalyzed efficiently by human cyclophilin 18, a peptidyl-prolyl *cis–trans* isomerase,[10] which identifies this reaction as a *cis–trans* isomerization at one or both of the two Xaa-Pro peptide bonds per monomer. The faster reaction is not affected by the presence of cyclophilin (data not shown).

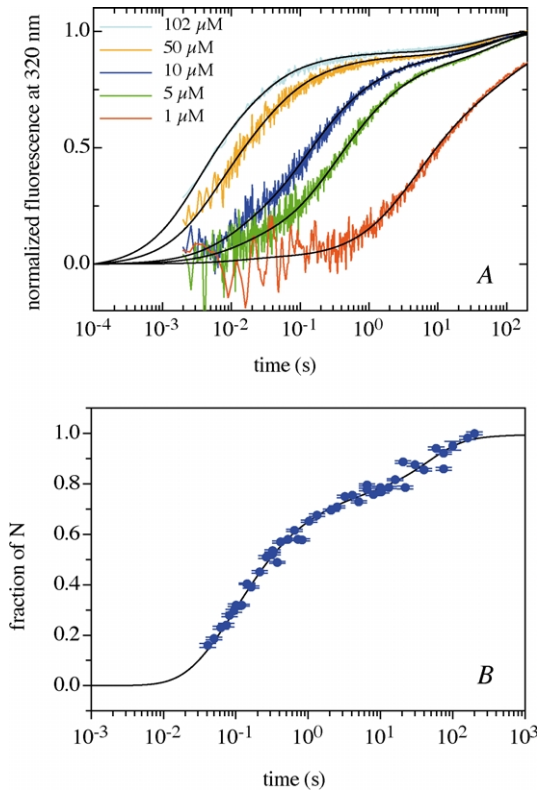


Figure 4. A, Fluorescence detected folding kinetics in the presence of 0.58 M GdmCl (pH 7.1), 20 °C at the indicated values of $[M]_0$ measured after stopped-flow mixing at an emission wavelength of 320 nm. B, Time-course of formation of native molecules measured in interrupted refolding experiments at pH 7.1, 20 °C, $[M]_0 = 10$ μM. The continuous lines in both panels represent the results from a global fit of the data. For global fitting, ten fluorescence-detected refolding traces at $[M]_0$ between 0.5 μM and 102 μM and the time-course of formation of native molecules shown in B, were fitted simultaneously to the kinetic model shown in equation (4).

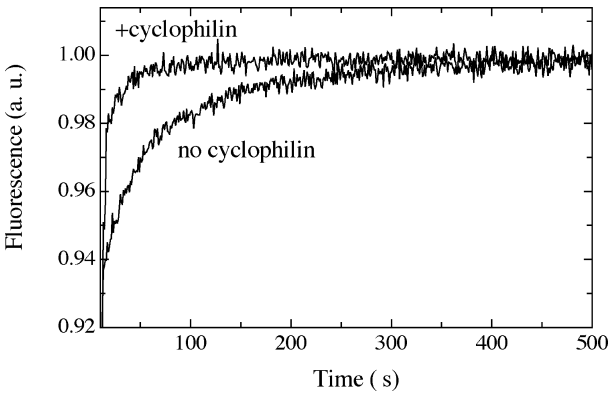


Figure 5. Effect of human cyclophilin 18 on the slow refolding reaction of the foldon domain. The plot compares refolding at pH 7.1, 0.58 M GdmCl, 20 °C in the absence and in the presence of 3.3 μM cyclophilin. The presence of cyclophilin increases the rate constant for the slow reaction from 7×10^{-3} s^{-1} to 4×10^{-2} s^{-1}.

We tested whether the faster, concentration-dependent reaction produces native protein or a folding intermediate by performing interrupted refolding experiments.[11–13] In these experiments, the protein is allowed to refold for a certain time (t_i). The protein is then transferred to unfolding conditions and the resulting unfolding kinetics are monitored. The native state has a characteristic stability and barrier for unfolding, which results in a characteristic rate constant for its unfolding reaction. This distinguishes it from partially folded intermediates. Interrupted refolding experiments measure the increase in amplitude of the unfolding reaction of the native protein as a function of the refolding time, t_i. This corresponds to the time-course of formation of the native state. Thus, interrupted refolding experiments can distinguish whether a folding reaction produces native protein or a folding intermediate.[12,13] Figure 4B shows that native molecules are formed in a fast and a slow reaction, which occur on the same timescale as the two fluorescence-detected folding reactions (Figure 4B). Obviously, both the faster, concentration-dependent process and the prolyl-isomerization limited process produce native protein. This suggests that the fast reaction reflects formation of the native trimer for molecules with all prolyl peptide bonds in the native *trans* conformation (fast-folding molecules, U_F). The slow process is due to folding of molecules with at least one non-native *cis* isomer (slow-folding molecules, U_S), as shown by its catalysis by cyclophilin (Figure 5).

Mechanism of folding and association

The assignment of the different kinetic phases to direct folding and prolyl isomerization steps enables us to characterize the folding/association

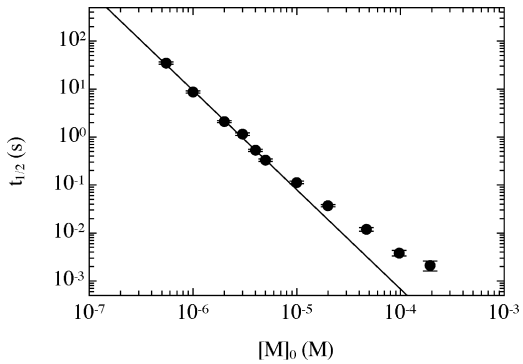


Figure 6. Effect of total monomer concentration ($[M]_0$) on the half-time of the fast-refolding reaction of the foldon domain at 0.58 M GdmCl (pH 7.1), 20 °C. The continuous line represents a fit of the data between 0.5 μM and 4 μM to equation (3). The fit gives a slope of -2.06 ± 0.05 indicating an apparent reaction order of 3 at low concentrations of protein.

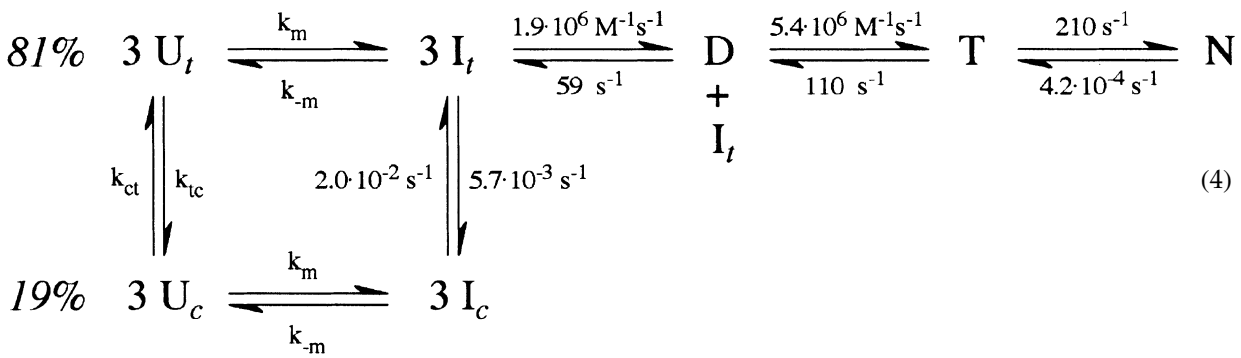
process of the foldon domain in more detail. The analysis of the concentration-dependence of the half-lives ($t_{1/2}$) of the folding reaction starting from U_F allows us to determine the apparent reaction order for the fast-folding pathway according to:

$$\log t_{1/2} = c - (n-1)\log [M]_0 \qquad (3)$$

where n is the reaction order and c is a reaction-specific constant.[14] Analysis of the concentration-dependence of folding thus yields information on contributions from concentration-dependent association reactions and from concentration-independent folding steps. We determined the half-time of the folding reaction from the concentration-dependence of the fluorescence-detected kinetics shown in Figure 4A. At high concentrations of protein, where folding and prolyl isomerization are well separated, we evaluated the half-lives of the fast reactions in order to obtain information on the direct folding reaction. Figure 6 shows that the slope of $\log t_{1/2}$ *versus* $\log[M]_0$ changes with protein concentration. At low concentrations of protein the slope is -2.06 ± 0.05, which corresponds to apparent third-order kinetics. At monomer concentrations above 5 μM the slope decreases significantly, indicating a change in the apparent reaction order. Between 100 μM and 200 μM initial monomer concentration there is only little effect of protein concentration on the folding kinetics. This shows that the fast-folding reaction approaches the first-order limit at high concentrations of protein.

The change from apparent third-order kinetics to first-order kinetics shows that folding is limited by association steps at low concentrations of protein and by a unimolecular folding reaction at high concentrations of protein. It is very unlikely that apparent third-order kinetics for a reaction in solution arise from a true trimolecular reaction.[15] Third-order reactions are usually caused by a rapid monomer–dimer pre-equilibrium followed by a second bimolecular association step to form the trimer. This mechanism gives rise to apparent third-order kinetics if the dimer is populated to only very low levels. The observation of apparent third-order kinetics at low concentrations of protein confirms the finding that the rapid collapse observed within the first millisecond of refolding is due to a conformational change in the monomer.

The weak concentration-dependence of the fluorescence-detected kinetics at high concentrations of protein implies that a unimolecular folding step becomes rate-limiting when all association reactions are fast. The half-time for formation of the native state of about 3–4 ms at high concentrations of protein indicates that the rate constant of the unimolecular step is around 200–300 s^{-1}. It is reasonable to assume that this unimolecular folding process is due to a structural rearrangement in the trimer and represents a late step in the formation of the native structure, and might be similar to the final steps during folding of a single-domain protein. However, we cannot exclude completely the possibility that the first-order reaction, which becomes rate-limiting at high concentrations of protein, occurs at the level of the burst phase intermediate or a partially folded dimer. These considerations lead to the minimal folding model for the foldon domain:



(4)

In this mechanism, U denotes the completely unfolded monomer, I is the monomeric burst phase intermediate, D is a partially folded dimer, T is a partially folded trimer and N the native trimer. The subscripts *c* and *t* indicate monomers with a least

one *cis* Xaa-Pro peptide bond and monomers with all *trans* peptide bonds, respectively. k_m and k_{-m} represent rate constants for formation and unfolding of the burst phase intermediate, respectively, which occurs on the submillisecond timescale. To test whether this folding mechanism can describe all experimental data quantitatively, we fitted the folding kinetics at different concentrations of protein (Figure 4A) together with the time-course of formation of native foldon at 10 μM (Figure 4B) globally to the mechanism shown in equation (4). The results show that both the concentration-dependence of fluorescence kinetics and the time-course of formation of native molecules are well described by the model (continuous lines in Figure 4A and B). The global fit allows the determination of the rate constants for all reactions that occur after the submillisecond burst phase, since the apparent reaction order changes from third-order to almost first-order. The results of the global fit are given in equation (4). In agreement with the lower limit of the estimate from the limiting value of the half-lives for folding at high concentrations of protein (Figure 6) the fit gives a rate constant (k_f) of structural rearrangement in the trimer of $210(\pm 20)\ s^{-1}$. This is in the same range as the rate constants observed for the fastest folding monomeric proteins of similar size.[16] The bimolecular rate constants for formation of the dimer and of the trimer are $1.9(\pm 0.5) \times 10^6\ M^{-1}\ s^{-1}$ and $5.4(\pm 0.3) \times 10^6\ M^{-1}\ s^{-1}$, respectively, which is about 100 to 200-fold slower than the expected diffusion limit for bimolecular reactions of chains of this size.[9] However, both association reactions on the foldon trimerization pathway are significantly faster than most bimolecular steps during folding of dimeric proteins like the well-characterized GCN4[17] leucine zipper and of many globular dimeric proteins.[18] Rate constants similar to the two bimolecular steps during trimerization of foldon have been reported for wild-type arc repressor,[9] the fastest folding naturally occurring dimeric protein known, and for some designed leucine zippers.[19] However, an engineered arc repressor variant[20] and a designed fragment of trp repressor[21] show bimolecular rate constants of about $3 \times 10^8\ M^{-1}\ s^{-1}$, which are the fastest association reactions reported for folding of small dimeric proteins to date. In the case of the engineered arc repressor, the rate-enhancement relative to the wild-type protein was achieved by replacing the intermolecular salt-bridge/hydrogen bonding network by hydrophobic residues.[20] The fast-folding trp repressor fragment is stabilized mainly by intermolecular hydrophobic interactions. Similar to wild-type arc repressor, the foldon domain is stabilized by intermolecular hydrogen bonds and by an intermolecular salt-bridge, which contributes substantially (17 kJ/mol) to trimer stability (S.M. *et al.*, unpublished results). In this respect, it is interesting to note that the association steps in the foldon domain have virtually the same rate constants as the dimerization step of wild-type arc repressor.

The unfolding rate constants of foldon under native conditions (0.58 M GdmCl) obtained from the fit reveal that the major barrier for unfolding is represented by the reaction from the native trimer (N) to the partially folded trimer ($k = 4.2(\pm 0.5) \times 10^{-4}\ s^{-1}$). This value corresponds well to the rate constant for unfolding of native foldon measured at high concentrations of denaturant and extrapolated to 0.58 M GdmCl (S.Gü. & T.K., unpublished results). Unfolding of the partially folded trimer (T) and the dimer (D) are significantly faster with rate constants of $110(\pm 40)\ s^{-1}$ and $59(\pm 4)\ s^{-1}$, respectively, obtained from the fit.

In the folding mechanism shown in equation (4), we assumed that only the monomers with native *trans* prolyl isomers can form productive dimers and trimers. According to studies on model peptides the Ile3-Pro4 (12% *cis*) and Ala6-Pro7 (8% *cis*) peptide bonds in foldon should lead to 19% unfolded monomers with at least one *cis* prolyl peptide bond.[22] This agrees well with $20(\pm 1)$% of slow-folding molecules observed in the interrupted refolding experiments (equation (4) and Figure 4B), if we assume that only monomers with both prolyl peptide bonds *in trans* can enter the productive folding pathway. The presence of $20(\pm 1)$% slow-folding molecules would, however, be observed if only the Ala6–Pro7 peptide bond, which is in a highly structured region of foldon (Figure 1), was essential for folding and if a *cis* bond at this position could be incorporated into partially folded dimers and trimers. For this mechanism, the presence of 22% slow-folding molecules would be expected. Our data do not allow us to discriminate between these mechanisms. However, the rate constants for the fast-folding pathway are not influenced significantly by the folding mechanism of the slow-folding molecules.

The determination of all rate constants on the folding/association pathway of the foldon domain enables us to calculate the population of each species during folding at various concentrations of protein (Figure 7). At an initial concentration of 1 μM monomer, only the burst phase intermediate (I) and the native state become populated significantly (>10%) during folding. At concentrations of protein above 5 μM, the dimer becomes populated transiently, in agreement with a change in reaction order around this concentration of protein (Figure 6). As a consequence, the fast-folding pathway changes from apparent three-state to apparent four-state with the burst phase intermediate and the dimer populated to significant amounts. Above a concentration of 50 μM monomer, also the trimeric intermediate becomes populated significantly. The significant population of dimeric and trimeric intermediates explains the low apparent half-time of the reaction at high concentrations of protein, which would suggest a unimolecular folding reaction faster than the $210(\pm 20)\ s^{-1}$ obtained from the global fit. Since the apparent half-time was determined from the fluorescence measurements (Figure 6), it will be influenced by

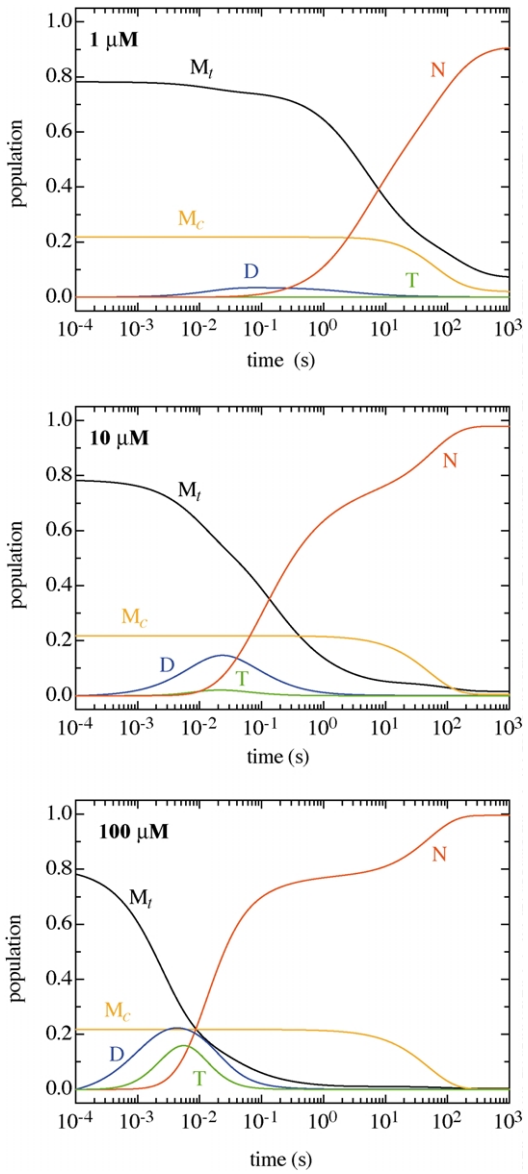


Figure 7. Simulation of the time-course of all kinetic species during folding of the foldon domain at 0.58 M GdmCl (pH 7.1), 20 °C. The results from the global fit of all kinetic data (see Figure 4) to the kinetic model shown in equation (4) were used to calculate the concentrations of the different species at the indicated values of $[M]_0$.

the kinetics of formation of dimers and trimers when these become populated significantly. These reactions are fast at high concentrations of protein, which results in apparently smaller and slightly concentration-dependent half-times for the folding reaction at high concentrations of protein. However, these half-times do not represent the true half-times for the first-order reactions.

During phage assembly in the cell, the estimated concentration of monomeric fibritin molecules is between 1 μM and 5 μM. Figure 7 shows that under these conditions neither the dimeric (D) nor the trimeric (T) intermediate becomes populated to significant amounts. At *in vivo* concentrations, the half-time for folding is between one and ten seconds, which is very fast compared to the generation time of the phage of around 30 minutes.

Fast folding of multimeric proteins compared to monomeric proteins

The folding mechanism of the foldon domain has several interesting differences compared to fast-folding reactions of small monomeric proteins. Small monomeric proteins usually do not populate partially folded states to significant amounts during folding, although high-energy intermediates were shown to be obligatory for folding of many apparent two-state folders.[23,24] Obviously, small monomeric proteins are able to avoid transient high concentrations of intermediates, which minimizes the probability of aggregation side-reactions and optimizes the shape of the free energy barriers for rapid folding.[25] In the case of the foldon domain, however, rapid formation of intermolecular interactions is essential for efficient folding. Formation of the burst phase intermediate (I) leads (i) to pronounced fluorescence changes of the single Trp residue, which is part of the β-hairpin, (ii) to chain compaction as indicated by efficient energy transfer from the two tyrosine residues to the tryptophan and (iii) to significant changes in the far-UV CD signal (data not shown). This suggests that the monomeric intermediate involves structure formation of the β-hairpin, which is in agreement with the results of NMR studies on a monomeric state of the foldon domain (A-state) observed in equilibrium at low pH. In this state, the β-hairpin forms essentially the same structure as in the folded trimer (S.M. *et al.*, unpublished results). This shows that the monomeric foldon domain has a high propensity to form a β-hairpin even in the absence of intermolecular interactions. The similarity between the burst phase intermediate and the A-state is supported by the identical fluorescence emission maxima of the two states (Figure 3B). Rapid formation of the β-hairpins in the monomers probably facilitates the subsequent association reactions, since the hairpins provide highly specific surfaces that allow fast formation of intermolecular interactions. This model is in agreement with the folding mechanisms of the dimeric GCN4 fragment, for which rapid formation of α-helical structure was shown to accelerate the subsequent bimolecular association step.[26] Obviously, optimized energy landscapes for folding of monomeric and oligomeric proteins are different. Monomeric proteins fold fastest when the intermediates are marginally less stable than the unfolded state,[25] whereas populated intermediates promote the subsequent concentration-dependent association reactions in oligomeric proteins. The rapid formation of a partially folded state in the 27 residue foldon monomer shows that small, single-domain proteins should be able to rapidly form partially folded intermediates, provided this was beneficial for efficient folding.

Conclusions

Our folding studies on the foldon domain provide the first detailed information on the folding mechanism of a fast-folding trimeric protein. Previous studies on trimeric proteins focused mainly on large proteins, which fold very slowly and are accompanied by irreversible aggregation reactions[5] or on collagens, which are limited in folding by prolyl *cis–trans* isomerization reactions.[27] Our results suggest that in the small 27 amino acid residue foldon domain, all folding and association steps are optimized for rapid formation of a stable trimer. The two consecutive association reactions show bimolecular rate-constants similar to the values found in the fastest-folding small dimeric proteins. Further, the first-order folding reaction, which becomes rate limiting at high concentrations of protein, has a rate constant comparable to those of fast-folding, small, single-domain proteins. In contrast to previously studied small multimeric proteins or protein fragments, where association involved mainly helix–helix interactions, the foldon domain trimerizes *via* β-hairpins. The high bimolecular rate constants for both association steps in the foldon domain demonstrate that β-hairpins allow very rapid and specific formation of intermolecular interactions during protein assembly.

Materials and Methods

Protein synthesis and purification

The foldon domain was synthesized chemically with an ABIMED economy peptide synthesizer EPS 221 (Abimed, Germany) using standard F-moc chemistry. Pre-coupled resin was purchased from Novabiochem, Switzerland. Amino acids were from Alexis Biochemicals, USA or from Iris Biotech, Marktredwitz, Germany. Solvents and other chemicals were from Fluka (Buchs, Switzerland). The protein was purified using HPLC with a C-8 reverse-phase preparative column (Hibar, LiChrosorb®100 RP-8 from Merck, Darmstadt, Germany). ^{13}C,^{15}N-labeled foldon for NMR spectroscopy was expressed in *E. coli* BL21(DE3) cells grown in ^{13}C,^{15}N-enriched M9 minimal medium. Protein purity was confirmed by nanospray mass spectrometry and analytical HPLC. Expression and purification of the foldon domain in *E. coli* was performed as described.[4]

NMR spectroscopy

NMR samples ([U-^{15}N]foldon 95% H_2O/5% ^2H_2O; [U-^{13}C,^{15}N]foldon 95% H_2O/5% ^2H_2O; [U-^{13}C,^{15}N]foldon 100% ^2H_2O) of 300 μl volume (Shigemi NMR microtubes) were prepared as 0.3 mM protein solutions at pH 7.1 in 5 mM sodium phosphate, 0.02% (w/v) NaN_3 without any further addition of salt. Residual alignment of [U-^{13}C,^{15}N]foldon for the measurement of residual dipolar couplings was introduced by lamellar ether/*n*-hexanol phases.[28] A set of standard triple and double-resonance assignment, quantitative *J*-coupling, nuclear Overhauser effect (NOE) spectroscopy (NOESY) and ^{15}N relaxation experiments similar to those described[29] were performed on a Bruker DRX 600 spectrometer at 25 °C. Standard data processing and analysis was carried out as described.[29]

Structure calculation

Experimental NOE distance, torsion angle and residual dipolar restraints derived from the NMR data are listed in Table 1. Structure calculations were carried out with a simulated annealing protocol using the program CNS.[30] The structural statistics for the best 40 structures are given in Table 1.

Protein Data Bank accession code

The structural statistics have been deposited in the Brookhaven Protein Data Bank with PDB accession code 1RFO.

Denaturant-induced equilibrium transitions

Denaturant-induced equilibrium transitions were recorded in an AMINCO Bowman series 2 spectrofluorimeter (SLM Aminco, USA) and with an Aviv 62ADS spectropolarimeter (Aviv, USA). All transitions were measured at 10 mM sodium phosphate (pH 7.1 at 20 °C). For fluorescence measurements at concentrations of total monomer ($[M]_0$) of 5 μM and 30 μM, the excitation wavelengths were 278 nm and 298 nm, respectively, at 2 nm bandwidth. Emission was recorded at 320 nm (2 nm bandwidth). CD measurements were performed at 228 nm with 0.5 cm (5 μM) or 0.1 cm (30 μM) path-lengths.

Equilibrium transition curves were analyzed assuming a two-state transition from native trimer (N) to unfolded monomer (U):

$$N \rightleftharpoons 3U \tag{5}$$

The resulting equilibrium constant is given by:

$$K_{eq} = \frac{[U]^3}{[N]} \tag{6}$$

and the total monomer concentration ($[M]_0$) can be expressed as:

$$[M]_0 = [U] + 3[N] \tag{7}$$

The fractions of monomers in the unfolded state (f_U) and in the native state (f_N) are given by:

$$f_U = \frac{[U]}{[U] + 3[N]} = \frac{[U]}{[M]_0} \tag{8a}$$

$$f_N = \frac{3[N]}{[U] + 3[N]} = \frac{3[N]}{[M]_0} = 1 - f_U \tag{8b}$$

Thus, equation (6) becomes:

$$K_{eq} = \frac{[U]^3}{[N]} = \frac{3f_u^3[M]_0^2}{1 - f_u} \tag{9}$$

To fit the equilibrium transitions to equation (9) we expressed the fraction of unfolded protein at a given denaturant concentration (x) as:[31]

$$f_U(x) = \frac{S_N(x) - S(x)}{(S_N(x) - S(x)) + (S(x) - S_U(x))} \tag{10}$$

where $S(x)$ corresponds to the measured signal at the given denaturant concentration, x. S_N and S_U represent the spectroscopic signals of the native and unfolded state, respectively, obtained from linear extrapolation of the baselines according to:[32]

$$S_N(x) = S_N(H_2O) + m_N x \tag{11a}$$

$$S_U(x) = S_U(H_2O) + m_U x \quad (11b)$$

Using:

$$\Delta G^0 = -RT \ln K_{eq} \quad (12)$$

and a linear denaturant-dependence of ΔG^0 :[32,33]

$$\Delta G^0(x) = \Delta G^0(H_2O) + mx \quad (13)$$

we can obtain the free energy for unfolding in water $\Delta G^0(H_2O)$ from fitting the transition curve to equations (9)–(13) in a single step, similar to the procedure described by Santoro & Bolen for a two-state transition of monomeric proteins.[34] This yields the following equation for a two-state unfolding transition from a native trimer to unfolded monomers:

$$S(x) = S_N(x) - \frac{S_N(x) - S_U(x)}{3[M]_0} \times \left(\sqrt[3]{e^{\frac{-\Delta G^0(H_2O)+mx}{RT}} \left(\frac{9}{2}[M]_0 + \sqrt{D} \right)} + \sqrt[3]{e^{\frac{-\Delta G^0(H_2O)+mx}{RT}} \left(\frac{9}{2}[M]_0 - \sqrt{D} \right)} \right) \quad (14a)$$

with:

$$D = \left(\frac{9}{2} \right)^2 [M]_0^2 + e^{\frac{-\Delta G^0(H_2O)+mx}{RT}} \quad (14b)$$

Equations (14a) and (14b) were used to fit the individual transition curves monitored by fluorescence or CD and to globally fit all data obtained by measuring various spectroscopic probes at different concentrations of total monomer ($[M]_0$).

Kinetic experiments

All stopped-flow experiments were performed with a SX18.MV stopped-flow instrument from Applied Photophysics (Leatherhead, UK) equipped with a Hamamatsu R1104 photomultiplier tube for single-wavelength kinetics and a Hamamatsu R6095 photomultiplier tube for all other measurements. For experiments using a cut-off filter (≥320 nm), the excitation wavelength was 278 nm (2 nm bandwidth). For single-wavelength refolding experiments, the excitation wavelength was 280 nm (4.5 nm bandwidth) for final concentrations of protein up to 10 μM and 295 nm (4.5 nm bandwidth) for higher concentrations. The emission bandwidth was 12 nm. For refolding experiments, the protein was allowed to unfold for at least ten hours in 6.38 M GdmCl, 10 mM sodium phosphate (pH 7.1) before the measurements. At final concentrations of protein above 10 μM monomer, the protein was unfolded at the same concentration of denaturant at pH 2. Unless stated otherwise, all measurements were performed in 10 mM sodium phosphate (pH 7.1 at 20 °C). Refolding was initiated by 11-fold dilution to a final concentration of GdmCl of 0.58 M. The fluorescence intensity of the unfolded state under refolding conditions (Figure 3) was determined by extrapolating the fluorescence signal of the unfolded foldon domain (measured in the stopped-flow instrument) to 0.58 M GdmCl. The effect of human cyclophilin 18 on the slow-refolding reaction was measured by manual mixing fluorescence measurements in 10 mM sodium cacodylate (pH 7.1) at $[M]_0 = 5$ μM foldon domain in the absence and in the presence of 3.3 μM cyclophilin. The use of sodium cacodylate instead of sodium phosphate had no effect on the folding kinetics.

Stopped-flow interrupted refolding experiments were used to monitor the formation of native molecules during refolding in 0.58 M GdmCl, 20 mM sodium phosphate (pH 7.0) at 20.0 °C. Completely unfolded foldon (in 3.4 M GdmCl, 20 mM sodium phosphate (pH 1.7), $[M]_0 = 60$ μM) was diluted sixfold into final conditions of 10 μM protein, 0.58 M GdmCl, 20 mM sodium phosphate (pH 7.0) to initiate refolding. After various times (t_i), refolding was interrupted by transferring the solution into final conditions of 6.7 M GdmCl, 20 mM sodium phosphate (pH 7.0), final protein concentration 1.7 μM. Native foldon unfolds with double-exponential kinetics under these conditions (S.Gü. & T.K., unpublished results) with $\tau_1 = 7.8(\pm 0.5)$ seconds (80% amplitude) and $\tau_2 = 0.24(\pm 0.02)$ seconds (20% amplitude). The relative amplitudes of the two reactions is independent of the refolding time. The amplitude of the major, slow-unfolding reaction ($\tau_1 = 7.8(\pm 0.5)$ seconds) was used as a measure for the amount of native protein that was present after the time t_i, when refolding was interrupted. The observed unfolding amplitudes after various times of refolding were normalized against the amplitude of completely refolded foldon to yield the fraction of native molecules that were present after t_i.

Data fitting and simulations

Data evaluation was carried out using the programs ProFit (Quantumsoft, Zurich, Switzerland) and Matlab (The MathWorks, Natick, MA, USA). Interrupted refolding experiments at $[M]_0 = 10$ μM and ten direct fluorescence-detected refolding traces with $[M]_0$ ranging from 0.5 μM to 102 μM were analyzed globally by nonlinear, least-squares curve fitting. The experimental data were fit to the numerical solution of the kinetic scheme depicted in equation (4). Rate constants and relative signal amplitudes of the different kinetic species were fitted as global parameters. The equilibrium constant between native and unfolded protein determined by a global fit of the equilibrium unfolding transitions measured by fluorescence and CD at various concentrations of protein (Figure 2) was used as an additional constraint for the fit. To ensure that the fit converged to the global minimum, it was repeated 60 times with randomly chosen starting values for the fitting parameters.

References

1. Tao, Y., Strelkov, S. V., Mesyanzhinov, V. V. & Rossmann, M. G. (1997). Structure of bacteriophage T4 fibritin: a segmented coiled coil and the role of the C-terminal domain. *Structure*, **5**, 789–798.
2. Letarov, A. V., Londer, Y. Y., Boudko, S. P. & Mesyanzhinov, V. V. (1999). The carboxy-terminal domain initiates trimerization of bacteriophage T4 fibritin. *Biochemistry (Moscow)*, **64**, 817–823.
3. Boudko, S. P., Londer, Y. Y., Letarov, A. V., Sernova, N. V., Engel, J. & Mesyanzhinov, V. V. (2002). Domain organization, folding and stability of bacteriophage T4 fibritin, a segmented coiled-coil protein. *Eur. J. Biochem.* **269**, 833–841.
4. Frank, S., Kammerer, R. A., Mechling, D., Schulthess, T., Landwehr, R. J. B., Guo, Y. *et al.* (2001). Stabilization of short collagen-like triple helices by protein engineering. *J. Mol. Biol.* **308**, 1081–1089.
5. Seckler, R. (2000). Assembly of multi-subunit

structures. In *Mechanisms of Protein Folding* (Pain, R., ed.), pp. 279–308, Oxford University Press, Oxford.
6. Stetefeld, J., Frank, S., Jenny, M., Schulthess, T., Kammerer, R. A., Boudko, S. P. *et al.* (2003). Collagen stabilization at atomic level: crystal structure of designed $(GlyProPro)_{10}$-foldon. *Structure*, **11**, 339–346.
7. Dosset, P., Hus, J. C., Blackledge, M. & Marion, D. (2000). Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J. Biomol. NMR*, **16**, 23–28.
8. Myers, J. K., Pace, C. N. & Scholtz, J. M. (1995). Denaturant *m* values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4**, 2138–2148.
9. Milla, M. E. & Sauer, R. T. (1994). P22 arc repressor: folding kinetics of a single-domain, dimeric protein. *Biochemistry*, **33**, 1125–1133.
10. Fischer, G., Wittmann, L. B., Lang, K., Kiefhaber, T. & Schmid, F. X. (1989). Cyclophilin and peptidyl-prolyl *cis–trans* isomerase are probably identical proteins. *Nature*, **337**, 476–478.
11. Schmid, F. X. (1983). Mechanism of folding of ribonuclease A. Slow refolding is a sequential reaction *via* structural intermediates. *Biochemistry*, **22**, 4690–4696.
12. Kiefhaber, T. (1995). Protein folding kinetics. In *Methods in Molecular Biology* (Shirley, B. A., ed.), vol. 40, pp. 313–341, Humana Press, Totowa, NJ.
13. Bieri, O. & Kiefhaber, T. (2000). Kinetic models in protein folding. In *Protein Folding: Frontiers in Molecular Biology* (Pain, R., ed.), 2nd edit., pp. 34–64, Oxford University Press, Oxford.
14. Moore, J. W. & Pearson, R. G. (1981). *Kinetics and Mechanisms*, Wiley, New York.
15. Gutfreund, H. (1995). *Kinetics for the Life Sciences*, Cambridge University Press, Cambridge.
16. Jackson, S. E. (1998). How do small single-domain proteins fold? *Fold. Des.* **3**, R81–R91.
17. Zitzewitz, J. A., Bilsel, O., Luo, J., Jones, B. E. & Matthews, C. R. (1995). Probing the folding mechanism of a leucine zipper peptide by stopped-flow circular dichroism spectroscopy. *Biochemistry*, **34**, 12812–12819.
18. Jaenicke, R. (1987). Folding and association of proteins. *Prog. Biophys. Mol. Biol.* **49**, 117–237.
19. Wendt, H., Berger, C., Baici, A., Thomas, R. M. & Bosshard, H. R. (1995). Kinetics of folding of leucine zipper domains. *Biochemistry*, **34**, 4097–4107.
20. Waldburger, C. D., Jonsson, T. & Sauer, R. T. (1996). Barriers to protein folding: formation of buried polar interactions is a slow step in acquisition of structure. *Proc. Natl Acad. Sci. USA*, **93**, 2629–2634.
21. Gloss, L. M. & Matthews, C. R. (1998). Mechanism of folding of the dimeric core domain of *Escherichia coli* Trp repressor: a nearly diffusion-limited reaction leads to the formation of an on-pathway intermediate. *Biochemistry*, **37**, 15990–15999.
22. Reimer, U., Scherer, G., Drewello, M., Kruber, S., Schutkowski, M. & Fischer, G. (1998). Side-chain effects on peptidyl-prolyl *cis/trans* isomerization. *J. Mol. Biol.* **279**, 449–460.
23. Bachmann, A. & Kiefhaber, T. (2001). Apparent two-state tendamistat folding is a sequential process along a defined route. *J. Mol. Biol.* **306**, 375–386.
24. Sánchez, I. E. & Kiefhaber, T. (2003). Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J. Mol. Biol.* **325**, 367–376.
25. Wagner, C. & Kiefhaber, T. (1999). Intermediates can accelerate protein folding. *Proc. Natl Acad. Sci. USA*, **96**, 6716–6721.
26. Zitzewitz, J. A., Ibarra-Molero, B., Fishel, D. R., Terry, K. L. & Matthews, C. R. (2000). Preformed secondary structure drives the association reaction of GCN4-p1, a model coiled-coil system. *J. Mol. Biol.* **296**, 1105–1116.
27. Boudko, S. P., Frank, S., Kammerer, R. A., Stetefeld, J., Schulthess, T., Landwehr, R. *et al.* (2002). Nucleation and propagation of the collagen triple helix in single-chain and trimerized peptides: transition from third to first order kinetics. *J. Mol. Biol.* **317**, 459–470.
28. Ruckert, M. & Otting, G. (2000). Alignment of biological macromolecules in novel non-ionic liquid crystalline media for NMR experiments. *J. Am. Chem. Soc.* **122**, 7793–7797.
29. Kahmann, J. D., Sass, H. J., Allan, M. G., Seto, H., Thompson, C. J. & Grzesiek, S. (2003). Structural basis for antibiotic recognition by the TipA class of multidrug-resistance transcriptional regulators. *EMBO J.* **22**, 1824–1834.
30. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
31. Tanford, C. (1968). Protein denaturation. Part B. The transition from native to denatured state. *Advan. Protein Chem.* **23**, 218–282.
32. Pace, C. N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* **131**, 266–280.
33. Greene, R. F. J. & Pace, C. N. (1974). Urea and guanidine–hydrochloride denaturation of ribonuclease, lysozyme, alpha-chyomtrypsinn and beta-lactoglobulin. *J. Biol. Chem.* **249**, 5388–5393.
34. Santoro, M. M. & Bolen, D. W. (1988). Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry*, **27**, 8063–8068.
35. Cornilescu, G., Marquardt, J. L., Ottiger, M. & Bax, A. (1998). Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J. Am. Chem. Soc.*, 6836–6837.
36. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. (1996). AQUA and PROCHECK-NMR programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, **8**, 477–486.
37. Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55.

Edited by F. Schmid

(Received 24 November 2003; received in revised form 30 January 2004; accepted 5 February 2004)

3.2 The foldon A-state

We quantify the electrostatic contributions to the stability of foldon and show that a partially folded state can be highly populated at low pH. Structure and folding of the A-state are evaluated with NMR spectroscopic methods.

3.2.1 Electrostatic contributions to foldon trimer stability

The foldon trimer is stabilized by hydrophobic interactions, inter-subunit hydrogen bonds between the three hairpins and by a central inter-subunit salt bridge between E5 and R15 (Figure 3–1). The co-titration of $^{13}C^{\gamma}$(E5) and $^{13}C^{\delta}$(R15) resonances validates the presence of this inter-subunit ion bond (Figure 3–2 A). The polar sidechain-backbone interactions D9-Q11H^N and D17-D17H^N are evident from the structure, from slow $^1H/^2H$ exchange of the amides at 25 °C and pH 7 ($t_{1/2}$>30 min) and from co-titration of the respective $^{13}C^{\gamma}$ sidechain and $^1H^N$ and ^{15}N backbone resonances (Figure 3–2 B and C). An intramolecular salt bridge D17-K16 proposed from the fibritin C-terminal crystal structure (Tao et al., 1997) is not found in solution, as evidenced by random coil chemical shifts of the K16 sidechain over the whole pH range.

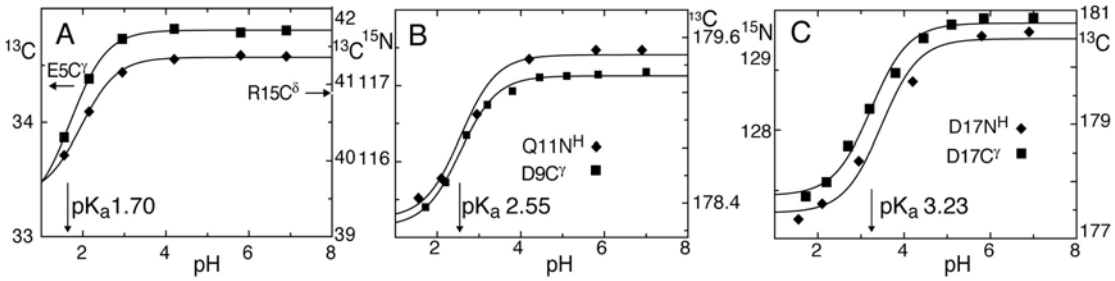


Figure 3–2. Polar interactions in foldon as evidenced by the co-titration of chemical shifts of both interaction partners for E5-R15, D9-Q11N^H and D17-D17N^H. Fits to the Henderson-Hasselbalch equation are shown, pK$_a$ values of the carboxylate groups are indicated.

The interactions E5-R15, D9-Q11H^N and D17-D17H^N contribute to the free energy of trimer unfolding (Figure 3–3), whereas the remaining acidic sidechain (E19, pK$_a$ 4.37) is solvent-exposed in the trimer. The pH dependent stabilization energy for each site i is calculated as $\Delta G_{FU}^{pH}(i) = -RT\ln\left(\frac{\left[H^+\right]+K_a^U(i)}{\left[H^+\right]+K_a^F(i)}\right)$ (Tollinger et al., 2003), where

K_a^U and K_a^F are the K_a values of the unfolded and the folded state (Figure 3–3). In unfolded states, pK_a values only have a weak amino acid sequence dependence (Tollinger et al., 2003) and random coil values of $pK_a^U(E) = 4.40$ and $pK_a^U(D) = 3.90$ can safely be assumed.

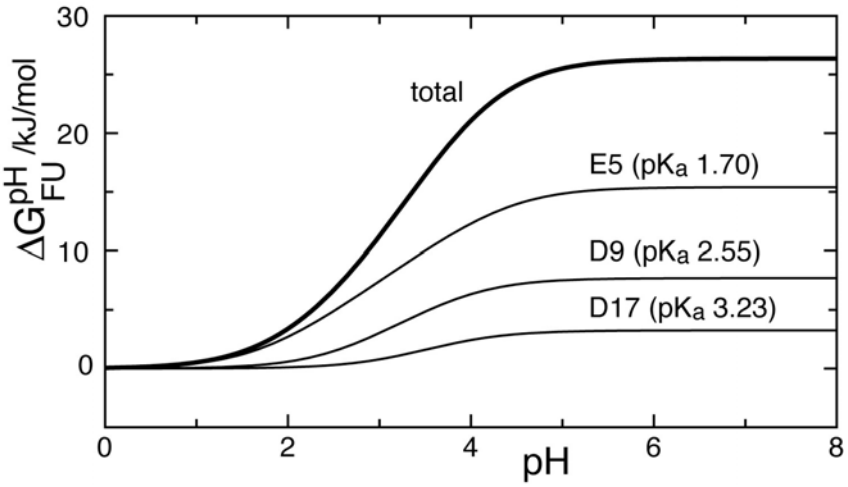


Figure 3–3. Contributions of ionic interactions in foldon to the free energy of unfolding. Free energies per monomer are shown for the inter-subunit E5-R15 salt bridge and for interactions of D9 and D17 with local backbone amide protons. The sum of these contributions is indicated. Random coil pK_a values are assumed for the unfolded state ($pK_a^U(E) = 4.40$, $pK_a^U(D) = 3.90$)

The inter-subunit salt bridge E5-R15 is highly solvent-shielded and results in a $pK_a(E5) = 1.70$, thus amounting to a stabilization energy of $2.303 \cdot RT(pK_a^F - pK_a^U) = -16$ kJ/mol per monomer. Titration of this inter-subunit ion bond leads to the appearance of a second folded spectral species at lower pH values and decreasing protein concentrations. This concentration dependence as well as relaxation data show that this A-state is monomeric.

3.2.2 Monomer structure and folding

The analysis of foldon assembly has shown that the trimer forms in two sequential bimolecular association reactions starting from a collapsed monomeric folding intermediate I (Güthe et al., 2004). Thus, the association scheme is

$$3U \underset{k_u}{\overset{k_f}{\rightleftharpoons}} 3I \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} D+I \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} T \quad .$$

The intermediate I is formed in the dead-time of stopped-flow mixing (about 1 ms) from the unfolded state U. Thus, k_f and k_u are too large to be derived from stopped-flow fluorescence data (Güthe et al., 2004). The intermediate has similar fluorescence properties as the monomeric acidic state of foldon. Therefore, this A-state can be regarded as the trapped folding intermediate, which is stabilized by diminishing electrostatic interactions in the association reaction. We demonstrate that the foldon A-state structure contains a stable single hairpin, which is highly twisted in the monomer, presumably favoring trimerization to a β-propeller over competing aggregation side-reactions. Thus, the β-hairpin obviously is the folding nucleus of foldon.

The understanding of α-helix-, β-hairpin- and loop-formation is of general interest, as it allows insights into the kinetics and thermodynamics of the earliest local protein folding events. While α-helix model systems have been described theoretically and experimentally for more than four decades (Zimm and Bragg, 1959; Zimm et al., 1959; Lifson, 1961), stable isolated β-hairpins as model systems for β-sheet formation have been described rather recently (Blanco et al., 1994). Hairpin sequences extracted from proteins often either have the tendency to polymerize into extended β-sheets or are devoid of any structure in solution (Blanco et al., 1998). The discovery of protein aggregates associated with folding diseases (Pan et al., 1993) has further sparked the interest in β-sheet formation.

Foldon has been suggested to exhibit an optimized folding and association pathway, as its only known biological function is to promote the folding and trimerization of fibritin (Boudko et al., 2002). In structural terms, this stands for the fast and stable formation of structures. Upon titration to pH 2, the foldon trimer can be disrupted by weakening the strong inter-subunit salt bridge E5-R15, resulting in a monomeric partially folded A-state (Figure 3–4). The structure of the A-state was determined and shows that the single β-hairpin of only 12 amino acids length (A12-L23) adopts its native structure with intact hydrogen bonds in the A-state, whereas the rest of the foldon sequence folds up synergistically to its native structure only during association.

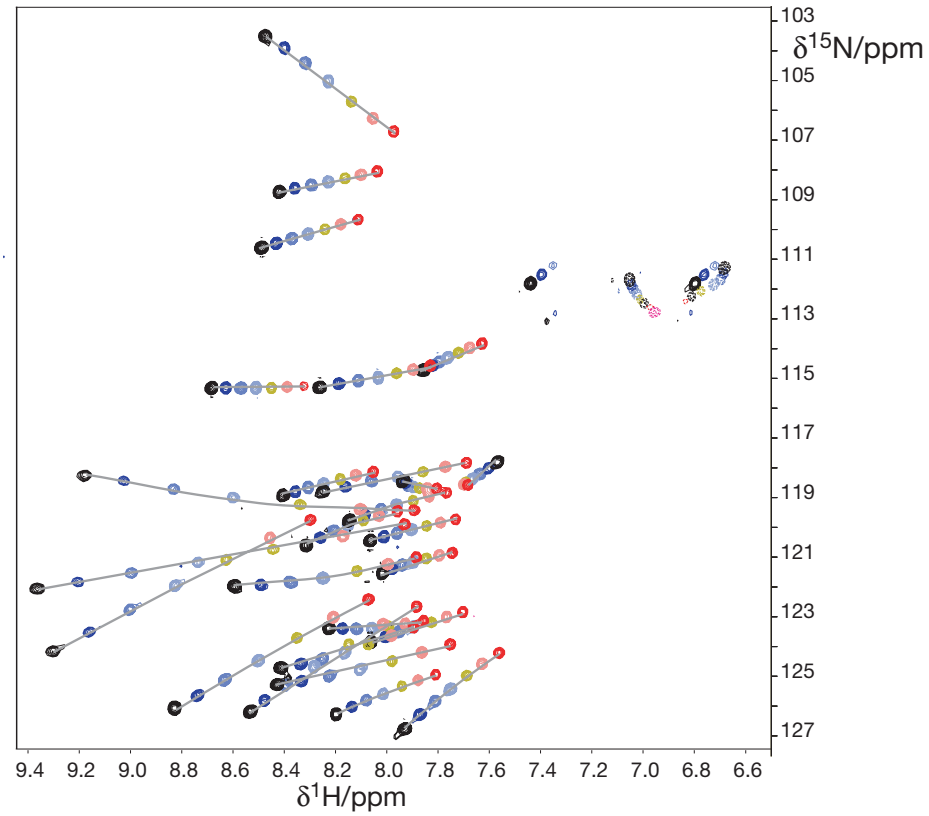


Figure 3–4. HSQC spectra of the foldon A-state between 25 °C and 85 °C in intervals ΔT=10 °C, at pH 2, 50 μM peptide concentration. Correlation spectra show fast chemical exchange between a folded and an unfolded species and yield the thermodynamics of hairpin folding from thermal chemical shift transitions.

We have assessed the thermodynamics and kinetics of foldon hairpin formation by NMR spectroscopy. The A-state hairpin is highly stable in aqueous solution as compared to other native hairpins (Blanco et al., 1994; Zerella et al., 1999). An

analysis of ^{15}N relaxation data shows that folding and unfolding of the A-state β-hairpin (resembling $1/k_f$ and $1/k_u$ in the above folding scheme) occur within few microseconds, as has been observed for other hairpins by detecting fluorescence or infrared spectral changes upon laser-induced temperature-jumps (Munoz et al., 1997; Xu et al., 2003).

Hairpins thus form on considerably slower timescales (roughly tenfold) than α-helices. Nevertheless such μs-dynamics phenomena have become accessible to molecular dynamics (MD) simulations. Thus, MD can be used to study folding and unfolding processes of small secondary structural units at atomic resolution. However, experimental tests of such predictions are difficult. We use the alignment of the A-state in a mechanically strained polyacrylamide matrix to monitor local disorder upon equilibrium unfolding in a residue specific manner and compare the results to predictions from MD simulations. Residual structural preferences are detected experimentally throughout thermal unfolding, including amino acid stretches without tertiary contacts. This limits the conformational freedom of the unfolded state and thus reduces the entropy loss upon protein folding.

The detailed results on the characterization of the foldon monomer are described in the following manuscript, which is in preparation for the Journal of Molecular Biology.

Original Publication

Meier, S., Güthe, S., Kiefhaber, T., and Grzesiek, S.

Structure and dynamics in the fibritin folding nucleus containing a stable β-hairpin

in preparation

Structure and dynamics in the fibritin folding nucleus containing a stable β-hairpin

Sebastian Meier[1], Sarah Güthe[2], Thomas Kiefhaber[2], and Stephan Grzesiek[1*]

[1]Division of Structural Biology and [2]Division of Biophysical Chemistry, Biozentrum der Universität Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

INTRODUCTION

Fibritin is an elongated, trimeric, 486 amino acid protein of the bacteriophage T4 head with structural, chaperone as well as sensory functions [1-3]. Fibritin consists of an N-terminal anchor domain (residues 1–46), a large central coiled-coil part (residues 47–456), and a small C-terminal globular domain (residues 457–486). The structures of C-terminal constructs comprising 120 and 75 amino acids of T4 fibritin have been solved by crystallography [4,5]. The C-terminal part of the fibritin trimer is formed from a trimeric coiled-coil domain, which is terminated by a C-terminal trimeric β-sheet propeller consisting of monomeric β-hairpin segments (Figure 1). This β-propeller is necessary for the correct folding of the holoprotein [6] and hence has been termed foldon. Trimerization occurs at temperatures of up to 75 °C and SDS concentrations of up to 2% at room temperature [6]. Due to the extreme stability of the trimer, foldon can be used as an artificial trimerization inducer or enhancer. Thus, collagen [7] and HIV gp140 [8] trimers have been stabilized via the attachment of the foldon domain.

Kinetic data indicate that the folding of foldon is a multi-step process [9], where trimerization starts from a collapsed monomeric intermediate formed on the submillisecond time scale. This monomeric intermediate has similar fluorescence properties (emission maximum at 330 nm) as the monomeric foldon A-state, which is obtained at pH values below 2 and concentrations of lower than ~200 μM. Trimerization proceeds from the collapsed monomer intermediate by two sequential bimolecular association steps with rate constants that are close to the fastest association rates observed for proteins in solution. The extremely fast association reflects the biological role of foldon as a very effective trimerization domain during phage assembly, where fibritin concentrations in the host are on the order of a few μM [9].

Clearly, the detailed understanding of protein folding is of fundamental importance [10]. The interest has further been sparked by the discovery of folding diseases where misfolded proteins reach a seemingly global energy minimum comprised of fibrillar

β-strand amyloid structures [11,12]. While model peptides for the folding of α-helices have been described extensively for several decades [13-15], the folding of isolated β-hairpins, representing the minimal model for β-sheet formation, has only been studied more recently [16,17]. The understanding is complicated by the various non-sequential interactions involved in hairpin- and β-sheet formation. The folding and unfolding of isolated β-hairpins has been shown to occur on a time scale of few microseconds, and the experimental evidence indicates that folding usually occurs as a broad two-state transition due to the relatively small folding enthalpies [18].

Solution NMR can monitor folding transitions and equilibria at atomic resolution. Traditionally, chemical shift and NOE information are used [19], but more recently also the weak alignment of proteins in anisotropic liquid phases [20] has become a tool to directly monitor average net orientations and order parameters of individual bonds by residual dipolar couplings (RDCs) [21-25]. The introduction of inert orienting media, such as strained polyacrylamide gels [26,27], has made it possible to obtain weak alignment and RDC information even under relatively harsh unfolding conditions. RDC techniques allow a quick characterization of the topology of folded and also of only partially folded states.

In the present study, we have used modern solution state NMR methods to characterize the details of the transition from the foldon trimer to the A-state monomer, to determine the monomer structure, and to characterize its thermal unfolding. The protonation of the E5 sidechain below pH 2, and the consequent opening of the E5-R15 salt bridge are responsible for the trimer-monomer transition. The A-state monomer forms a β-hairpin with native backbone structure, including native hydrogen bonds and thermally stable sidechain contacts. Relaxation data show that the hairpin amides undergo μs motions, presumably representing the overall folding and unfolding transition. The determination of RDCs during thermal unfolding in mechanically strained polyacrylamide gels permits the direct observation of local order of NH^N and $C^\alpha H^\alpha$ moieties and a breakdown into different contributions from hydrogen bonds and sidechain contacts. The apparent two-state folding of the hairpin can thus be dissected into residue specific phenomena, which can be compared to predictions from various molecular dynamics simulations and statistical

mechanical models. Local structures in the turn and near large hydrophobic sidechains prove to be highly resistant to thermal unfolding and are therefore likely to initiate correct hairpin folding by a hydrophobic collapse steered to the native hydrogen bond register.

RESULTS and DISCUSSION

Native trimer structure

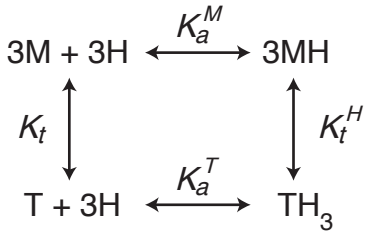
The isolated foldon trimer in solution at pH 7 (Figure 1) adopts a structure that is very similar to the foldon part of the larger protein constructs in crystalline form [9]. Within the trimer, the foldon monomers (residues G1–L27) consist of an extended N-terminal region (G1-Q11), a β-hairpin (A12-L23), and a C-terminal 3_{10} helix (L23–L27). The extended N-terminal region contains a polyproline II helix between residues P4 and P7 and packs against one side of the β-hairpin by hydrophobic contacts. The central β-hairpin is short, with five residues per strand (A12-K16 and E19-L23) and strongly twisted. According to a definition used by Hermans and coworkers [28], the right hand twist amounts to an angle of more than 60° between two cross-strand $C^{\alpha}C^{\alpha}$ vectors of sequential hydrogen bonded residue pairs (e.g. $C^{\alpha}C^{\alpha}$ V14-V21 and $C^{\alpha}C^{\alpha}$ K16-E19, see Figure 1A). This twist leads to a very compact β-propeller topology and presumably disfavors aggregation into extended sheets. The hairpin is stabilized by a cluster of hydrophobic residues at its N- and C-terminal ends (A12–V14, V21–L23), and a diagonal interaction between Y13 and W20 seems to promote the hairpin twist. Both hairpin strands are connected by a type I' β–turn consisting of residues K16 to E19. Together with R15, residues K16, D17, and E19 form a charged cluster in the turn region of the hairpin.

The foldon trimer interface consists of many hydrophobic interactions along the trimer symmetry axis, involving residues Y2 and I3 at the N-terminus, V14 in the hairpin, as well as L23, F26, and L27 in the C-terminal 3_{10} helix (Figure 1C). The crucial inter-subunit salt bridges (see below) between residues E5 and R15 and the inter-subunit hydrogen bonds between Y13 and R15 further contribute to the extreme stability of the trimer.

The monomeric A-state

When changing the pH from neutral to acidic, a second species appears in the NMR spectra of foldon, which is in slow chemical exchange with the foldon trimer state (Figure 2A). This second species becomes increasingly populated between pH 4.3 and pH 1.6 (Figure 2C) and apparently corresponds to the acidic state (A-state) of foldon. Lowering the protein concentration while keeping the pH at 2 leads to a relative intensity increase of this A-state species with respect to the trimer state (Figure 2B). This fact, the ^{15}N-relaxation data of the A-state (see below), its fast exchange with an unfolded species over a broad temperature range, and the high similarity of its fluorescence spectra with the monomeric folding intermediate indicate that the A-state of foldon is monomeric. The intensity of spectral lines corresponding to trimer and monomer becomes approximately equal at pH 1.7 for 250 μM total protein concentration (Figure 2C). Thus, the trimer concentration is about 1/3 of the monomer concentration under these conditions. As trimer and monomer coexist in slow chemical exchange, their separate sets of resonances could be assigned individually by standard heteronuclear 2D NMR experiments. In this way, complete backbone and sidechain chemical shift assignments were obtained. No further spectral species are detectable for a threshold of about 5% of the monomer A-state population. Thus, we conclude that other intermediates, such as dimers, are not significantly populated.

The reason for the disintegration of the trimer at low pH values can be traced back to the breaking of the inter-subunit E5-R15 salt bridge (Figure 1D). A titration of the sidechain carboxylate ^{13}C resonances for all glutamic and aspartic acid residues (Figure 2D) and of the aliphatic E5-$^{13}C^{\gamma}$ and R15-$^{13}C^{\delta}$ resonances (Figure 2E) shows that the pK_a for E5 is strongly shifted to a value of 1.7, and that the E5 and R15 sidechain resonances co-titrate according to this pK_a value. Thus, the protonation state of the E5-R15 salt bridge closely follows the monomer-trimer equilibrium as judged from the monomer and trimer resonance intensities in Figure 2C.
The following simplified reaction scheme expresses this dependence of the monomer-trimer equilibrium on the monomer protonation state:



(1)

where $K_a^M = \frac{[M][H]}{[MH]}$, $K_a^T = \frac{[T][H]^3}{[TH_3]}$, $K_t^H = \frac{[MH]^3}{[TH_3]}$, and $K_t = \frac{[M]^3}{[T]}$.

An analysis of the spectral data yields values for all four equilibrium constants: thus, the midpoint of the protonation equilibrium of E5 in the trimer is at pH 1.7, yielding a value for K_a^T of $10^{-5.1}$ M^3, whereas for the free glutamic acid sidechain in the monomer a K_a^M value of $10^{-4.4}$ M can be assumed. Neglecting an appreciable population of deprotonated monomer M at pH 1.7, the monomer-trimer equilibrium constant in the protonated state can be calculated as $K_t^H = [MH]^3/[TH_3] = [0.125\ mM]^3/[0.125\ mM/6] = 9.4 \cdot 10^{-8}$ M^2, because the observed monomer (MH) and total trimer (T + TH_3) intensities are approximately equal for a total foldon concentration of 250 μM, and $[T] = [TH_3]$ at this pH value. Finally, the circular reaction scheme in Equation 1 leads to the condition that $K_t = K_t^H \cdot (K_a^M)^3 / K_a^T = 7.4 \cdot 10^{-16}$ M^2.

The difference in the protonation equilibrium constants for the E5 carboxylate in the monomer and trimer correspond to a stabilization energy of $RT \cdot (\ln K_a^T - 3 \cdot \ln K_a^M) = 48$ kJ per mole trimer at physiological pH and 25 °C. This substantial strength of the inter-subunit salt-bridge can be explained by the highly hydrophobic interior of the foldon trimer, which also leads to pronounced blue shifts in the fluorescence emission spectrum of the interior residue W20 [9].

When both monomer and trimer species are present at low pH, no line broadening from intermediate exchange can be observed in the ^1H-^{15}N correlation spectra, even for small chemical shift differences ($|\Delta\nu(^1H)| < 50$ Hz) between the two forms. This indicates that the exchange time between monomer and trimer forms must be significantly longer than 20 ms. Indeed, results of chemical exchange spectroscopy

show that the trimerization occurs on a time scale >800 ms at room temperature, whereas an onset of trimerization is observed within ≈800 ms above 40 °C.

The A-state monomer structure

The resonances of the A-state monomer show strong chemical shift dispersion for hairpin residues A12–L23 (e.g. Figure 2B). For this region, the secondary chemical shifts indicate an intact β-sheet structure. In particular, the pronounced downfield shifts of the $^{1}H^{N}$ amide resonances (V14, K16, D17, V21) are a sign of strong and intact β-sheet hydrogen bonds [29,30]. In contrast, residues outside of the hairpin region, i.e. S0–A12 and S24–L27, exhibit random coil ^{1}H, ^{13}C, and ^{15}N chemical shifts. Therefore, according to the chemical shifts the A-state monomer adopts a defined structure within in the hairpin, but random coil conformations elsewhere.

An analysis of $^{1}D_{NH}$ and $^{1}D_{C\alpha H\alpha}$ RDCs (Figure 3A,B) obtained for the A-state monomer in strained polyacrylamide gels corroborates this picture. RDCs in the hairpin are large and consistent with a rigid structure, whereas outside of this region RDCs adopt smaller negative ($^{1}D_{NH}$) or positive values ($^{1}D_{C\alpha H\alpha}$), consistent with a largely mobile polypeptide chain that orients relative to an external director [25]. In order to test the structural similarity of foldon A-state monomer and trimer, orientation tensors were derived by linear fits [31] from the trimer structure and from all $^{1}D_{NH}$ and $^{1}D_{C\alpha H\alpha}$ RDCs observed within segments of 10 consecutive amino acids. The position of these decapeptide segments was moved through the entire foldon sequence, i.e. spanning Y2–Q11, I3–A12, …, G18–L27, and orientation tensors were derived for each. Theoretical RDCs were then back-calculated from these 17 orientation tensors and compared to the observed RDCs. Figure 3C shows the NMR Q-factors ($= \mathrm{rms}(D^{obs} - D^{calc})/\mathrm{rms}(D^{obs})$ [32]) for the different decapeptides as a function of their average position. Clearly, the Q-factors have distinct minimal values (~0.36) when the decapeptides are taken from the region between G10 and S24. Outside of this region, the Q-factors increase strongly, e.g. to values around 0.7 at the N-terminus. This simple comparison proves that the A-state monomer structure is very similar to the trimer structure within the hairpin, but that it is different and unstructured elsewhere. Figure 3A,B visualizes this finding by showing calculated

$^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ RDCs for the entire foldon based on an orientation tensor that was derived from RDCs of the hairpin region only (residues A12–L23). Within the hairpin, the calculated and observed RDCs are almost identical. Elsewhere the calculated RDCs vary strongly and deviate significantly from the observed RDCs, which are more uniform than expected for a folded form and moderate in size.

Despite of the inherent dynamics of the acid collapsed monomeric state and the low protein concentration (< 0.2 mM) that is necessary to populate the monomer at pH 2, it was possible to determine its three-dimensional structure (Figure 4) from a limited set of distance restraints and RDCs (Table 1). The structure is well defined with a heavy atom backbone RMSD of 0.29 Å for residues G10–L23 representing the hairpin and a small, kinked N-terminal extension (G10, Q11). Within the hairpin, the β-sheet hydrogen bonds show a well-defined in-line geometry. This is directly evident from the NH^N RDCs for V14/L22, V21/R15, and K16/W20 (Figure 3A), which have pairwise very similar values due to the identical orientation of donor and acceptor amide groups within the hydrogen bonds. As NOEs that contradict this structural model are not found, we conclude that misfolded conformations are not highly populated.

Tertiary NOE interactions between the hairpin and the unstructured N- and C-terminal parts of foldon monomer are not observed. The absence of such tertiary contacts is also evident from details in the chemical shifts. Strong ring-current effects from contacts to the aromatic sidechains in the hairpin are absent in the monomer, e.g. the ^1H chemical shift of the A6 methyl group adopts a random coil value of 1.2 ppm in the monomer, but is strongly shifted to –0.6 ppm in the trimer because of the packing against the sidechains of Y13 and W20. The trimer formation also induces rearrangements of the sidechains within the hairpin. In particular, the variations in the NOE patterns indicate that the sidechains of Y13, R15, W20, and L23 change their conformation and the sidechains pack into a more compact hydrophobic core upon trimerization (Figure 4B,C).

The structure of the foldon monomer shows that the β-sheet hairpin conformation is intact even without further tertiary monomer or quartery trimer interactions. Thus,

foldon has an autonomously folding β-hairpin motif. The foldon A-state most likely resembles the monomeric folding intermediate, which is stabilized at low pH by weakening the association reaction. Fluorescence data, indeed, underline the similarity between the A-state and the on-path folding intermediate. The monomer with its intact isolated hairpin thus appears as a stable, low energy intermediate in folding of the foldon trimer. Early formation of hairpins in protein folding has been reported also e.g. for a staphylococcal nuclease [33] and for protein L and protein G [34].

Quantitative analysis of β-hairpin formation

The presence and stability of secondary structure elements can be conveniently monitored by NMR, due to the strong structure-dependence of chemical shifts. Chemical shift melting curves are highly reproducible and therefore suitable for quantitative thermodynamic analysis. Especially $^{1}H^{\alpha}$ data have previously been used to assess the thermodynamics of hairpin formation [35-37]. We observe thermal equilibrium unfolding of the foldon monomer by $^{1}H^{\alpha}$, $^{13}C^{\alpha}$, ^{15}N and $^{1}H^{N}$ chemical shifts obtained from two-dimensional heteronuclear correlation spectra.

All sites in the hairpin exhibit cooperative two-state transitions upon thermal unfolding, as observed by all of the monitored chemical shifts (Figure 5). As the two states are in fast chemical exchange, observed shifts are a weighted average of both, folded and unfolded states. Thus, the temperature dependence of the measured chemical shifts could be fitted to the functional form

$$\begin{aligned}\delta X &= p_u \cdot \delta X_u + p_f \cdot \delta X_f = \\ &= \delta X_u + (\delta X_f - \delta X_u)[\exp(-\Delta G^0 / RT)]/[1 + \exp(-\Delta G^0 / RT)] \quad , \end{aligned} \tag{2}$$

where δX is the observed chemical shift, δX_u and δX_f are the chemical shifts of the unfolded and folded state respectively, p_u and p_f are the populations of the unfolded and folded state with $p_u+p_f = 1$ and ΔG^0 is the free energy of folding obtained from the Gibbs-Helmholtz formula [38] as

$$\Delta G^0 = (\Delta H^0 + \Delta C^0{}_p (T - 298)) - T(\Delta S^0 + \Delta C^0{}_p \ln(T/298)).$$

The foldon A-state β-hairpin is highly populated even at room temperature (Figure 5). Therefore, values for δX_f can be obtained from the thermal chemical shift transitions and have been validated by comparison to chemical shifts of the trimer. Values for δX_u are also obtained from thermal chemical shift transitions and have been validated by comparison to random coil chemical shifts (Figure 5). The fits are not highly robust with respect to ΔC_p, which is small for peptide folding and is commonly approximated as $\Delta C_p \approx 0$ to minimize errors in fitting of the other thermodynamic parameters of peptide folding [37].

$^1H^\alpha$ and $^{13}C^\alpha$ shifts strongly depend on the backbone (ϕ/ψ) angles and have been used to monitor secondary structure populations. $^1H^\alpha$ and $^{13}C^\alpha$ resonance frequencies of the A-state hairpin were fitted to the Gibbs-Helmholtz equation with a two-state model for each site, which displays significant conformational shift changes (>0.1 ppm) upon thermal unfolding (Table 2). This yields apparent folding enthalpies $\Delta H^0 = -50.7$ kJ/mol and average entropies of folding $\Delta S^0 = -152$ J/(mol·K) in H_2O (Table 2) for the hairpin strand residues (13-16,19-22). The foldon hairpin is populated to more than 80 % at room temperature in the A-state. Thus, the A-state hairpin is a highly stable native-like hairpin with its thermodynamics in aqueous solution comparable to designed hairpins stabilized by the addition of 50% methanol [36] or 30% trifluoroethanol [37].

Strand residues in the foldon A-state β-hairpin have melting temperatures $T_m = 332.6 \pm 2.7$ K, whereas apparent melting temperatures for the β–turn resonances of D17 and G18 are increased by $\Delta T_m = 10$ K (Table 2), as evidenced from the behavior of the $^{13}C^\alpha$ and $^1H^\alpha$ shift of D17 (Figure 5A) or the frequently used glycine methylene proton dispersion [39] (Figure 5B). An increased thermal stability of up to $\Delta T_m = 15$ K in various turns has been observed from $^1H^\alpha$ chemical shifts in engineered hairpins [37] and interpreted as indication of local stability and hairpin nucleation from the turn, which is also predicted by kinetic [40] and thermodynamic [41] zipper models. Indeed, the sequences DG and NG have the highest propensities for type I' β–turns in the RCSB Protein Data Bank [42].

The possibility has been pointed out that the folded structure of β-hairpins is a molten-globule state stabilized in an extended backbone conformation by sidechain interactions without intact hydrogen bonds [19]. The $^1H^N$ shift dispersion in the foldon A-state hairpin indicates the presence of strong hydrogen bonds at room temperature. $^1H^N$ shift transitions show the melting of hydrogen bonds during the thermal unfolding (Figure 5C) of the foldon hairpin by cooperative transitions. Obtaining reliable thermodynamic parameters is, however, hindered by the complicated behavior of $^1H^N$ chemical shifts, which results from dependencies on secondary structure and on hydrogen bonding, both within the polypeptide chain and with the solvent. $^1H^N$ resonances of K16 and E19 near the turn have values almost identical to the resonances in the trimer and point to particularly highly populated hydrogen bonds with similar strengths as in the native structure. Theoretical predictions have proposed that these hydrogen bonds near the turn get stabilized at higher temperatures and are more highly populated than the neighboring, energetically stronger in-line hydrogen bonds [41].

^{15}N chemical shifts depend on direct and adjacent hydrogen bonds, as well as on backbone and sidechain conformations [43] and are therefore not well suited to specifically monitor backbone conformations or hydrogen bonds, but will yield an apparent average of both effects. The strong down- and upfield shifts of D17 and G18 at room temperature are characteristic of type I' β–turn topology [44]. Accordingly, in all low-energy structures calculated of the monomer, residues D17 and G18 are in canonical type I' β–turn conformations, although no dihedral angle restraints are imposed on the turn during structure calculations.

Relaxation data and folding kinetics

The unfolded and the folded hairpin are in fast chemical exchange on the chemical shift time scale and follow an apparent two-state folding reaction

$$U \underset{k_u}{\overset{k_f}{\rightleftharpoons}} F \quad . \qquad (3)$$

The exchange rate constant $k_{ex} = 1/\tau_{ex}$ is defined as $k_{ex} = k_f+k_u = k_f/p_f = k_u/p_u$ [45], where p_f and p_u are the equilibrium populations of states F and U, respectively, with $p_f+p_u = 1$. In the fast chemical exchange limit ($k_{ex} >> \Delta\omega$) between folded and unfolded state state with resonance frequency difference $\Delta\omega = (\omega_f-\omega_u)$, the chemical exchange contributes additively to the transverse relaxation rate R_2 observed in NMR experiments [45] as

$$R_{2ex} = p_f \cdot p_u \cdot (\omega_f-\omega_u)^2 \cdot \tau_{ex} . \qquad (4)$$

In this fast exchange regime, populations of the two states and the exchange time τ_{ex} cannot be derived independently from the relaxation data. However, thermal transitions of the chemical shifts yield populations and chemical shift differences of the two states, thus making a derivation of τ_{ex} accessible. Values of $\Delta\omega(^{15}N) = \omega_f-\omega_u$ were obtained from thermal transitions of ^{15}N resonances (Figure 4D and Table 3) in the foldon hairpin. Thus, the dynamics in the hairpin can be quantified by obtaining the R_{2ex} terms from a Lipari-Szabo analysis of ^{15}N NMR R_1 and R_2(CPMG) relaxation data and $\{^1H\}$-^{15}N steady state NOEs [46] at room temperature (25 °C) and near the hairpin transition temperature T_m (55 °C).

Fits using model free spectral density functions (data not shown) indicate low mobility on the ps-ns timescale for all hairpin residues at 25 °C in agreement with the high hairpin population. Generalized order parameters on this timescale assume values $S^2 = 0.6$-0.85 at 25 °C and $S^2 \approx 0.4$ at 55 °C in the A-state hairpin. Where significant $\Delta\omega(^{15}N) = \omega_f-\omega_u$ values lead to R_{2ex} terms, µs-motions can be quantified (Table 3, Figure 6). This applies to six A-state hairpin residues, which do not overlap with trimer resonances at 25 °C, and to all hairpin resonances with sufficiently large $\Delta\omega(^{15}N)$ (residues 14-23) at 55 °C (Table 3). The values of $|\Delta\nu(^{15}N)| = |\Delta\omega(^{15}N)|/2\pi$ from thermal chemical shift transitions and values of R_{2ex} obtained from relaxation data are shown for hairpin residues in Table 3. Estimates for p_f and p_u obtained from thermal transition curves are $p_f = 0.85$, $p_u = 0.15$ at 25 °C and $p_f \approx p_u \approx 0.5$ at 55 °C. As the factor $p_f \cdot p_u$ is very robust to errors in the population estimates, especially near the transition midpoint, exchange times can be obtained with high reliability.

At 25 °C average τ_{ex} values of 6.4±1.0 μs are thus obtained for hairpin residues, corresponding to the rate constants $k_f = 0.13\pm0.02 \cdot 10^6$ s^{-1} and $k_u = 0.030\pm0.004 \cdot 10^6$ s^{-1} for hairpin folding and unfolding, respectively. At 55 °C average τ_{ex} values are 1.5±0.7 μs corresponding to $k_f = k_u = 0.47\pm0.24 \cdot 10^6$ s^{-1}. G18 in the turn exhibits particularly fast exchange in the foldon hairpin with $\tau_{ex} = 0.5$ μs at 55 °C, possibly due to its conformational freedom in dihedral angle space. Very similar exchange times of 6.6 ± 0.5 μs at 25 °C are obtained from $^1H^N$ relaxation rates for a smaller unlabeled hairpin construct of residues 12-23 only (data not shown). These estimates for the unlabeled dodecapeptide were obtained with a spin-echo 1D pulse sequence for residues V14, K16 and V21 due their large $^1H^N$ resonance frequency changes upon hydrogen bond breakage (Figure 2B) at pH 2.

R_{2ex} contributions to relaxation can generally arise from folding, local dynamics or aggregation. For short stretches like the A-state hairpin, local internal motions do not occur for several residues in a concerted manner unless linked to the total folding-unfolding of the hairpin [47]. As the foldon hairpin does not aggregate but specifically trimerizes in slow chemical exchange and as the R_{2ex} value is obtained for all residues in the hairpin, the observed exchange reflects the global folding of the hairpin. This is consistent with folding times of few μs, which have been obtained with optical spectroscopy for hydrophobic cluster formation in GB1 [18] and for engineered hairpins like trpzip4 and peptide I [48] upon laser-induced temperature jumps. Accordingly, ^{13}C NMR relaxation analysis has shown conformational exchange on the microsecond time scale for model hairpins [47,49].

Residual dipolar couplings in thermal unfolding

Experimental and computational analysis of β-hairpin model systems have pointed to four major sources of β-hairpin stability: local turn propensities, β-strand propensities, favorable sidechain contacts between the strands and hydrogen bonds between the strands [50]. While local structures in partially folded and unfolded states have been observed with NMR spectroscopic methods, a more complete picture requires information on long-range structure and can be obtained by the measurement

of residual dipolar couplings. We measure $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ RDCs between 40 °C and 80 °C in inert polyacrylamide gels to gain information on the structural order of peptide groups and sidechains, respectively. At temperatures lower than 40 °C the monomer was insufficiently populated to measure RDCs with adequate reliability. NMR spectra of the foldon A-state in polyacrylamide gels are identical to those in solution, indicating that the gels cause no major structural changes.

The sequence- and temperature-dependence of $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ RDCs during thermal A-state unfolding are depicted in Figure 7. Large RDCs are mainly observed for the well-folded hairpin region in the A-state, whereas regions outside the hairpin adopt smaller but non-zero RDCs. This resembles maximal structural order in the foldon A-state in the hairpin and smaller, but detectable smaller order N- and C-terminal of the hairpin. Alignment tensors obtained from separate fits of $^1D_{C\alpha H\alpha}$ and $^1D_{NH}$ to the hairpin structure show that $^1D_{C\alpha H\alpha}$ are ~15 % smaller than expected from the $^1D_{NH}$ at 40 °C. This indicates that the local order of NH^N vectors is bigger than the local order of $C^\alpha H^\alpha$ vectors, in accordance with highly populated hydrogen bonds in the folded β–hairpin, but rather loose hydrophobic contacts at 40 °C. $^1D_{NH}$ RDCs for V14/L22 and V21/R15 adopt a very similar profile during thermal unfolding, further indicating that the strong hydrogen bonds between V14 an V21 (see above) impose a time average in-line geometry on the donor and acceptor peptide group (Figure 7).

RDCs assume non-zero values in unfolded stretches without tertiary contacts N- and C-terminal of the hairpin. Furthermore RDCs converge to non-zero values in the hairpin upon thermal unfolding. Non-zero RDCs in sterically aligned unfolded states are observable even for random coils and result from non-spherical distributions of individual NH^N and $C^\alpha H^\alpha$ bond vectors in any aligned polypeptide chain [25]. In random chains, however, the non-zero RDC is only weakly influenced by the residue position, resulting in a smooth RDC value along the peptide chain without dispersion of the RDCs [25]. As RDCs in the complete sequence of foldon are well dispersed, distinct orientations are observed and the chain exhibits deviations from random coil behavior, both in the absence of tertiary contacts near the chain termini (Figure 7A) and upon thermal unfolding (Figure 7A,B). Largest RDCs outside the foldon hairpin are observed for the residues 4-8 near prolines P4 and P7, as can be understood from

the limited conformational freedom of prolines. In this proline rich stretch, a small relative decrease of RDCs (~0.3) is observed between 40 °C to 80 °C due to stable local dihedral angle preferences, whereas RDCs of all β-strand residues show an average relative decrease by ~0.8 over the same temperature range due to the large loss of structural ordering upon hairpin unfolding.

Conformational preferences are also evident for large hydrophobic residues upon thermal unfolding of the β-hairpin (Figure 7B). In particular, high temperature $^{1}D_{C\alpha H\alpha}$ dipolar couplings indicate a preference for extended conformations similar to the orientation in the hairpin for residues Y13 and W20 (Figure 7). Comparable $^{1}D_{C\alpha H\alpha}$ RDCs are observed for L22, which forms a hydrophobic cluster with Y13 and W20 in the hairpin. Large residues close to the N- and C-terminus, specifically Y2 and F26, exhibit similar stable nonrandom structures (Figure 7A). These effects on large sidechains presumably arise from local sidechain-backbone interactions, which force large sidechains into extended conformations and restrict the conformational space accessible to polypeptide chains [51]. Residual structures have been reported in single extracted hairpin strands [36] and short peptides [23] and most probably arise from local steric sidechain-backbone clashes, resulting in local dihedral angle preferences.

Thermal transitions of $^{1}D_{C\alpha H\alpha}$ dipolar couplings in the hairpin strands fitted to two-state behavior (Figure 7B) yield average melting temperatures of 330.7 ± 3.9 K in very good agreement with thermal transition curves obtained for chemical shifts (Table 2). Residues near the hairpin ends (e.g. Q11, A12) get disordered at clearly lower temperatures than the average (Figure 7B). Residues at the type I' β–turn exhibit thermal $^{1}D_{NH}$ and $^{1}D_{C\alpha H\alpha}$ transitions at increased temperatures (E19) or are restricted to nonrandom values for the whole temperature range (K16) (Figure 7B). This indicates that the turn adopts a non-random and non-extended, but rather native-like time average conformation, possibly due to the high propensity of the turn sequence for type I' β–turn conformations. In addition to such local intrinsic preferences, the turn may be stabilized in a bent conformation upon the melting of hydrogen bonds via temporary hydrophobic sidechain contacts, resulting from the stronger hydrophobic force at higher temperatures. This is in agreement with MD simulations, which show that transition temperatures of the hydrophobic collapse in

molecular dynamics simulations are higher than the transition temperatures for the formation of backbone secondary structures in hairpins [41]. The interplay between turn propensities and a nearby hydrophobic cluster in hairpins has been pointed out, where a smaller separation of the hydrophobic cluster and the turn is entropically more favorable [50,52]. This might account for the high structural stability in the short foldon hairpin with tight type I' β–turn and nearby hydrophobic cluster.

Conclusion

The foldon domain has been proposed to have an evolutionarily optimized folding and association pathway [53], as its only known biological function is to promote the correct folding of fibritin. Previous kinetic investigations have quantified the fast and efficient association reactions, starting from a monomeric intermediate with partially buried hydrophobic sidechains [9]. We analyze the structure, folding kinetics and stability of this partially folded monomeric state trapped at pH 2.

Destabilizing a strong inter-subunit salt bridge at low pH destabilizes the foldon trimer and populates a monomeric partially folded structure, thus allowing a detailed equilibrium NMR study. The monomeric A-state exhibits the same fluorescence maximum as the folding intermediate at pH 7 [9]. We find that the foldon A-state contains a β-hairpin with intact and stable hydrogen bonds and is highly populated in aqueous solution. It exhibits a strong twist, presumably avoiding competing aggregation side reactions into extended sheets. The intact hairpin is likely to provide suitable surfaces for a fast and specific association of foldon under physiological conditions. Kinetic data have shown that this trimerization occurs in two sequential bimolecular steps via association reactions with rates similar to those of the fastest protein dimerizations [9].

Formation of the foldon intermediate occurs within the stopped-flow dead time ($\tau_{folding}$<1 ms). Chemical exchange in the foldon hairpin at equilibrium was quantified by ^{15}N relaxation data and occurs in average times of $\tau_{ex}\approx$6.4 μs at 25 °C and $\tau_{ex}\approx$1.5 μs at 55 °C, comparable to folding times obtained by kinetic experiments for other hairpins [18].

The foldon A-state hairpin is highly populated in aqueous solution at room temperature. Thermal transitions of A-state β-hairpin unfolding monitored with chemical shifts and residual dipolar couplings both yield average transition temperatures around 330 K. Chemical shifts and dipolar couplings indicate increased local structural stability at the turn, as has been proposed for various hairpins from thermal $^{1}H^{\alpha}$ chemical shift transitions [37] and theoretical models [41]. The stability most likely arises from an interplay of transient hydrophobic cross-strand interactions and local dihedral angle propensities [50,52]. Residual dipolar couplings additionally show such local conformational preferences for unfolded segments without tertiary contacts and for hairpin residues at temperatures of 80 °C. These non-random conformations have been predicted by MD and DFT calculations and are explained by steric clashes of sidechains with their local peptide backbone [51,54,55].

METHODS

Purification

U–^{13}C, U–^{15}N-labeled foldon was expressed as a thioredoxin fusion protein in *E.coli* BL21(DE3) from a plasmid based on pET-32a (Novagen) [7]. The fusion protein carries an N-terminal 6xHis-tag and a thrombin cleavage site between thioredoxin and the C-terminal foldon. The non-native thrombin cleavage site residues $G^{-1}S^{0}$ are highly flexible and disordered at both pH 7 and pH 2. The complete sequence of the foldon construct is GS^{0} GYIPEAPRDG^{10} QAYVERKDG^{20} VLLSTFL^{27}.

Foldon was purified via affinity chromatography on a nickel-NTA column (Novagen) before and after thrombin cleavage and was taken up in 5 mM potassium phosphate buffer of pH 7. Protein concentrations were determined using the method of Gill & von Hippel [56] with an extinction coefficient of 2.56 $cm^{-1}mg^{-1}ml$.

Heteronuclear NMR

NMR samples were prepared in $^1H_2O/^2H_2O$ 19:1 (v/v) and adjusted to the indicated pH values with a Russel electrode by addition of 1 M HCl. All spectra were recorded on a BRUKER DRX 600 spectrometer. Weak alignment of 0.3 mM U–^{13}C, U–^{15}N foldon in acrylamide gels of 7% (w/v) was achieved by stretching gels in NEW-ERA sample tubes to an aspect ratio of 2.9 as described [57]. $^1D_{NH}$ were measured with the IPAP experiment [58], $^1D_{C\alpha H\alpha}$ were obtained via a gradient-selected constant time HSQC without ^{13}C decoupling during acquisition. Errors are estimated as <0.3 Hz for $^1D_{NH}$ and <0.8 Hz for $^1D_{C\alpha H\alpha}$ dipolar couplings from reproducing the measurements. Temperature calibration was performed as described on a standard polyethylenglycole sample [59]. ^{1}H, ^{15}N, and ^{13}C chemical shifts were referenced relative to the ^{2}H lock resonance of water for all temperatures. Shifts of sidechain carboxylate resonances were measured at 25°C with a gradient selected H(CA)CO experiment optimized for sidechain methlyene groups. ^{15}N relaxation data were recorded as described [60] on a sample of 0.3 mM U–^{13}C, U–^{15}N foldon and fitted to model-free spectral density functions with the program TENSOR [61]. R_2 data were obtained using Carr-Purcell-Meiboom-Gill (CPMG) spin-echo experiments with delays of 1 ms between the 180 ° pulses.

Data analysis and structure calculation

All data were processed with NMRPipe [62] and analyzed with PIPP [63]. Fits of thermal transitions according to the Gibbs-Helmholtz formula [38] were performed by Monte Carlo fitting followed by a robust minimization of χ^2 with ProFit (QuantumSoft, Switzerland). pK_a values were obtained from ^{13}C chemical shift measurements at 25 °C by fitting the Henderson-Hasselbalch equation to carboxyl or sidechain chemical shifts in ProFit. Relaxation data [60,64] were analyzed by fitting a Gaussian curve to spectral peaks with the nLINLS tool contained in NMRPipe and by Monte Carlo fits to exponential decays in MATLAB.

Structure calculations were performed using the program CNS [65] using a standard simulated annealing protocol. Structure representations were generated with MOLMOL [66].

ACKNOWLEDGEMENTS

We thank Klara Rathgeb-Szabo for protein purification. This work was supported by SNF grant 31-43'091.95 to S.G.

Tables

Table 1. Statistics of the foldon A-state NMR structure[1)]

RMSDs from experimental distance constrains[2)] [Å]	
all (175)	0.068 ± 0.002
RMSDs from dihedral constraints [°] (28)[3)]	1.0± 0.2
RMSDs from $^{3}J_{HNHA}$ coupling constraints [°] (20)	1.3± 0.1
Deviation from the idealized covalent geometry	
bonds [Å]	0.0060 ± 0.0001
angles [°]	0.69 ± 0.02
impropers[5)] [°]	0.50 ± 0.03
Coordinate precision [Å][6)]	
backbone non-hydrogen atoms	0.29
all non-hydrogen atoms	0.75
Percentage of non-gly, non-pro residues in Ramachandran regions[7)]	
core	92.2
allowed	7.8
generous	0.0
disallowed	0.0

[1)] The statistics were obtained from a subset of the 10 best energy structures out of 100 calculated following the standard CNS [65] simulated annealing protocol with dipolar restraints incorporated. The number of the various constraints is given in parentheses.
[2)] Distance constraints comprise 76 intraresidual NOEs, 61 sequential NOEs ($|i-j| = 1$), 15 short-range NOEs ($1<|i-j|\leq 5$) and 23 long-range NOEs ($|i-j|>5$).
[3)] Consisting of 15 ^{1}H-^{15}N, 14 $^{1}H^{\alpha}$-$^{13}C^{\alpha}$ dipolar one-bond couplings from residue 8-22.
[4)] The dihedral angle constraints comprise 16 ϕ and 12 ψ angles obtained from TALOS [44].
[5)] The improper torsion restraints serve to maintain planarity and chirality.
[6)] The coordinate precision is defined as the average rms difference between the individual simulated annealing structures and the mean coordinates. Values are reported for residues 10-23, i.e. for core residues, which show maximal $\{1H\}$–^{15}N heteronuclear NOE values in the foldon A-state.
[7)] These values are calculated with the program PROCHECK-NMR [67] for residues 10-23.

Table 2. Transition temperatures T_m and apparent folding enthalpies ΔH^0 from thermal unfolding of the foldon hairpin observed by $^{13}C^{\alpha}$ and $^{1}H^{\alpha}$ shifts[1].

	T_m/K	ΔH^0/kJ/mol	ΔS^0/J/(mol·K)
Turn[2]	342.6 ± 0.8	-56.5 ± 7.6	-165.4 ± 22.0
Strands[3]	332.6 ± 2.7	-50.7 ± 5.2	-152.4 ± 16.4

[1] Values are obtained by fits of chemical shifts to the Gibbs-Helmholtz equation. Average values and standard deviations are given.
[2] Fitted resonances comprise $^{13}C^{\alpha}17$, $^{1}H^{\alpha}17$ and the glycine methylen proton dispersion $\Delta^{1}H^{\alpha}$ 18, whereas the secondary shift of $^{13}C^{\alpha}18$ is too small for reliable fits.
[3] Fitted resonances comprise $^{13}C^{\alpha}$ and $^{1}H^{\alpha}$ resonances of residues 13-16 and 19-22. The secondary shift was too small to yield reliable fits for $^{13}C^{\alpha}$ 13 and $^{1}H^{\alpha}$ 15,16.

Table 3. Folding times obtained for foldon hairpin residues.

	25 °C				55 °C			
res	**$\lvert\nu_f-\nu_u\rvert$/Hz[1)]**	**R_{2ex}/Hz[2)]**	**ΔR_{2ex}/Hz[3)]**	**τ_{ex}/μs[4)]**	**$\lvert\nu_f-\nu_u\rvert$/Hz[1)]**	**R_{2ex}/Hz[2)]**	**ΔR_{2ex}/Hz[3)]**	**τ_{ex}/μs[4)]**
13	21.1	-[5)]	-	-	36.1	-[5)]	-	-
14	185.7	(not fit)[6)]	-	-	170.7	0.72	0.10	2.5
15	103.7	0.33	0.10	6.1	118.7	0.24	0.04	1.7
16	292.6	(overlap)	-	-	307.6	0.98	0.04	1.1
17	361.3	4.79	0.04	7.3	376.3	1.41	0.02	1.0
18	364.7	4.40	0.03	6.6	349.7	0.65	0.12	0.5
19	153.2	0.64	0.15	5.4	138.2	0.19	0.06	1.0
20	87.7	-[5)]	-	-	102.7	0.23	0.03	2.5
21	214.4	1.56	0.26	6.7	229.4	0.29	0.02	0.6
22	87.9	(overlap)	-	-	102.9	0.16	0.03	1.5
23	269.7	1.99	0.10	5.4	284.7	0.60	0.10	0.8

[1)]Values are derived from two-state transitions of nitrogen shifts. $\lvert\nu_f-\nu_u\rvert$ is the ^{15}N shift difference between the folded and the unfolded state in Hz. Values are temperature-corrected assuming a temperature independent shift of the folded state and a linear temperature factor –0.5 Hz/°C for the unfolded state. These assumptions derive from nearly unchanged native trimer chemical shifts between 25 ° C and 55 °C and from the behavior of unfolded residues near the N- and C-terminus of the monomer. The corrections have a very weak influence on τ_{ex}.
[2)] Exchange rates R_{2ex} obtained from model free spectral density fitting of T_1, T_2 and $\{^{1}H\}$–^{15}N NOE data (Figure 6) at pH 2.
[3)] Errors of R_{2ex} from the spectral density fitting.
[4)] The exchange time is obtained as $\tau_{ex} = \dfrac{R_{2ex}}{4\pi^2 p_f p_u (\nu_f - \nu_u)^2}$.
[5)] Analysis of relaxation data did not result in R_{2ex} terms for nuclei with small $\lvert\nu_f-\nu_u\rvert$.
[6)] The model-free spectral density fitting of V14 relaxation data at 25 °C did not converge.

Figure captions

Figure 1: The structure of the T4 fibritin foldon trimer. (A) Hydrogen bond topology of the trimeric foldon β-hairpin propeller. (B) Backbone worm representation of the mean solution structure. (C) Inter-subunit contacts within the foldon trimer. The three monomers are shown in backbone worm representation as orange, blue and beige. Residues that make inter-subunit contacts are indicated in spacefill with CPK colors. For clarity the R15···E5 salt bridge is only shown between two monomers. (D) Secondary structure topology of a foldon monomer. The H-bonds are depicted by a dotted line between the amide proton donor (filled hexagons) and the oxygen acceptor (open hexagons).

Figure 2: (A,B) Concentration-dependent dissociation of the foldon trimer at pH 2 and 25 °C in solutions with 5 mM phosphate and 20 mM chloride at 40 μM (A) and 160 μM (B) peptide concentration. Trimer and monomer coexist in equilibrium in slow exchange. Amide resonance assignments for both species are given in (B) with trimer resonances indicated with **T**. (C) Trimer and monomer peak intensities as a function of pH at 25 °C and 0.25 mM total peptide concentration. (D) Titration of sidechain carboxylate groups in trimeric foldon at 25 °C. Solid lines are fits of the Henderson-Hasselbalch equation to chemical shifts. pK_a values are: E5: 1.70, D9: 2.55, D17: 3.23, E19: 4.37. (E) Co-titration of E5 and R15 sidechains.

Figure 3: Analysis of $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ RDCs in the foldon monomer. (A) $^1D_{NH}$, (B) $^1D_{C\alpha H\alpha}$ RDCs. Solid lines represent measured couplings, dotted lines RDCs calculated from the average trimer structure. Residual orientation was obtained within mechanically strained acrylamide gels for 0.4 mM foldon monomer, 5 mM phosphate, 20 mM chloride at 45 °C and pH 2. (C) Q-factors of the combined $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ RDCs of segments of 10 consecutive amino acids moved through the foldon sequence as a function of the average residue number (see text).

Figure 4: Solution structure of the A-state foldon monomer. (A) Backbone worm representation of the 10 lowest energy structures of the foldon monomer. (B) The

hairpin part of the 10 lowest energy structures of the foldon monomer in the A-state in stick representation. Sidechains of residues Y13, W20 and L22, which form a hydrophobic cluster, are shown in red. For comparison, the hairpin part of the 10 lowest energy structures of foldon within the native trimer (1RFO) is shown in (C).

Figure 5: Residue specific thermal transitions in the hairpin as monitored by backbone chemical shifts of $^{13}C^{\alpha}$ (A), $^{1}H^{\alpha}$ (B), $^{1}H^{N}$ (C) and ^{15}N (D) at pH 2. Solid lines are fits to the Gibbs-Helmholtz equation assuming two-state exchange. The derived thermodynamic parameters of (A,B) are given in Table 2, $\Delta\omega(^{15}N) = (\omega_f - \omega_u)$ for the quantification of exchange times are given in Table 3.

Figure 6: (A) R_1, R_2(CPMG) and $\{^{1}H\}-^{15}N$ data of the foldon A state at 25 °C (red) and 55 °C (black). (B) Resulting chemical exchange rates R_{2ex} as obtained by Lipari-Szabo analysis.

Figure 7: (A) $^{1}D_{NH}$ and $^{1}D_{C\alpha H\alpha}$ RDCs in the foldon A-state between 40 °C and 80 °C. Scales are adjusted to $^{1}D_{C\alpha H\alpha}^{\max,solid-state} = -1.98 \cdot {}^{1}D_{NH}^{\max,solid-state}$. Random chain behavior is approximated as an average over RDCs at 80 °C (solid black line). (B) Temperature course of $^{1}D_{NH}$ and $^{1}D_{C\alpha H\alpha}$ fitted to two-state models. Thermal transitions indicative of local order upon thermal unfolding are shown in red for residues near the turn and in blue for large hydrophobic sidechains.

References:

1. Coombs DH, Eiserling FA: Studies on Structure, Protein-Composition and Assembly of Neck of Bacteriophage-T4. *Journal of Molecular Biology* 1977, 116:375-405.
2. Terzaghi BE, Terzaghi E, Coombs D: Role of the Collar-Whisker Complex in Bacteriophage T4d Tail Fiber Attachment. *Journal of Molecular Biology* 1979, 127:1-14.
3. Conley MP, Wood WB: Bacteriophage-T4 Whiskers - Rudimentary Environment-Sensing Device. *Proceedings of the National Academy of Sciences of the United States of America* 1975, 72:3701-3705.
4. Tao YZ, Strelkov SV, Mesyanzhinov VV, Rossmann MG: Structure of bacteriophage T4 fibritin: A segmented coiled coil and the role of the C-terminal domain. *Structure* 1997, 5:789-798.
5. Strelkov SV, Tao Y, Shneider MM, Mesyanzhinov VV, Rossmann MG: Structure of bacteriophage T4 fibritin M: a troublesome packing arrangement. *Acta Crystallographica Section D-Biological Crystallography* 1998, 54:805-816.
6. Letarov AV, Londer YY, Boudko SP, Mesyanzhinov VV: The carboxy-terminal domain initiates trimerization of bacteriophage T4 fibritin. *Biochemistry-Moscow* 1999, 64:817-823.
7. Frank S, Kammerer RA, Mechling D, Schulthess T, Landwehr R, Bann J, Guo Y, Lustig A, Bächinger HP, Engel J: Stabilization of short collagen-like triple helices by protein engineering. *Journal of Molecular Biology* 2001, 308:1081-1089.
8. Yang XZ, Lee J, Mahony EM, Kwong PD, Wyatt R, Sodroski J: Highly stable trimers formed by human immunodeficiency virus type 1 envelope glycoproteins fused with the trimeric motif of T4 bacteriophage fibritin. *Journal of Virology* 2002, 76:4634-4642.
9. Güthe S, Kapinos L, Möglich A, Meier S, Grzesiek S, Kiefhaber T: Very fast folding and association of a trimerization domain from bacteriophage T4 fibritin. *Journal of Molecular Biology* 2004, 337:905-915.
10. Onuchic JN, Wolynes PG: Theory of protein folding. *Current Opinion in Structural Biology* 2004, 14:70-75.
11. Booth DR, Sunde M, Bellotti V, Robinson CV, Hutchinson WL, Fraser PE, Hawkins PN, Dobson CM, Radford SE, Blake CC, et al.: Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature* 1997, 385:787-793.
12. Pan KM, Baldwin M, Nguyen J, Gasset M, Serban A, Groth D, Mehlhorn I, Huang Z, Fletterick RJ, Cohen FE, et al.: Conversion of alpha-helices into beta-sheets features in the formation of the scrapie prion proteins. *Proceedings of the National Academy of Sciences of the United States of America* 1993, 90:10962-10966.
13. Zimm BH, Doty P, Iso K: Determination of the Parameters for Helix Formation in Poly- Gamma-Benzyl-L-Glutamate. *Proceedings of the National Academy of Sciences of the United States of America* 1959, 45:1601-1604.
14. Zimm BH, Bragg JK: Theory of the Phase Transition between Helix and Random Coil in Polypeptide Chains. *Journal of Chemical Physics* 1959, 31:526-535.

15. Lifson S: Theory of Helix-Coil Transition in Polypeptides. *Journal of Chemical Physics* 1961, 34:1963-&.
16. Blanco FJ, Rivas G, Serrano L: A short linear peptide that folds into a native stable beta-hairpin in aqueous solution. *Nature Structural Biology* 1994, 1:584-590.
17. Gellman SH: Minimal model systems for beta sheet secondary structure in proteins. *Curr Opin Chem Biol* 1998, 2:717-725.
18. Munoz V, Thompson PA, Hofrichter J, Eaton WA: Folding dynamics and mechanism of beta-hairpin formation. *Nature* 1997, 390:196-199.
19. Lacroix E, Kortemme T, Lopez de la Paz M, Serrano L: The design of linear peptides that fold as monomeric beta-sheet structures. *Curr Opin Struct Biol* 1999, 9:487-493.
20. Tjandra N, Bax A: Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* 1997, 278:1111-1114.
21. Shortle D, Ackerman MS: Persistence of native-like topology in a denatured protein in 8 M urea. *Science* 2001, 293:487-489.
22. Ackerman MS, Shortle D: Robustness of the long-range structure in denatured staphylococcal nuclease to changes in amino acid sequence. *Biochemistry* 2002, 41:13791-13797.
23. Ohnishi S, Shortle D: Observation of residual dipolar couplings in short peptides. *Proteins-Structure Function and Genetics* 2003, 50:546-551.
24. Alexandrescu AT, Kammerer RA: Structure and disorder in the ribonuclease S-peptide probed by NMR residual dipolar couplings. *Protein Sci* 2003, 12:2132-2140.
25. Louhivuori M, Paakkonen K, Fredriksson K, Permi P, Lounila J, Annila A: On the origin of residual dipolar couplings from denatured proteins. *Journal of the American Chemical Society* 2003, 125:15647-15650.
26. Tycko R, Blanco FJ, Ishii Y: Alignment of biopolymers in strained gels: A new way to create detectable dipole-dipole couplings in high-resolution biomolecular NMR. *Journal of the American Chemical Society* 2000, 122:9340-9341.
27. Sass HJ, Musco G, Stahl SJ, Wingfield PT, Grzesiek S: Solution NMR of proteins within polyacrylamide gels: Diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes. *Journal of Biomolecular NMR* 2000, 18:303-309.
28. Wang L, O'Connell T, Tropsha A, Hermans J: Molecular simulations of beta-sheet twisting. *Journal of Molecular Biology* 1996, 262:283-293.
29. Wagner G, Pardi A, Wüthrich K: Hydrogen-Bond Length and H-1-NMR Chemical-Shifts in Proteins. *Journal of the American Chemical Society* 1983, 105:5948-5949.
30. Cordier F, Grzesiek S: Direct observation of hydrogen bonds in proteins by interresidue (3h)J(NC ') scalar couplings. *Journal of the American Chemical Society* 1999, 121:1601-1602.
31. Sass J, Cordier F, Hoffmann A, Cousin A, Omichinski JG, Lowen H, Grzesiek S: Purple membrane induced alignment of biological macromolecules in the magnetic field. *Journal of the American Chemical Society* 1999, 121:2047-2055.
32. Cornilescu G, Marquardt JL, Ottiger M, Bax A: Validation of Protein Structure from Anisotropic Carbonyl Chemical Shifts in a Dilute Liquid Crystalline Phase. *Journal of the American Chemical Society* 1998, 120:6836-6837.

33. Walkenhorst WF, Edwards JA, Markley JL, Roder H: Early formation of a beta hairpin during folding of staphylococcal nuclease H124L as detected by pulsed hydrogen exchange. *Protein Science* 2002, 11:82-91.
34. McCallister EL, Alm E, Baker D: Critical role of beta-hairpin formation in protein G folding. *Nature Structural Biology* 2000, 7:669-673.
35. Ramirez-Alvarado M, Blanco FJ, Serrano L: De novo design and structural analysis of a model beta-hairpin peptide system. *Nature Structural Biology* 1996, 3:604-612.
36. Maynard AJ, Sharman GJ, Searle MS: Origin of beta-hairpin stability in solution: Structural and thermodynamic analysis of the folding of model peptide supports hydrophobic stabilization in water. *Journal of the American Chemical Society* 1998, 120:1996-2007.
37. Santiveri CM, Santoro J, Rico M, Jimenez MA: Thermodynamic analysis of beta-hairpin-forming peptides from the thermal dependence of H-1 NMR chemical shifts. *Journal of the American Chemical Society* 2002, 124:14903-14909.
38. Atkins PW: *Physical chemistry* edn 6th. Oxford, England: Oxford University Press; 1998.
39. Searle MS, Griffiths-Jones SR, Skinner-Smith H: Energetics of weak interactions in a beta-hairpin peptide: Electrostatic and hydrophobic contributions to stability from lysine salt bridges. *Journal of the American Chemical Society* 1999, 121:11615-11620.
40. Dill KA, Fiebig KM, Chan HS: Cooperativity in protein-folding kinetics. *Proceedings of the National Academy of Sciences of the United States of America* 1993, 90:1942-1946.
41. Klimov DK, Thirumalai D: Mechanisms and kinetics of beta-hairpin formation. *Proceedings of the National Academy of Sciences of the United States of America* 2000, 97:2544-2549.
42. Ramirez-Alvarado M, Blanco FJ, Niemann H, Serrano L: Role of beta-turn residues in beta-hairpin formation and stability in designed peptides. *Journal of Molecular Biology* 1997, 273:898-912.
43. Xu XP, Case DA: Probing multiple effects on 15N, 13C alpha, 13C beta, and 13C' chemical shifts in peptides using density functional theory. *Biopolymers* 2002, 65:408-423.
44. Cornilescu G, Delaglio F, Bax A: Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *Journal of Biomolecular NMR* 1999, 13:289-302.
45. Millet O, Loria JP, Kroenke CD, Pons M, Palmer AG: The static magnetic field dependence of chemical exchange linebroadening defines the NMR chemical shift time scale. *Journal of the American Chemical Society* 2000, 122:2867-2877.
46. Lipari G, Szabo A: Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules.2. Analysis of Experimental Results. *Journal of the American Chemical Society* 1982, 104:4559-4570.
47. Nesmelova I, Krushelnitsky A, Idiyatullin D, Blanco F, Ramirez-Alvarado M, Daragan VA, Serrano L, Mayo KH: Conformational exchange on the microsecond time scale in alpha-helix and beta-hairpin peptides measured by 13C NMR transverse relaxation. *Biochemistry* 2001, 40:2844-2853.

48. Xu Y, Oyola R, Gai F: Infrared study of the stability and folding kinetics of a 15-residue beta-hairpin. *Journal of the American Chemical Society* 2003, 125:15388-15394.
49. Friedrichs MS, Stouch TR, Bruccoleri RE, Mueller L, Constantine KL: Structural and Dynamic Properties of a Beta-Hairpin-Forming Linear Peptide.2. C-13 NMR Relaxation Analysis. *Journal of the American Chemical Society* 1995, 117:10855-10864.
50. Espinosa JF, Munoz V, Gellman SH: Interplay between hydrophobic cluster and loop propensity in beta-hairpin formation. *Journal of Molecular Biology* 2001, 306:397-402.
51. Street AG, Mayo SL: Intrinsic beta-sheet propensities result from van der Waals interactions between side chains and the local backbone. *Proceedings of the National Academy of Sciences of the United States of America* 1999, 96:9074-9076.
52. Munoz V, Henry ER, Hofrichter J, Eaton WA: A statistical mechanical model for beta-hairpin kinetics. *Proceedings of the National Academy of Sciences of the United States of America* 1998, 95:5872-5879.
53. Boudko SP, Londer YY, Letarov AV, Sernova NV, Engel J, Mesyanzhinov VV: Domain organization, folding and stability of bacteriophage T4 fibritin, a segmented coiled-coil protein. *European Journal of Biochemistry* 2002, 269:833-841.
54. Rossmeisl J, Kristensen I, Gregersen M, Jacobsen KW, Norskov JK: Beta-sheet preferences from first principles. *Journal of the American Chemical Society* 2003, 125:16383-16386.
55. Yang WY, Larios E, Gruebele M: On the extended beta-conformation propensity of polypeptides at high temperature. *Journal of the American Chemical Society* 2003, 125:16220-16227.
56. Gill SC, von Hippel PH: Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* 1989, 182:319-326.
57. Chou JJ, Gaemers S, Howder B, Louis JM, Bax A: A simple apparatus for generating stretched polyacrylamide gels, yielding uniform alignment of proteins and detergent micelles. *Journal of Biomolecular NMR* 2001, 21:377-382.
58. Ottiger M, Delaglio F, Bax A: Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *Journal of Magnetic Resonance* 1998, 131:373-378.
59. Raiford DS, Fisk CL, Becker ED: Calibration of Methanol and Ethylene-Glycol Nuclear Magnetic-Resonance Thermometers. *Analytical Chemistry* 1979, 51:2050-2051.
60. Kay LE, Nicholson LK, Delaglio F, Bax A, Torchia DA: Pulse Sequences for Removal of the Effects of Cross-Correlation between Dipolar and Chemical-Shift Anisotropy Relaxation Mechanism on the Measurement of Heteronuclear T1 and T2 Values in Proteins. *Journal of Magnetic Resonance* 1992, 97:359-375.
61. Dosset P, Hus JC, Blackledge M, Marion D: Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *Journal of Biomolecular NMR* 2000, 16:23-28.
62. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A: Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes. *Journal of Biomolecular NMR* 1995, 6:277-293.

63. Garrett DS, Gronenborn AM, Clore GM: Automated and Interactive Tools for Assigning 3d and 4d NMR-Spectra of Proteins - Capp, Stapp and Pipp. *Journal of Cellular Biochemistry* 1995:71-71.
64. Kay LE, Torchia DA, Bax A: Backbone Dynamics of Proteins as Studied by N-15 Inverse Detected Heteronuclear NMR-Spectroscopy - Application to Staphylococcal Nuclease. *Biochemistry* 1989, 28:8972-8979.
65. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, et al.: Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallographica Section D-Biological Crystallography* 1998, 54:905-921.
66. Koradi R, Billeter M, Wüthrich K: MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* 1996, 14:51-55, 29-32.
67. Laskowski RA, Rullmann JAC, MacArthur MW, Kaptein R, Thornton JM: AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *Journal of Biomolecular NMR* 1996, 8:477-486.

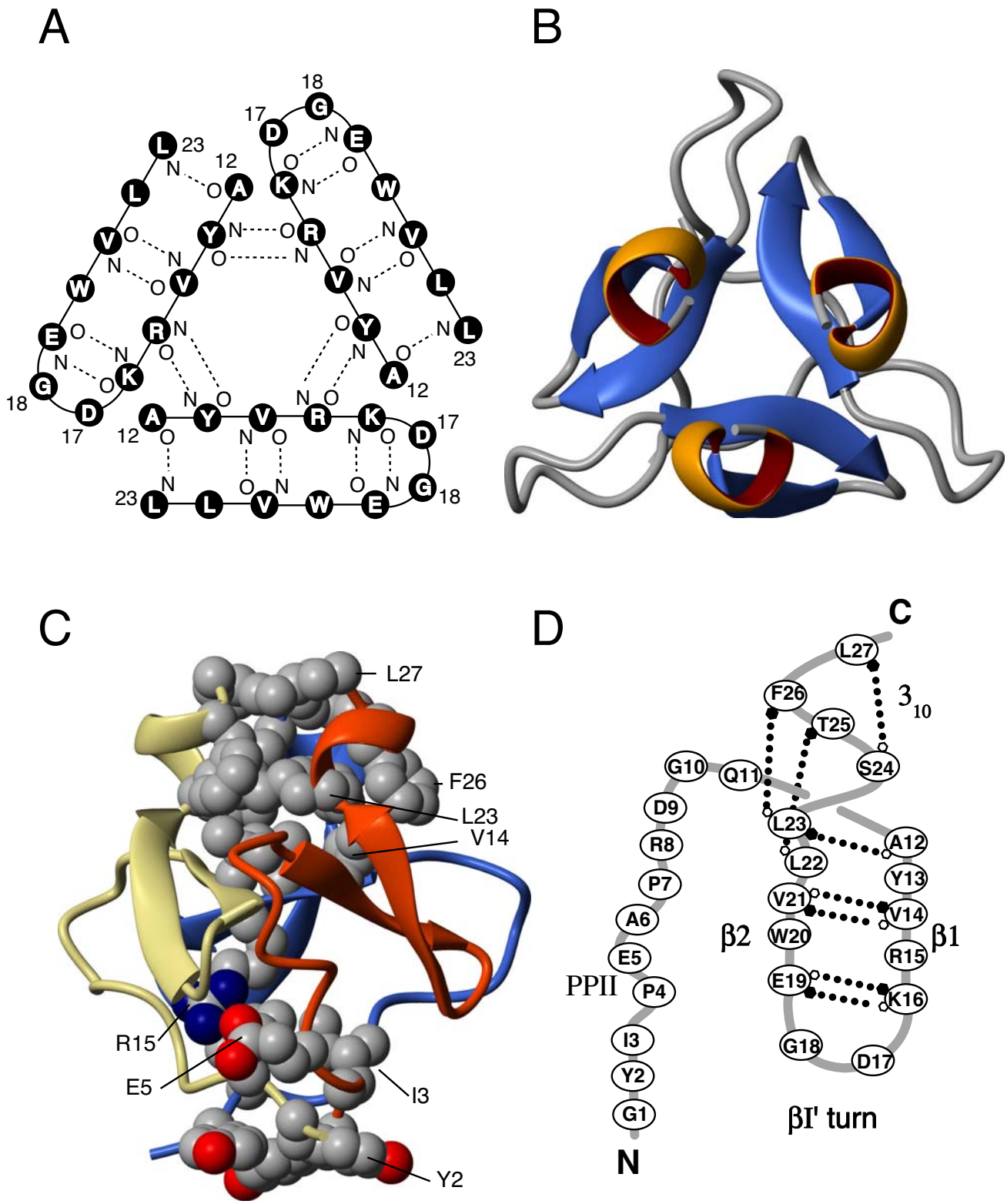


Figure 1 - Meier, Güthe, Kiefhaber and Grzesiek

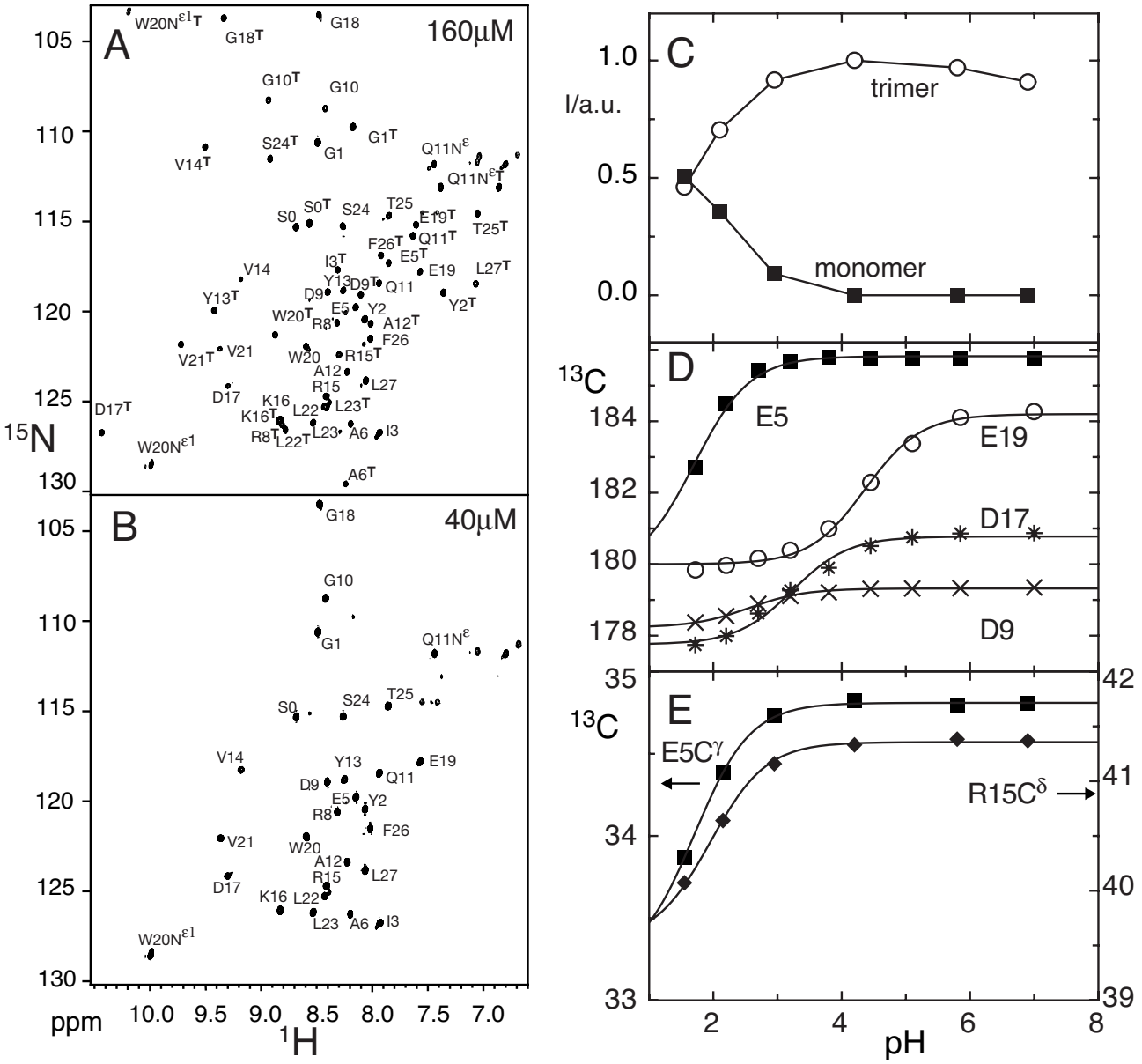


Figure 2 - Meier, Güthe, Kiefhaber and Grzesiek

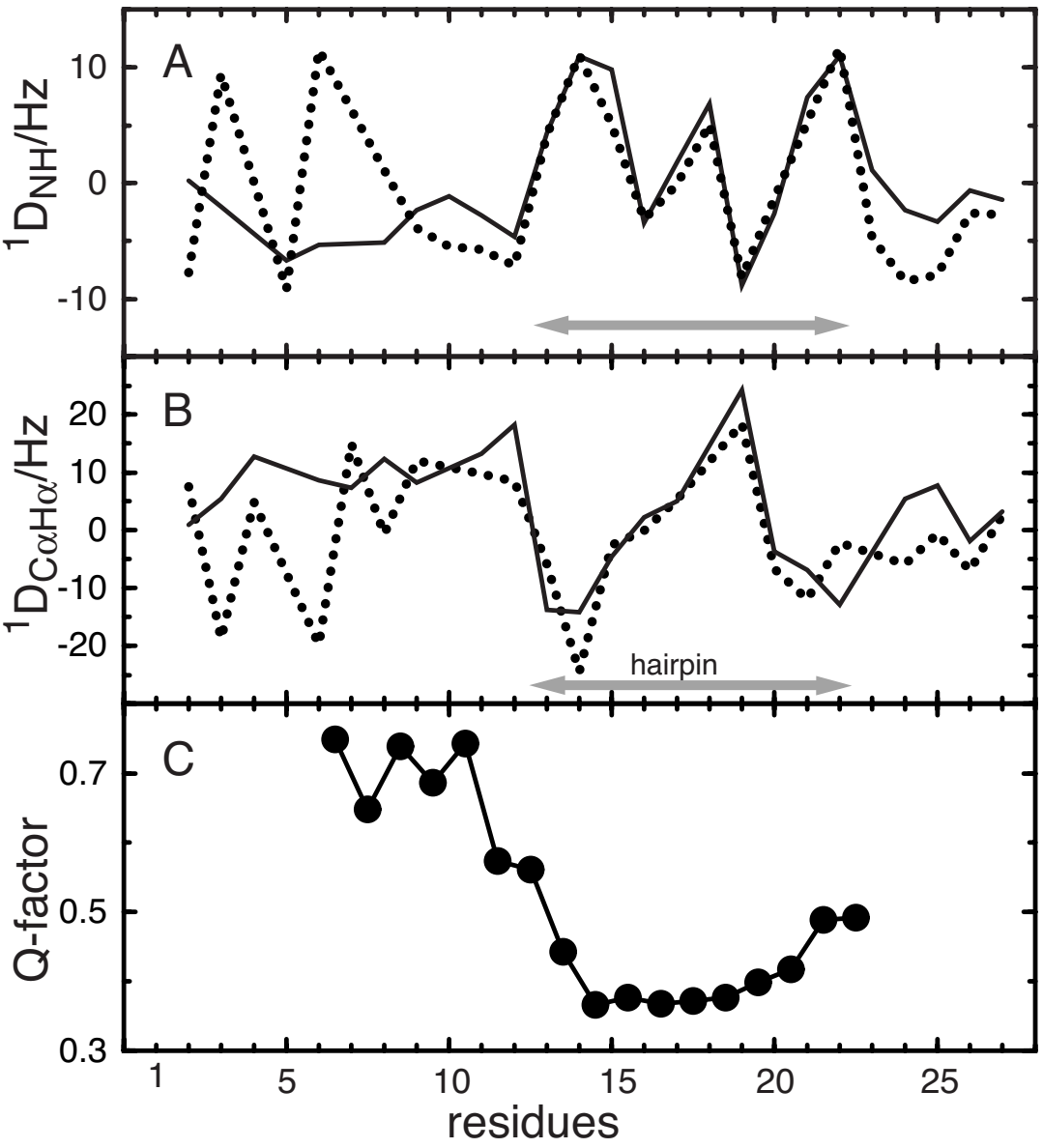


Figure 3 - Meier, Güthe, Kiefhaber and Grzesiek

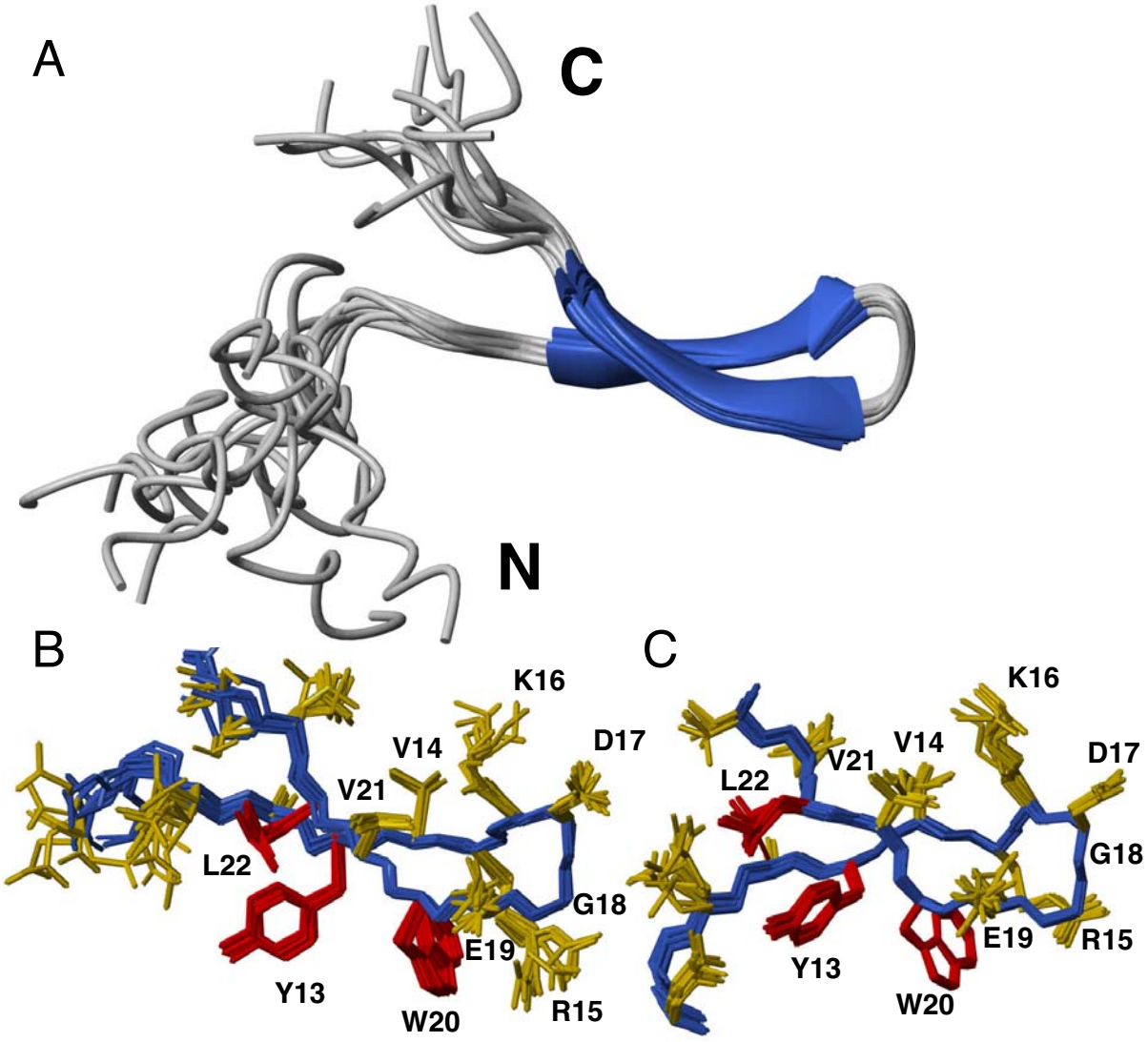


Figure 4 - Meier, Güthe, Kiefhaber and Grzesiek

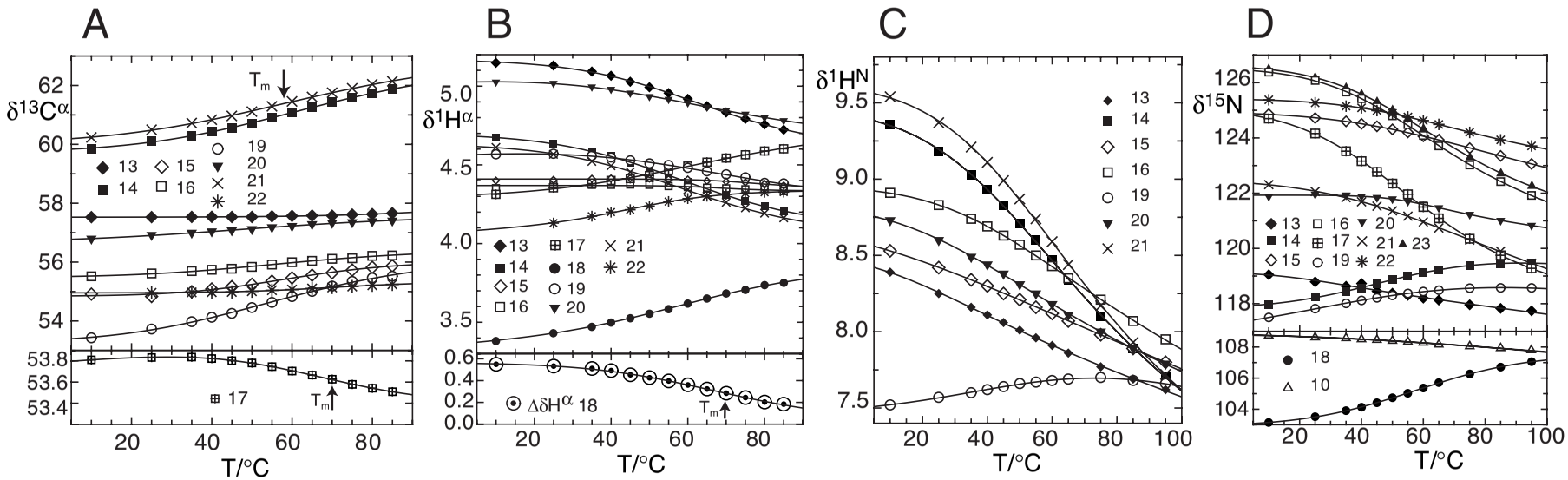


Figure 5 - Meier, Güthe, Kiefhaber and Grzesiek

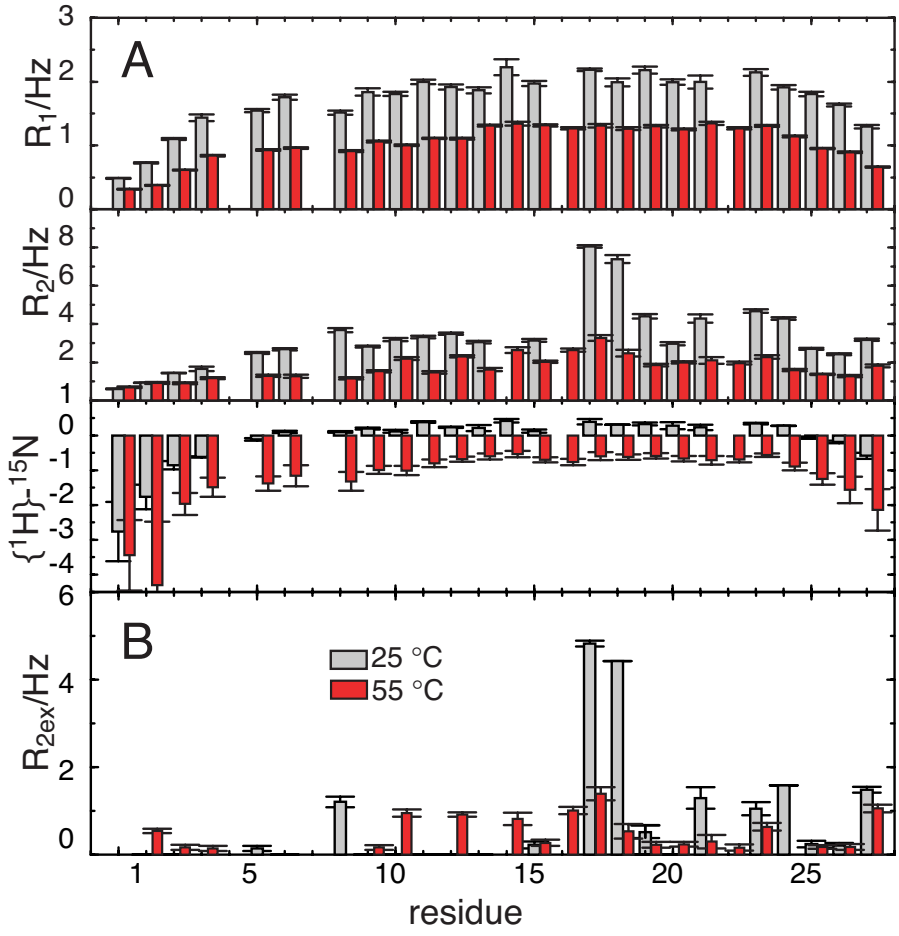


Figure 6 - Meier, Güthe, Kiefhaber and Grzesiek

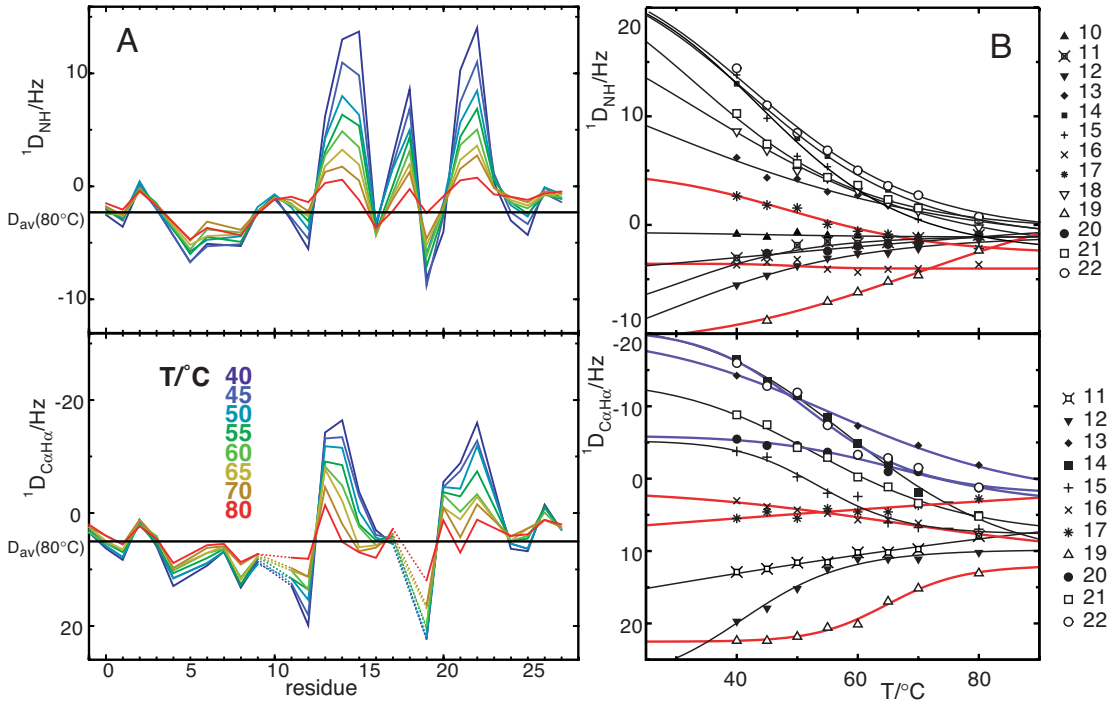


Figure 7 - Meier, Güthe, Kiefhaber and Grzesiek

Chapter 4

Structural biology and folding of a minicollagen cysteine rich domain from *Cnidaria* nematocysts

Cnidaria (greek "knide" = nettle) are aquatic animals represented by sessile polyps like corals and by vagrant medusae like jellyfish. *Cnidaria* are diploblastic, i.e. they contain only two cell layers, the endoderm and the ectoderm, and miss the mesoderm of higher organisms (Hessinger et al., 1988). In addition, they possess structural and signaling proteins, which have homologues in higher organisms (Sun et al., 1997; Hobmayer et al., 2000; Ozbek et al., 2002a). Thus, ancestors of today's *Cnidaria* are also likely evolutionary ancestors of the triploblastic invertebrate and subsequent vertebrate lineages (Hessinger et al., 1988; Benton and Ayala, 2003).

The defining characteristic of *Cnidaria* is the presence of nematocysts for offensive and defensive action (Hessinger et al., 1988). Nematocysts are products of a *post-Golgi* vesicle and are therefore technically neither cells nor organelles, but secretory products. Upon a stimulus these stinging capsules perform a specialized exocytosis process, which is one of the fastest natural processes known and occurs at accelerations of up to 40 000 g (Holstein and Tardent, 1984). In this process, a tube with a multiple length of the nematocyst is expelled to deliver toxins to the target or to capture or entangle prey. This ultra-fast exocytosis process results from an extreme intracellular osmotic pressure of 150 bar, which is built up by the deposition of poly-γ−glutamate and counter cations during maturation of the *post-Golgi* vesicle (Weber, 1990).

The wall stabilizing this high-pressure resistant structure of *Cnidaria* nematocysts is mainly built of minicollagens (Kurz et al., 1991) and the nematocyst outer wall antigen (NOWA) (Engel et al., 2002) (Figure 4–1). Minicollagens are the smallest known collagens to date. Central $(Gly\text{-}X\text{-}Y)_n$ collagen motives occur in repeats of only n = 14-16 (Kurz et al., 1991). Adjacent to the collagen part there is a polyproline helix on both sides, followed by N- and C-terminal domains rich in cysteine and

proline, which are referred to as minicollagen cysteine rich domains (MCRDs). Theses domains are also found as eightfold repeats at the C-terminus of NOWA, which is a modular protein of 774 amino acids (Figure 4–1).

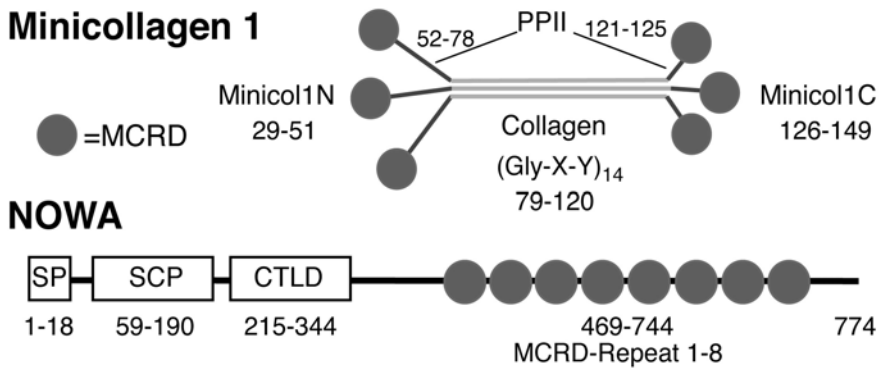


Figure 4–1. Domain organization of mincollagens and NOWA, which are the major constituents of the nematocyst wall. Cysteine rich domains (MCRD) are highlighted.

Minicollagens are expressed as soluble trimers (Ozbek et al., 2002b), whereas the collageneous matrix in nematocyst walls is disulfide-crosslinked (Blanquet and Lenhoff, 1966). The simultaneous appearance of MCRDs in NOWA and minicollagens points to a concerted action of these proteins via their MCRDs in the wall hardening of nematocysts. The wall hardening must be due to a reshuffling of disulfide bonds from intra- to intermolecular disulfide bonds between MCRD domains (Engel et al., 2001), as cysteine residues in minicollagens occur only in the MCRDs.

Structures of the minicollagen cysteine rich domain have not been reported before. We have solved the solution structure of the C-terminal MCRD of minicollagen-1 from hydra. The peptide was synthesized via solid phase N-(9-fluorenyl)methoxycarbonyl (F-moc) strategy by our collaborators Hans Peter Bächinger (Portland, Oregon), Elena Pokidysheva and Jürgen Engel (Biozentrum Basel).

The peptide of only 24 amino acids length contains six cysteines and six prolines. Due to the lack of protease cleavage sites, the disulfide pattern of the six cysteines could not be determined prior to the structure determination. In this thesis, the structure has been determined based on residual dipolar couplings at natural isotope abundance of

^{15}N and ^{13}C in addition to NOEs and dihedral angle restraints obtained from chemical shifts. We obtain a high-resolution structure of MCRD, which reveals the disulfide pattern. The domain is structured and stabilized by the three disulfide bonds Cys2–Cys18, Cys6–Cys14 and Cys10–Cys19, thus forming a disulfide pattern between cysteines 1–5, 2–4 and 3–6 (Figure 4–2). A three-dimensional similarity search of the PDB with the program DALI (Holm and Sander, 1994) did not yield any related structures. This indicates that the MCRD motif represents a new fold and reflects its highly specialized function in the disulfide-mediated crosslinking of the nematocyst wall.

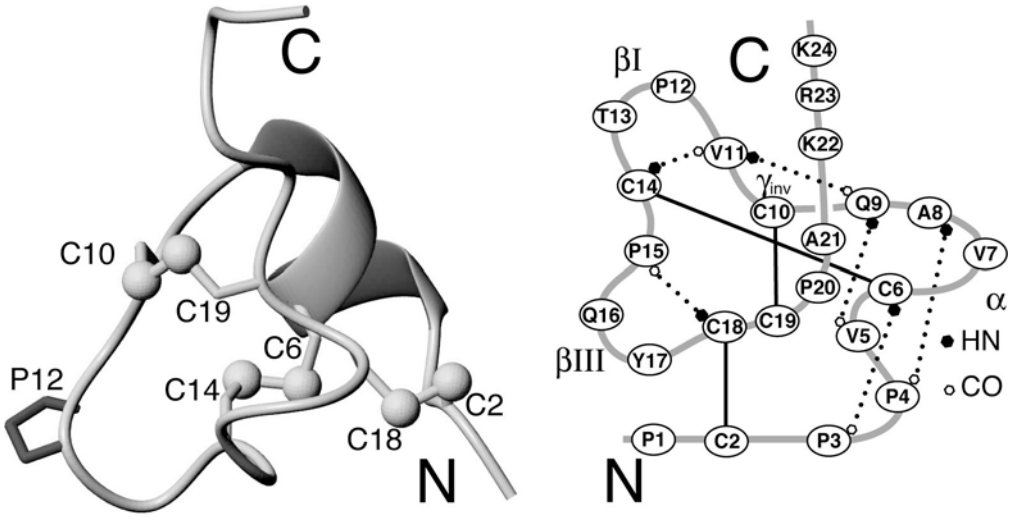


Figure 4–2. Structure and disulfide pattern in MCRD. The backbone has to bend several times to achieve the disulfide pattern and is thus built from several turns, each containing a cysteine residue.

Proteins with short sequences often employ high densities of disulfide bonds to generate a folded topology (Zhang and Snyder, 1989). Among them, MCRD with six cysteines in a structured core of only 20 amino acids has one of the highest cysteine densities known in protein domains. The MCRD domain forms fast and in high yields upon total reduction and subsequent oxidative refolding. As the simultaneous closing of all three native disulfide bonds is highly unlikely, kinetic intermediates are expected to exist during oxidation and folding. Two alternative models for the formation of disulfide bonds in proteins with several cysteines have been proposed: The folded precursor mechanism proposes that native disulfide bond formation is promoted by the appropriate orientation of cysteines in a pre-folded reduced intermediate (Boudko and Engel, 2004). However, lines of evidence often point to a stochastic model, where disulfide bonds form and rearrange quickly, but native

structures are subsequently stabilized during the conformational search (Creighton, 1997).

We show that the complete reduction of disulfide bonds in MCRD indeed leads to the loss not only of the fold, but also of local turn structures and hydrogen bonds, thus underlining the importance of disulfide bonds for the folding of small domains. Indications for the formation of unstructured intermediates during oxidative refolding are obtained by kinetic data from real-time folding observed with one-dimensional NMR spectra. Furthermore, equilibrium unfolding with glutathione redox buffer indicates the presence of a susceptible disulfide bond for intermolecular disulfide reshuffling during MCRD crosslinking to form the matrix of the nematocyst wall.

A bundle of the 10 lowest energy structures out of 100 calculated has been deposited in the PDB under accession number 1SP7. Chemical shift assignments are deposited in the BMRB under accession number 6200.

Detailed results and a thorough discussion of the structural biology and folding of MCRD are given in the following two accepted manuscripts (Meier et al., 2004b; Meier et al., 2004c).

Original Publications

I. Meier, S., Häussinger, D., Pokidysheva, E., Bächinger, H. P., and Grzesiek, S.

Determination of a high-precision NMR structure of the minicollagen cysteine rich domain from Hydra and characterization of its disulfide bond

FEBS Letters, in press (2004).

II. Meier, S., Pokidysheva, E., Milbradt, A., Renner, C., Häussinger, D., Bächinger, H. P., Moroder, L., Grzesiek, S., Holstein, T., Özbek, S., and Engel, J.

The structure of the Cys-rich terminal domain of Hydra minicollagen, which is involved in disulfide networks of the nematocyst wall.

Journal of Biological Chemistry, in press (2004).

Determination of a high-precision NMR structure of the minicollagen cysteine rich domain from *Hydra* and characterization of its disulfide bond formation

Sebastian Meier[1], Daniel Häussinger[1], Elena Pokidysheva[2], Hans Peter Bächinger[3], and Stephan Grzesiek[1]

[1]Division of Structural Biology and [2]Division of Biophysical Chemistry, Biozentrum der Universität Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland
[3]Shriners Hospital for Children and Department of Biochemistry and Molecular Biology, Oregon Health & Science University, Portland, OR 97239, USA

Key words: nematocyst, disulfide bond, heteronuclear NMR, residual dipolar couplings

Abstract

A high-precision solution structure of the C-terminal minicollagen cysteine rich domain of *Hydra* has been determined using modern heteronuclear and weak alignment NMR techniques at natural isotope abundance. The domain consists of only 24 amino acids, six of which are prolines and six are cysteines bonded in disulfide bridges that constrain the structure into a new fold. The redox equilibrium of the structure has been characterized from a titration with glutathione. No local native structures are detectable in the reduced form. Thus oxidation and folding are tightly coupled.

Mailing address:
Stephan Grzesiek
Department of Structural Biology, Biozentrum
University of Basel, CH-4056 Basel, Switzerland
Phone: ++41 61 267 2100, FAX: ++41 61 267 2109
Email: Stephan.Grzesiek@unibas.ch

file: mcrd6.doc
last saved: 17/4/04 12:50 PM
word count: 3286
char count: 17849

1. Introduction

Minicollagens from *Hydra* are the smallest known collagens to date [1]. They are found in the walls of nematocysts, which are explosive organelles formed from a post-Golgi vesicle in *Hydra*, jellyfish, corals and other *Cnidaria*. The explosive discharge of nematocysts is a specialized exocytosis process which achieves accelerations of > 40000 g and is one of the fastest natural processes [2]. This process is enabled by an estimated intramolecular osmotic pressure of 150 bar, which results from an enrichment of poly-γ-glutamate in the capsule matrix during maturation [3].

Nematocyst walls are stabilized by a collageneous matrix [4], which depends on disulfide crosslinking [5] to withstand the extreme osmotic pressure. The *Hydra* capsule wall mainly consists of the proteins minicollagen and NOWA [6-8] (Fig. 1). Both proteins contain highly homologous cysteine- and proline-rich domains, henceforward called minicollagen cysteine rich domains (MCRDs) [9]. Minicollagens are expressed as soluble precursor forms with intramolecular disulfide bonds in their MCRDs [6,8]. In nematocyst development a wall hardening occurs, during which the intramolecular MCRD disulfide bonds are most likely reshuffled to intermolecular bonds, thereby crosslinking molecules within the capsule wall [6]. The simultaneous appearance of MCRDs in minicollagen and NOWA points to a concerted function in wall hardening. A more detailed comparison and discussion of the biological context is given in an accompanying paper [10].

Here we report the NMR determination of the solution structure of the C-terminal MCRD domain of minicollagen 1. The domain consists of only 24 amino acids, six of which are cysteines and six are prolines. The well-defined structure forms in high yields within two hours when disulfide bond formation is catalyzed by glutathione. No local native structures are detectable in the reduced form of the MCRD. Thus oxidation and folding are tightly coupled.

2. Materials and Methods

The MCRD Minicol1C (Fig. 1, underlined) was chemically synthesized without isotope enrichment and purified after oxidation as described elsewhere [10]. NMR samples were prepared to 5 mM peptide concentration in 5 mM phosphate, 2 mM NaN_3, 95% H_2O/ 5% D_2O, pH 6.5. A 2 mM NMR sample of identical pH and ionic strength was lyophilized and redissolved in D_2O. For the measurement of residual dipolar couplings (RDCs) a 3 mM peptide sample was prepared in 10 mg/ml Pf1 phage solution (Asla Biotech) with 5 mM phosphate of pH 6.5 and 350 mM NaCl to tune down electrostatic interactions of the positively charged MCRD (pI 8.74) with the phages [11], resulting in a maximal coupling $A_{zz}^{NH} = 11.5$ Hz.

2.1 NMR spectroscopy

All multidimensional NMR measurements were performed at 15 °C on Bruker DRX 600 and 800 spectrometers. ^{1}H, ^{13}C and ^{15}N assignments were obtained from TOCSY, NOESY as well as ^{15}N and ^{13}C natural abundance HSQC and HMBC spectra. $^{1}D_{HN}$ and $^{1}D_{CH}$ RDCs were obtained by nondecoupled HSQCs. Spectra were processed with NMRPipe [12] and analyzed with PIPP [13].

2.2 Structure determination

In addition to NOEs, the input for the structure calculations contained backbone $^{1}D_{HN}$, backbone and sidechain $^{1}D_{CH}$ RDCs as well as dihedral angle restraints obtained with TALOS [14] (Table 1). All five X-Pro peptide groups were found to be in *trans* conformation by specific, strong sequential $d_{\alpha\delta}$ NOEs. Structures were calculated in CNS [15] with a simulated annealing protocol using RDCs without *a priori* knowledge of the orientation tensor via the ISAC routine [16]. Structure representations were generated by MOLMOL [17].

2.3 Reductive unfolding and oxidative folding

MCRD (0.3 mM) was reduced completely with 6 mM tris(2-carboxyethyl)-phosphine hydrochloride (TCEP) at pH 7.5. Oxidative refolding was started by the addition of 40 mM oxidized glutathione (GSSG) in 5 mM phosphate buffer, pH 7.5. Refolding was monitored by 1D ^{1}H NMR experiments and fitted to monoexponentials with ProFit (Quantumsoft).

Equilibrium unfolding of 0.3 mM MCRD in 9 mM GSSG, 100 mM phosphate (pH 7.0, 25 °C) was induced by the addition of varying amounts of reduced glutathione (GSH) in 100 mM phosphate (pH 7.0) and monitored by 1D ^{1}H NMR after 12 h incubation in airtight Shigemi tubes. Concentrations of GSH and GSSG were validated from the integrals of the respective ^{1}H NMR spectra.

3. Results

3.1 Assignment, structure determination and disulfide pattern of MCRD

Sequence specific ^{1}H, ^{13}C, and ^{15}N assignments were obtained from TOCSY and NOESY as well as heteronuclear HSQC and HMBC spectra acquired at natural isotope abundance (Fig. 2). All cysteine residues are bonded in disulfide bridges as evidenced by their $^{13}C^{\beta}$ shifts of 40 to 43.5 ppm [18]. The homogeneous spectra show only one single protein species (Fig. 2). All nonlabile protons could be assigned for structure calculations. Water exchange results in broadening of amide protons of Cys2 and Gln16 (Fig. 2), which is abolished at lower pH values.

In total, 331 distance constraints were obtained from NOESY spectra in H_2O and D_2O (Table 1). In addition, 37 $^{1}D_{HN}$ and $^{1}D_{CH}$ RDCs for residues 2-21 were measured at natural isotope abundance in weakly oriented samples and incorporated in the structure calculations (Fig. 3, Table 1) to further improve the structural quality. The resulting high resolution structure with a heavy atom backbone RMSD of 0.16 Å as well as pairwise interresidual

$d_{H\beta iH\beta k}$ NOEs in all disulfide bonds prove the disulfide pattern Cys2-Cys18, Cys6-Cys14 and Cys10-Cys19 (Fig. 4).

The N-terminus of MCRD (^{1}PCP^3) is an elongation of the preceding polyproline II helix in minicollagen 1 and adopts (ϕ/ψ)-values of (-60° – -90°/125° – 155°) similar to conformations in the ideal PPII helix (-75°/145°). The subsequent short α-helix (residues 4-8) is followed by an inverse γ-turn (9-11), a type I β-turn (11-14) and a type III β-turn (15-18). The highly conserved Pro12 stabilizes the βI turn topology as canonical residue i+1. The structure of MCRD (including the helix) can be viewed as a consecutive arrangement of turns, where each turn contains one cysteine that is directed away from the turn to form the aforementioned long-range disulfide pattern.

Backbone $H^N \rightarrow O'$ hydrogen bonds in the turns from residues 6→3, 8→4, 9→5, 11→9, 14→11, 18→15 are evident from the structure, upfield shifted $^1H^N$ chemical shifts (Fig. 2) and reduced $^1H/^2H$ exchange. A schematic overview of turns, hydrogen bonds and disulfide bridges is shown in Fig. 4.

$^1H^N$ T_2 relaxation times >120 ms of 2 mM MCRD at 25 °C show that MCRD is a monomer in solution and therefore does not contribute to the noncovalent stabilization of the minicollagen trimer by intermolecular interactions. Accordingly, the reduction of the MCRD disulfide bonds was found to have no influence on the transition temperature of 45 °C for minicollagen 1 trimerization [8].

3.2 Reductive unfolding and oxidative refolding

MCRD is readily reduced by an excess of TCEP or glutathione at pH 7.5. Upon reduction, all hydrogen bonds in the turns are abolished as evident from a change of amide $^1H^N$ resonance shifts to random coil values of >7.9 ppm for all residues (Fig. 5A). NOESY spectra give additional evidence that the structure is largely in extended or in random coil conformation, as the strong d_{HNHN} NOEs of the turns are abolished. Moreover, the ^{15}N

resonance of Thr13 assumes a random coil value (114.3 ppm), whereas position i+2 of a βI turn usually has characteristic ^{15}N upfield shifts [14] (103.7 ppm in the native MCRD structure, Fig. 2). Thus the disulfide pattern in MCRD obviously does not result from a non-covalently folded precursor, in which cysteines are already properly preoriented. Rather oxidation and folding occur simultaneously, and the disulfide bonds stabilize the natively folded structure.

Oxidative folding of completely reduced MCRD was started by the addition of a >100-fold excess of oxidized glutathione as its natural oxidant. Refolding was easily monitored by 1D NMR spectra of upfield shifted $^1H^N$ turn resonances (Fig. 5A) which do not overlap with the glutathione signal. The 1D signals apparently correspond either to the native folded species or a single unfolded state, whereas other intermediates are not highly populated. Folding proceeds to a steady state within about 2 h (Fig. 5A). Fitting of signal intensities to monoexponentials shows a decay rate of 1.13 +/- 0.06 h^{-1} for resonances corresponding to the unfolded species, whereas the native signals are formed at a slightly slower rate of 0.78 +/- 0.03 h^{-1} (Fig. 5B). This may indicate a hidden, albeit lowly populated intermediate, where for example a non-random coil and nonnative oxidized ensemble rearranges slowly to the native oxidized state [19].

In order to obtain information on the redox potentials of the disulfides, equilibrium unfolding of MCRD was studied by a titration of the ratio of reduced (GSH) versus oxidized (GSSG) glutathione [20]. The ratio of reduced versus oxidized MCRD, $[MCRD_{red}]/[MCRD_{ox}]$, was obtained from an integration of the corresponding 1D $^1H^N$ and $^1H^\delta$ resonances of threonine-13 and tyrosine-17, respectively (Fig. 5C). Assuming that the loss of the folded oxidized structure is caused by the cooperative reduction of n disulfide bridges, the reaction can be described by

$MCRD_{red} + n\ GSSG \leftrightarrow MCRD_{ox} + n\ (2\ GSH)$

corresponding to an equilibrium of the form $K_{eq} = [MCRD_{ox}]/[MCRD_{red}]\cdot([GSH]^2/[GSSG])^n$ [20]. The cooperativity n can be obtained from the slope of a Hill plot of $\log([MCRD_{ox}]/[MCRD_{red}])$ vs. $\log([GSH]^2/[GSSG])$ as shown in Fig. 5C. The transition from the folded, oxidized to the unfolded, reduced form is non-cooperative (n = 0.92) at low $[GSH]^2/[GSSG]$ (< 100 mM), whereas it becomes more cooperative (n = 2.5) at higher $[GSH]^2/[GSSG]$ ratios. This can be explained by a model where at low reducing potential only one disulfide opens, whereas for stronger reducing potentials all three disulfides open. Apparently, the opening of the first disulfide is sufficient to account for the signal loss of the folded resonances. The midpoint of the redox reaction ($[MCRD_{ox}]/[MCRD_{red}] = 1$) is found at a $[GSH]^2/[GSSG]$ ratio of 30 mM. This corresponds to an apparent, effective redox potential E'_o(MCRD) of -185 mV, where the Nernst relation $E'_o(MCRD) = E'_o(glutathione) - RT/2F \ln([GSH]^2/[GSSG])_{midpoint}$ and a standard redox potential for glutathione E'_o(glutathione) of -230 mV at 25°C [21] have been used. Judging from this effective equilibrium constant, MCRD should be predominantly folded with all disulfide bonds intact at physiological concentrations of GSH and GSSG in microsomes ($[GSH]^2/[GSSG] < 3$ mM [22,23]).

4. Discussion

The MCRD domain adopts a new fold that is highly defined by its three disulfide bridges. This structure arises during oxidative folding without populating other structured intermediates to high concentrations. The repeating cysteine pattern CXXXC in MCRD with three intervening residues between the cysteines has been shown to be highly unfavorable for nearest-neighbor disulfide bond formation [24]. Accordingly, the MCRD structure only

contains longe-range disulfide bonds, which presumably lock into the native conformation during oxidative refolding after the formation of the central, nonaccessible disulfide bridge Cys6-Cys14. The wealth of proline residues as well as the formation of hydrogen bonds may restrict the conformational search and stabilize local structure in MCRD during this folding process.

The structure of oxidized MCRD is not unfolded in 8 M urea as evident from ^{1}H NMR spectra (data not shown), whereas even local structures are abolished by reduction. This underlines the high importance of the disulfide bonds relative to hydrophobic interactions in stabilizing the fold and local native structures. In consequence the cysteine pattern, but not the hydrophobic residues of MCRD domains are highly conserved [7].

The MCRD domain fulfills its biological role by intermolecular disulfide bond reshuffling [7] . This process is most likely catalyzed [25] and probably involves at least one of the two solvent exposed disulfide bridges Cys2-Cys18 and Cys10-Cys19. Of these, the N-terminal Cys2-Cys18 should be particularly accessible in the N-terminal MRCDs of minicollagen. The equilibrium unfolding with the natural reductant GSH shows the presence of such a susceptible disulfide bond in MCRD.

Data deposition

The atomic coordinates of the 10 lowest energy CNS conformers have been deposited at the RCSB Protein Data Bank (www.rcsb.org) under PDB accession number 1SP7.

Acknowledgement

We thank Jürgen Engel and Suat Özbek for initiating this project and for helpful discussions. This work was supported by SNF grant 31-61’757.00 to S.G.

References

[1] Engel, J. (1997) Science 277, 1785-1786.
[2] Holstein, T. and Tardent, P. (1984) Science 223, 830-833.
[3] Weber, J. (1990) J. Biol. Chem. 265, 9664-9.
[4] Lenhoff, H.M., Kline, E.S. and Hurley, R. (1957) Biochim. Biophys. Acta 26, 204-5.
[5] Blanquet, R. and Lenhoff, H.M. (1966) Science 154, 152-3.
[6] Engel, U., Pertz, O., Fauser, C., Engel, J., David, C.N. and Holstein, T.W. (2001) Embo J. 20, 3063-3073.
[7] Engel, U., Ozbek, S., Streitwolf-Engel, R., Petri, B., Lottspeich, F., Holstein, T.W., Oezbek, S. and Engel, R. (2002) J. Cell Sci. 115, 3923-34.
[8] Ozbek, S., Pertz, O., Schwager, M., Lustig, A., Holstein, T. and Engel, J. (2002) J. Biol. Chem. 277, 49200-49204.
[9] Kurz, E.M., Holstein, T.W., Petri, B.M., Engel, J. and David, C.N. (1991) J. Cell Biol. 115, 1159-69.
[10] Pokidysheva, E. et al. (2004) J. Biol. Chem., (under revision).
[11] Hansen, M.R., Mueller, L. and Pardi, A. (1998) Nat. Struct. Biol. 5, 1065-1074.
[12] Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) J. Biomol. NMR 6, 277-293.
[13] Garrett, D.S., Powers, R., Gronenborn, A.M. and Clore, G.M. (1991) Journal of Magnetic Resonance 95, 214-220.
[14] Cornilescu, G., Delaglio, F. and Bax, A. (1999) J. Biomol. NMR 13, 289-302.
[15] Brunger, A.T. et al. (1998) Acta Crystallogr. D 54, 905-921.
[16] Sass, H.J., Musco, G., Stahl, S.J., Wingfield, P.T. and Grzesiek, S. (2001) J. Biomol. NMR 21, 275-80.
[17] Koradi, R., Billeter, M. and Wuthrich, K. (1996) J. Mol. Graphics 14, 51-&.
[18] Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L. and Sykes, B.D. (1995) J. Biomol. NMR 6, 135-40.
[19] Welker, E., Narayan, M., Wedemeyer, W.J. and Scheraga, H.A. (2001) Proc. Natl. Acad. Sci. U S A 98, 2312-6.
[20] Hawkins, H.C., de Nardi, M. and Freedman, R.B. (1991) Biochem. J. 275 (Pt 2), 341-8.
[21] Fasman, G.D. (1976) Handbook of Biochemistry and Molecular Biology, CRC Press, Cleveland.
[22] Hwang, C., Sinskey, A.J. and Lodish, H.F. (1992) Science 257, 1496-502.
[23] Bass, R., Ruddock, L.W., Klappa, P. and Freedman, R.B. (2004) J. Biol. Chem. 279, 5257-62.
[24] Zhang, R.M. and Snyder, G.H. (1989) J. Biol. Chem. 264, 18472-9.
[25] Ozbek, S., Engel, U. and Engel, J. (2002) J. Struct. Biol. 137, 11-14.
[26] Cornilescu, G., Marquardt, J.L., Ottiger, M. and Bax, A. (1998) J. Am. Chem. Soc. 120, 6836-6837.
[27] Laskowski, R.A., Rullmann, J.A.C., MacArthur, M.W., Kaptein, R. and Thornton, J.M. (1996) J. Biomol. NMR 8, 477-486.

Figure Legends

Fig. 1: Top: Domain organization of *Hydra* minicollagen 1 and NOWA. MCRDs are indicated as circles. Bottom: sequence alignment of minicollagen and NOWA MCRDs. For simplicity residues of the Minicol1C MCRD used in this study are renumbered from 1 to 24.

Fig. 2: Natural abundance ^{1}H-^{15}N HSQC spectrum of Minicol1C MCRD (5 mM, experimental time 1.5 h). Cys2 is broadened beyond detection due to solvent exchange but is identified at pH 4 at the indicated position.

Fig. 3: Top: 1D traces of heteronuclear natural abundance spectra for the measurement of RDCs. Bottom: comparison of measured RDCs and RDCs derived from the calculated average structure of Minicol1C MCRD.

Fig. 4: The structure of Minicol1C MCRD. Left: stereo view in an all-heavy-atom representation of the 10 lowest energy conformers out of 100 structures calculated. The backbone is displayed in dark orange and sidechains in cyan, except for cysteine sidechains in yellow. The charged C-terminal sequence $K^{22}R^{23}K^{24}$ is disordered at pH 6.5 in the absence of additional salt. Center: ribbon representation in the same orientation as stereo view shows that disulfide bonds Cys10-Cys19 and Cys2-Cys18 are solvent accessible, whereas the central disulfide bond Cys6- Cys14 is inaccessible. Right: disulfide and hydrogen bond topology.

Fig. 5: Oxidative folding and reductive unfolding of MCRD. (A) 1D spectra of MCRD folding initiated by addition of 40 mM GSSG to 0.3 mM MCRD previously reduced with TCEP. Spectra were recorded at the indicated times after starting the oxidation reaction. Backbone $^{1}H^{N}$ (plain numbers) and sidechain resonances are indicated. (B) Time course of

glutathione catalyzed refolding as monitored by the indicated resonance intensities derived from the spectra in (A). (C) Hill plot of equilibrium reductive unfolding monitored in a solution of varying GSH/GSSG content (see text). Solid lines correspond to linear fits of the low and high $[GSH]^2/[GSSG]$ regions of the correlation.

Table 1: Statistics of the MCRD NMR structure[1)]

RMSDs from experimental distance constrains [Å]	
all (331) [2)]	0.044 ± 0.001
RMSDs from dipolar coupling constraints (37)[3)]	1.2 Hz
NMR quality factor Q[4)]	0.20
RMSDs from dihedral constraints [°] (15)[5)]	0.83± 0.03
Deviation from the idealized covalent geometry	
bonds [Å]	0.0053 ± 0.0002
angles [°]	0.74 ± 0.01
impropers[6)] [°]	0.72 ± 0.01
Coordinate precision [Å][7)]	
backbone non-hydrogen atoms	0.16
all non-hydrogen atoms	0.56
Percentage of non-gly, non-pro residues in Ramachandran regions[8)]	
core	93.3
allowed	6.7
generous	0.0
disallowed	0.0

[1)] The statistics were obtained from a subset of the 10 lowest energy structures out of 100 calculated with a CNS [15] simulated annealing protocol using dipolar restraints implemented in form of the ISAC routine [16]. The number of constraints is given in parentheses. Coordinate precision and Ramachandran plot quality are reported for the core of the structure (residues 2-21) excluding the flexible N- and C-terminus.

[2)] NOEs comprise 60 intraresidual NOEs, 132 sequential NOEs ($|i-j|=1$), 114 short range NOEs ($1<|i-j|\leq 5$) and 25 long range NOEs ($|i-j|>5$). Hydrogen bond constraints were not imposed.

[3)] RDC data comprise 13 ^{1}H-^{15}N, 18 ^{1}H-$^{13}C^{\alpha}$ and 6 sidechain ^{1}H-^{13}C methyl and methine dipolar one-bond couplings. Only RDCs of the nonflexible MCRD core residues are incorporated.

[4)] The NMR quality factor Q is defined as the ratio of the rmsd between observed and calculated couplings and the rms of the observed couplings [26].

[5)] The dihedral angle constraints comprise 4 ϕ and 9 ψ angles obtained for residues 2-21 with ^{1}H, ^{15}N and ^{13}C natural abundance chemical shift assignments as inputs for TALOS [14], as well as 2 χ^{1} angles obtained from NOESY peak intensities for two residues with very few long-range NOE data (Cys2 in a PCPP sequence and solvent exposed Gln9).

[6)] The improper torsion angle restraints serve to maintain planarity and chirality.

[7)] The coordinate precision is defined as the average root mean square distance between the individual simulated annealing structures and the mean coordinates. Values are reported for core residues[1)].

[8)] Values are calculated with the program PROCHECK-NMR [27] for the residues 2-21[1)]. The inverse γ turn topology of Cys10 (Fig. 4) is the only of these residues to fall outside the core region of the Ramachandran plot.

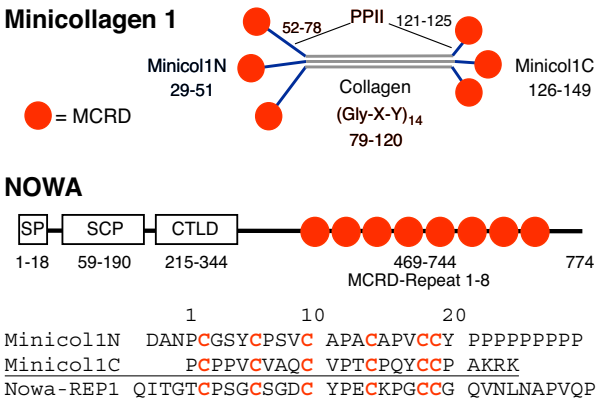


Figure 1: Meier et al.

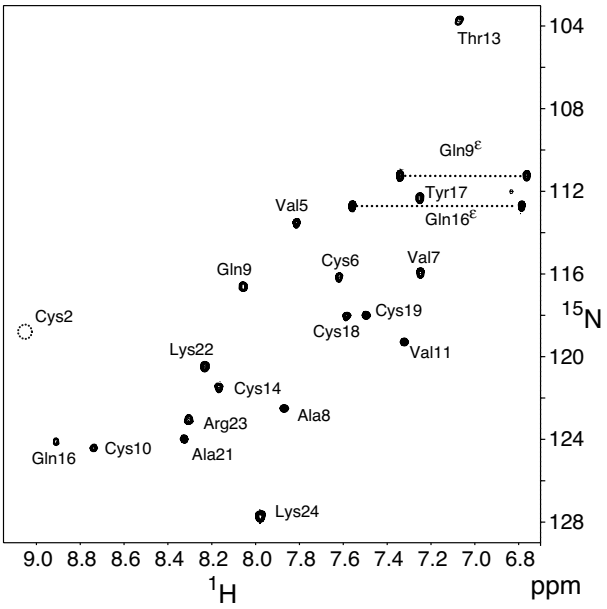


Figure 2: Meier et al.

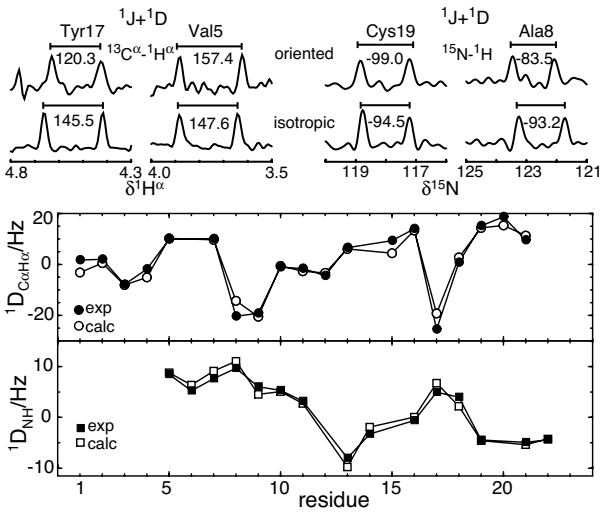


Figure 3: Meier et al.

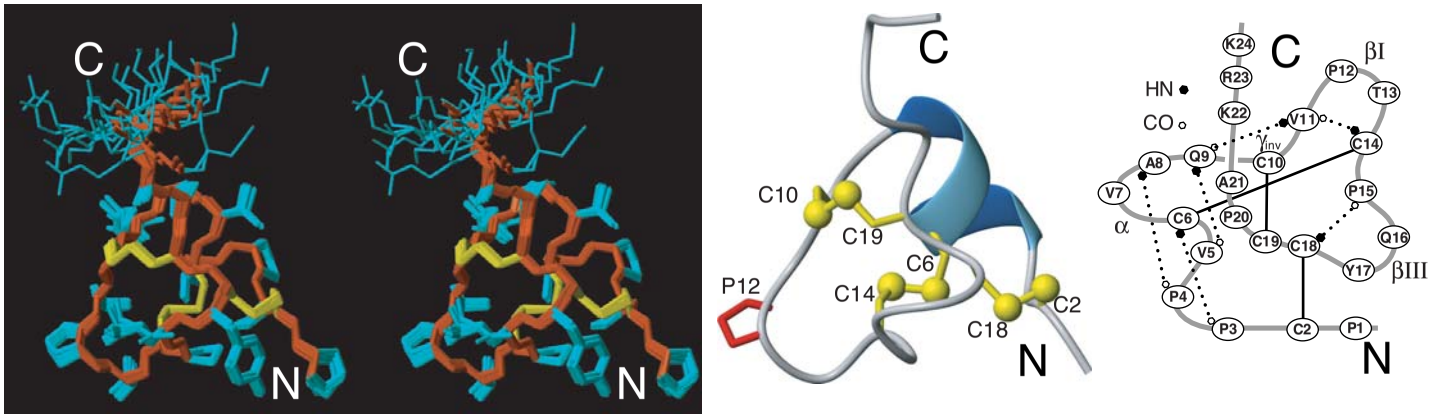


Figure 4: Meier et al.

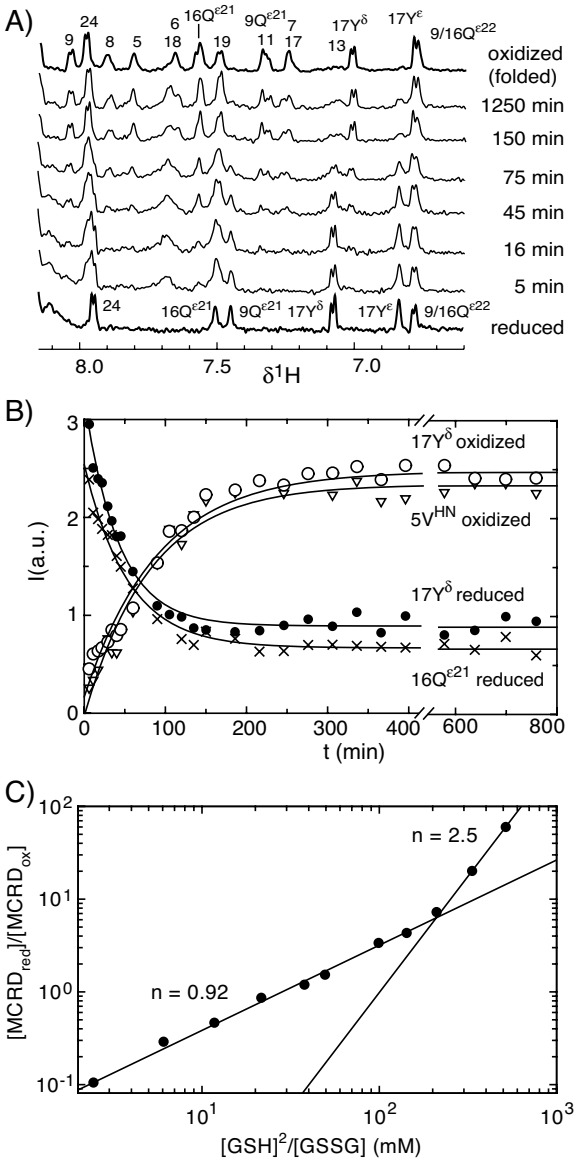


Figure 5: Meier et al.

The structure of the Cys-rich terminal domain of *Hydra* minicollagen, which is involved in disulfide networks of the nematocyst wall

Elena Pokidysheva[1*], Alexander Milbradt[2*], Sebastian Meier[3*], Christian Renner[2#], Daniel Häussinger[3], Hans Peter Bächinger[4], Luis Moroder[2], Stephan Grzesiek[3#], Thomas W. Holstein[5], Suat Özbek[1#] and Jürgen Engel[1]

[1]Department of Biophysical Chemistry, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

[2]Max-Planck-Institute for Biochemistry, D-82152 Martinsried, Germany

[3]Department of Structural Biology, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

[4]Shriners Hospital for Children and Department of Biochemistry and Molecular Biology, Oregon Health & Science University, Portland, OR 97239, USA

[5]Institute of Zoology, Technical University of Darmstadt, Schnittpahnstrasse 10, D-64287 Darmstadt, Germany

*Note: these authors contributed equally to this work

[#]To whom correspondence should be addressed: 1) E-mail: suat.oezbek@unibas.ch

2) E-mail: stephan.grzesiek@unibas.ch

3) E-mail: renner@biochem.mpg.de

Running title: structure of minicollagen Cys-rich domain

Abstract

The minicollagens found in the nematocysts of *Hydra* constitute a family of invertebrate collagens with unusual properties. They share a common modular architecture with a central collagen sequence ranging from 14-16 Gly-X-Y repeats flanked by polyproline / hydroxyproline stretches and short terminal domains that show a conserved cysteine pattern (CXXXCXXXCXXXCXXXCC). The minicollagen cysteine-rich domains (MCRD) are believed to function in a switch of the disulfide connectivity from intra- to intermolecular bonds during maturation of the capsule wall. The solution structure of the C-terminal fragment including a MCRD of minicollagen-1 was determined in two independent groups by ^{1}H NMR. The corresponding peptide comprising the last 24 residues of the molecule was produced synthetically and refolded by oxidation under low protein concentrations. Both presented structures are identical in their fold and disulfide connections (Cys2-Cys18, Cys6-Cys14, Cys10-Cys19) revealing a robust structural motif that is supposed to serve as the polymerization module of the nematocyst capsule.

Introduction

Minicollagens of nematocysts in *Hydra*, corals and other cnidaria are very unusual proteins with structural properties not shared by other invertebrate or vertebrate collagens. From both the N- and the C-terminus of the collagen triple helix emerge three polyproline-II-type helices, which consist of 5 to 23 proline or hydroxyproline residues (1-3). Each of the PPII-helices is terminated by a small Cys-rich domain, termed minicollagen Cys-rich domain (MCRD). The N- and the C-terminal MCRDs are homologous and share the cysteine pattern CXXXCXXXCXXXCXXXCC. A small propeptide region preceding the N-terminal MCRD is cleaved off during expression, and mature minicollagen has a rather symmetrical appearance with closely similar structural elements at both sides. This bipolar nature of minicollagen is unique among all other known collagens (4-7) and suggests a special function.

A unique function of minicollagen is also suggested by its restricted appearance in the capsule wall of nematocysts. Nematocysts are complex explosive organelles, which basically consist of a capsule, an inverted tubule armed with spines and an operculum. The tubule is connected to the capsule wall and is twisted in many turns inside the osmotically charged capsule matrix. Following stimulation, the internal tube is expelled, the osmotic pressure is released and capsule contents including toxins are released at the tubule end. This specialized form of exocytosis proceeds with ultra-fast rates and accelerations comparable to those of a fired bullet (8).

The capsule wall resists more than 150 atmospheres of osmotic pressure in the charged state. For *Hydra* nematocysts it consists mainly of two proteins: minicollagen and nematocyst outer wall antigen (NOWA) (9-11). The two proteins can be dissolved from capsule preparations only under reducing conditions. Already at a time at which the amino acid sequence of minicollagens was unknown it was found that these collagens formed disulfide cross-linked polymers that were insoluble in SDS but easily soluble in the presence of a reducing agent (12,13). Following the discovery of a family of minicollagens (3) and recombinant expression of minicollagen-1 (9) it was found that the proteins are expressed in a soluble precursor form present in the endoplasmatic reticulum and post-Golgi vacuoles in *Hydra*. They are converted to the disulfide linked assembly form of the nematocyst wall upon wall compaction, during which

a dense and well-defined capsule wall is formed (9). The morphological changes, a loss of accessibility to antibodies against minicollagen-1 and a parallel loss of solubility under non-reducing conditions suggested a close link between disulfide polymerization and the condensation of wall proteins. Both processes provide an explanation for the unusually high tensile strength of the mature nematocyst wall.

Minicollagen-1 of *Hydra* recombinantly expressed in mammalian cells contains internal disulfide bonds in its MCRDs but no interchain disulfide cross-links between chains (9). The trimeric collagen molecules dissociated into single chains when heated to 45°C under non-reducing conditions (10). Reduction did not influence this transition temperature indicating that only the collagen domain is responsible for trimerization. Recombinant minicollagen-1 was found to form aggregates in electron micrographs but no disulfide bridges were formed spontaneously under *in vitro* conditions (9). This material behaved like the precursor form *in vivo* and could be solubilized by SDS or other denaturants under non-reducing conditions. It was therefore concluded that disulfide isomerases or other external parameters are required for a disulfide reshuffling process by which internal disulfide bridges are converted to intermolecular links (14).

More recent results suggest, that the structural function of minicollagens in wall hardening is complemented by NOWA (11). Altogether ten domains were identified in this protein, namely a SCP-domain (smart00198), a C-type lectin domain (CTLD, smart00034) and 8 C-terminal domains with homology to the MCRDs. In particular, the pattern of six cysteines is shared by the corresponding domains in NOWA and minicollagen. The presence of homologous Cys-rich domains in both proteins suggested a joint function in disulfide-mediated polymerization. Supportive evidence was obtained by analysis of breakdown products of native nematocyst capsules after limited sonification without reduction (Özbek et al., in preparation). Minicollagen and NOWA formed disulfide cross-linked units that like the entire capsule were readily dissolved under even mild reducing conditions.

Disulfide reshuffling processes and formation of disulfide-linked complexes are common mechanisms in the extracellular space and some recently explored systems may be referenced (15-19). The minicollagen/NOWA system may stand as a prototype for the controlled formation of a highly stable matrix layer by disulfide linkage. To understand the assembly mechanism, the 3-dimensional structure of the

MCRDs involved is needed. The core of the MCRD consists of only 20 residues with 6 closely positioned Cys residues. At the start of this work it was not known whether the very small domains were autonomous or stable only as part of a larger structure. The disulfide connections were also not known and could not be explored by limited proteolysis because of the lack of suitable cleavage sites. These problems were approached by solving the structure of the C-terminal Cys-rich domain of minicollagen-1 by NMR spectroscopy. A new fold was found, with disulfide connections between Cys-2 and Cys-18, Cys-6 and Cys-14, and Cys-10 and Cys-19. Oxidative folding occurs at high rates in the presence of a GSSG/GSH redox buffer with formation of almost exclusively one single isomer. A highly conserved Pro residue located between Cys-10 and Cys-14 induces and probably directs correct disulfide connections. The presented structure proposes likely candidates for disulfide reshuffling supposed to be a key reaction during nematocyst morphogenesis.

Experimental Procedures

Peptide Synthesis and oxidation

Preparation 1

The 6xStBu protected peptide Ac-PCPPVCVAQCVPTCPQYCCPAKRK-NH_2 was synthesized on Fmoc-Rink-Amide PEGA resin by the Fmoc/tBu strategy using double couplings with Fmoc-Xaa-OH/HBTU/HOBt/DIEA (4:4:4:8), intermediate Fmoc cleavage with 20 % piperidine in DMF and acetic anhydride/DIEA (4:8) for N-terminal acetylation. After resin cleavage/deprotection with TFA/phenol/H_2O/thioanisole/1,2-ethanedithiol (82.5:5:5:5:2.5) the product was isolated by RP-HPLC; yield: 6 %; HPLC: t_R = 12.5 min (>98 %); ESI-MS: *m/z* = 1055.0 $[M+3H]^{3+}$; 1581.8 $[M+2H]^{2+}$; M_r = 3162.31 calculated for $C_{136}H_{230}N_{32}O_{29}S_{12}$. The Cys protecting groups were cleaved in TFE/H_2O with tributylphosphine (60 equiv) at RT for 5 h and the resulting fully deprotected peptide was oxidized at pH 8.0 in the presence of GSSG/GSH (9 equiv, 10:1) under air atmosphere at 7 °C. The crude product was purified by preparative size exclusion column chromatography; yield: 20 %; HPLC: t_R = 12.3 min (>98 %); ESI-MS: *m/z* = 1314.8 $[M+2H]^{2+}$, 876.4 $[M+3H]^{3+}$; M_r = 2627.23 calculated for $C_{112}H_{176}N_{32}O_{29}S_6$.

Preparation 2

The linear peptide was synthesized with an ABI 433A synthesizer. Couplings were carried out on a Fmoc-Lys(Boc)-PEG-PS resin (Perseptive Biosystems, 0.21 mmol/g) using Fmoc-amino acids (Fmoc-Arg(Pbf), Fmoc-Lys(Boc), Fmoc-Tyr(tBu), Fmoc-Cys(Trt), and Fmoc-Gln(Trt)) (Anaspec). HATU (O-(7-azabenzotriazol-1-yl)-1.1.3.3-tetramethyluronium hexafluorophosphate, (Perseptive Biosystems) (4.0 eq.) / Diisopropylethylamine mediated peptide couplings. The peptide was cleaved from the resin with trifluoroacetic acid, containing thioanisol (5 % v/v), ethanedithiol (3 % v/v), and anisole (2 % v/v). After precipitation in diethylether, the crude peptide was purified by preparative HPLC (Vydac® C18, 10-15 µm, 300Å, 250 x 50 mm, W.R. Grace, MD, USA). The peptide was characterized by electrospray

ionization/quadrupole/time-of-flight mass spectrometry and amino acid analysis. A mass of 2590.9 Da was determined and the expected amino acid composition was found.

The lyophilized purified peptide was dissolved in 25 mM Tris/HCl, pH 6.0, containing 100 mM NaCl at a concentration of 0.02 mM. Folding was allowed to proceed under N_2 for 72 to 96 hours at 4 ^{o}C. Oxidation was induced by changing the pH to 8.0-8.1 with saturated Tris, and the addition of reduced and oxidized glutathione (10:1 molar ratio) to a final concentration of 1 mM. The solution was exposed to air at 4 ^{o}C for four to ten days. Oxidation was stopped by the addition of trifluoroacetic acid to give a pH of 1.3. The oxidized peptide was purified by preparative HPLC. The yield of oxidized peptide was 103 mg (16 % of theoretical yield). Mass spectroscopy of the oxidized peptide showed a molecular mass of 2585.1 Da.

NMR Data Collection

Structure 1

NMR experiments for conformational analysis were carried out at 283 K on Bruker DRX500, DMX750 and DX900 spectrometers using a 3 mM sample of the peptide dissolved in H_2O/D_2O (9:1) mixture at pH 3.5. Resonance assignments were performed according to the method of Wüthrich (38). 173 experimental interproton distance constraints were extracted from 2D-NOESY (39) experiments with mixing times between 75 and 200 ms. Five hydrogen bonds were identified from temperature shifts and H/D exchange. Acceptor carbonyl groups were identified in initial structure calculations. The NOE intensities were converted into interproton distance constraints using the following classification: very strong (vs) 1.7-2.3 Å, strong (s) 2.2-2.8 Å, medium (m) 2.6-3.4 Å, weak (w) 3.0-4.0 Å, very weak (vw) 3.2-4.8 Å and the distances of pseudo atoms were corrected as described by Wüthrich (38). Distance geometry (DG) and molecular dynamics-simulated annealing (MD-SA) calculations were performed with the INSIGHTII 98.0 software package (Accelrys, San Diego, CA) on Silicon Graphics O2 R5000 computers (SGI, Mountain view, CA) as described recently (25). In brief, hundred structures were generated by distance geometry and refined with MD-SA steps. The experimental constraints were applied at every stage of the calculations.

The coordinates and structural restraints have been deposited in the Brookhaven Protein Data Bank under accession number 1SOP.

Structure 2

Chemically synthesized C-terminal MCRD of minicollagen-1 was purified after oxidation according to purification procedure 2. The determination of the NMR structure (PDB accession number 1SP7) was carried out in 5 mM sodium phosphate buffer, pH 6.5 at 15°C by using homonuclear and heteronuclear techniques and information from weak alignment. Details of this structure determination will be published separately. Structure representations were generated with MOLMOL (43).

Analytical ultracentrifugation

Sedimentation equilibrium experiments were performed on a Beckman Optima XL-A analytical ultracentrifuge (Beckman Instruments) equipped with 12-mm Epon double-sector cells in an An-60 Ti rotor. The MCRD was analyzed in 5mM Tris buffer (pH 7) without salt or containing 100 mM NaCl at 20°C. The peptide concentrations were adjusted to 0.2-0.8 □M. Sedimentation equilibrium scans were carried out at 48,000 rpm. Molecular masses were evaluated from ln*A versus* r^2 plots, where *A* is the absorbance at 277nm and r is the distance from the rotor center. A partial specific volume of 0.73 ml/g was used for all calculations.

Results

The MCRD constitutes a conserved sequence module in nematocyst minicollagens and in NOWA

The sequence of minicollagen–1 with domain indications and alignment of cysteine-rich domains from different minicollagens and NOWA are represented in Figure 1. Minicollagen-1 consists of an N-terminal MCRD, N-terminal polyproline region, a central collagen sequence, a shorter C-terminal polyproline stretch followed by a C-terminal MCRD (Fig. 1A). The preceding propeptide is cleaved off during recombinant expression of minicollagen-1 and probably also in *Hydra* (9). As already mentioned in the introduction the overall sequence homology of the Cys-rich domains is not very high but the cysteine pattern is identical for all minicollagens and for the eight C-terminal domains of NOWA with the only variation being in the number of residues spacing the first two Cys residues. The sequence of the C-terminal minicollagen-1 MCRD, which has been investigated in the present work, is underlined and the numbering of residues corresponds to the synthetic peptides used in this study. Beside the cysteines there is only one conserved residue, which is Pro12 (shown in purple in Fig. 1B).

Peptide synthesis and oxidation

As the MCRD occurs in different molecular contexts, at the N- and C-terminal extensions of minicollagen molecules as well as eight times repeated at the C-terminus of NOWA, we speculated that it might constitute an isolated domain with the capacity of independent folding. Peptide synthesis was carried out for the C-terminal MCRD of minicollagen-1 starting with the last proline residue of the C-terminal polyproline stretch and including the charged C-terminus of the full-length molecule. The formation of disulfide bonds occurs in the presence of redox buffer at 100 μM peptide concentrations to avoid aggregation by intermolecular disulfide bonds (see Experimental Procedures). The final product showed a single peak in mass spectroscopic analysis with the reduced and oxidized MCRDs having a difference in molecular weight of 6 Da, thereby strongly indicating the formation of three intramolecular disulfide bonds (Table I). Disulfide bonds can be shown to be all-intramolecular in mass spectroscopic analysis. Analytical

ultracentrifugation confirmed the absence of significant aggregation or multimerization in solutions from 0.2 to 0.8 mM total peptide concentration (Table I), which was further supported by NMR $^{1}H^{N}$ T_2 relaxation times of more than 100 ms at 25 °C, indicative of the prevalence of a monomeric state in solutions of 1.6 mM MCRD.

The NMR Structure of the MCRD

The solution structure of the MCRD was determined independently by a group in Basel and Munich (see affiliations in the author list) from two separate peptide preparations (see Experimental Procedures). Both structures show an identical tightly packed globular fold (Figures 2 and 3), which consists of a short N-terminal α-helix between Val5 and Gln9 followed by an inverse γturn (Gln9-Val11), a type I β turn (Val11-Cys14) and a type III β turn (Pro15-Cys18). Thus cysteines 6, 10, 14 and 18 are directly located in turns whereas cysteine 2 is located in a proline-rich N-terminal sequence and cysteine 19 is oriented presumably by the β III turn and a C-terminal proline. All proline residues in the MCRD are in trans conformation as evidenced by specific NOEs.

The only conserved residue Pro12 (Fig. 1) imposes a β I turn topology on residues 11 to 14 due to its fixed φ angle of –60 °. Hydrogen bonds are established in the turns between Val5 (O) and Gln9 (H^{N}), Gln9 (O) and Val11 (H^{N}), Val11 (O) and Cys14 (H^{N}), and Pro15 (O) and Cys18 (H^{N}) (Fig. 3), respectively, as derived both from the calculated structures and from the slow $^{1}H/D$ exchange upon lyophilization of MCRD and redissolving in $^{2}H_2O$. The positively charged C-terminus Lys22, Arg23, Lys24 is flexibly disordered and does not contribute to the MCRD structure (Fig. 2).

The first disulfide bond, Cys2-Cys18, clasps the N- and C-termini of the domain while the Cys6-Cys14 bond connects the N-terminal α-helix to the type I β turn starting with Val11, thus forming the core of the MCRD structure. The third disulfide bond Cys10-Cys19 is more exposed to the C-terminal surface of the domain and constrains the polypeptide backbone into two consecutive turns. The more surface-exposed N- and C-terminal disulfide bridges represent the most likely candidates for intermolecular disulfide exchange reactions. Complete reduction of the disulfide bonds with an excess of tris(2-carboxyethyl)-phosphine

hydrochloride (TCEP) at pH 7.5 results in complete unfolding without retainment of local conformational preferences as well assessed by ^{1}H NMR.

Reoxidation of reduced MCRD with an excess of oxidized glutathione at pH 7.5 proceeds extremely fast and is completed within 2 hours as observed by ^{1}H NMR spectroscopy (Fig. 4). This is presumably due to the small domain size and the effect of Pro12 upon disulfide bridge formation (see discussion). Fluorescence spectroscopy allows monitoring of the refolding due to the quenching of Tyr17 fluorescence by a nearby disulfide bridge and yields a folding half time of 1.5 hours at pH 8 upon oxygen saturation of the solution in agreement with NMR data (not shown).

Model of trimeric minicollagen-1

It has already been mentioned that minicollagen-1 at native conditions appears to be a non-covalent trimer. A schematic representation of the trimeric minicollagen molecule including the connections of N- and C-terminal MCRDs to the polyproline type II helices is shown in Figure 5. A model of trimeric minicollagen-1 was proposed earlier (9) in which the MCRD was assumed to have a linear structure. After elucidation of the structure of the MCRD we are now able to draw a schematic model in scale with the known dimensions of the collagen triple helix and the polyproline-II helix (4). An open question is the geometry of the connections between the collagen and polyproline parts of the molecule and between MCRDs and polyprolines, respectively. The angles and flexibility at the junction sequence APLP in the N-terminus and the single Ala spacer in the C-terminus (Fig. 1, shown in black) or by any Pro residue that potentially can occur in *cis* conformation are not known. The lack of association between MCRDs even at relatively high concentrations would suggest that they also do not associate in minicollagen without the help of an isomerase or another catalytic system. Polyproline or polyhydroxyproline-II helices are also known to be monomeric. For these reasons the polyproline arms with their MCRD heads are displayed as non-interacting entities in Fig. 5. This is confirmed by previous biochemical data showing that minicollagen-1 trimers are not disulfide-crosslinked and the triple helix is stabilized solely by the collagen sequence composition (10).

The ways that MCRDs are linked to the main molecule are shown schematically (red arrows) in the upper (N-terminal) and the lower (C-terminal) part of Figure 5. The polyproline-II helices approach the MCRDs from the N-terminus for the C-terminal domain and from C-terminus for the N-terminal domain. The connections between the polyproline-II helices and the MCRD might alter the accessibility of particular disulfide bridges in the MCRD. As can be seen from comparison of Figures 3 and 5, in the N-terminal MCRD the Cys10 – Cys19 disulfide is less exposed. An opposite situation is observed in the C-terminus where the polyproline-II stretch hides the Cys2 - Cys18 bridge. Accordingly the suggested candidates for intermolecular disulfide exchange are Cys2-Cys18 for the N-terminal and Cys10-Cys19 for the C-terminal MCRD.

Discussion

Small Cys-rich domains are widely distributed building blocks of extracellular proteins in essentially all phyla including plants and bacteria. The most abundant domain type is the epidermal growth factor domain (EGF, smart00181). Including its variants EGFca (smart00179) and EGF-like (smart0001) several thousands of different EGF-domains are known in proteins of different functions (see Smart database, http://smart.embl-heidelberg.de). In most cases EGF-domains are arranged in arrays with other domains. Laminin (20) and fibrillin (21) are two out of very many examples. Many Cys-rich domains, however, also exist as single autonomous proteins and the epidermal growth factor domain is a well-known example. This fact provides a bridge to the numerous low-molar mass Cys-rich proteins, which are also products of larger precursor forms but express their function as toxins, antimicrobial peptides or other small bioactive agents. Variations of sequences, cysteine patterns and three-dimensional structures are however very large for this diverse class of small proteins and clear homologies exist for small groups only (22). Small Cys-rich peptides with antimicrobial or toxic functions are part of the innate immunity and defense system of invertebrate animals. They are often stored in secretory granules and released in response to parasites via exocytosis. Nematocyst discharge in cnidarians represents a specialized form of exocytosis from a giant post-Golgi vesicle. The appearance of a structural motif related to Cys-rich peptide toxins in proteins involved in the formation of a wall polymer, which is associated with the nematocyst membrane, might hint at a phylogenetic link between this group of defensive molecules and the evolution of the nematocyst.

The common feature of small Cys-rich domains is the prevalence of disulfide bridges in structure stabilization. This is already suggested by the large fraction of cysteines, which is 4/22 for gomesin (23), about 6/40 for EGF domains, 8/42 for crambin (24) and hellethionin (25) as compared to 6/20 for the Cys-rich domain of minicollagen-1. The later domain has one of the highest ratios of cysteines to total residues known so far. Comparably dense cysteine patterns are only described for small conotoxin peptides like PnIVA and PnIVB (26). The peptide whose structure was elucidated in the present work is 4 residues longer than the essential core. The structure shows that the first proline and the last three residues are randomly

oriented and probably not required for correct oxidative folding. The remaining structure is well defined and data demonstrate formation of a single topoisomer under optimized oxidative conditions.

In the reduced state no conformational preferences were found in MCRD by NMR. Similar observations were reported for many other Cys-rich domains underlining the importance of disulfide bonds for stabilization. As shown for tachyplesin mutational replacements of Cys- residues by alanine lead to a loss of structure (27). Interestingly the global fold of tachyplesin was rescued by hydrophobic interactions between pairs of tyrosines or phenylalanines if cysteines were replaced by these residues (27).

Although the equilibrium structure of MCRD is predominantly determined by the covalent interactions between Cys residues kinetic intermediates must be responsible for correct disulfide pairings. A folded-precursor mechanism (28,29,36) and a quasi-stochastic mechanism (30,31) were proposed for other monomeric proteins. True pathways are probably between these two extremes. For the Cys-rich domain of minicollagen formation of a type I β turn by Pro12 and hydrogen bonding between Cys14 (H^N) and Val11 (O) is most likely a very important intermediate step. As can be seen from the structure (Figs. 2 and 3) this bend and the βturn bring the cysteines into close vicinity for correct intramolecular disulfide bridging.

It is amazing that few non-covalent interactions in the MCRD of minicollagen and in many other short peptides lead to correct connectivities between several Cys residues, which statistically could also interact in very many different intra- and intermolecular modes. To establish the disulfide connections of MCRD pairwise bridges between adjacent Cys-residues have to be prevented. The three intervening amino acid residues found in MCRDs render pair-wise bridges unfavorable (32) and this feature is most likely essential for proper folding.

In small disulfide rich proteins identical folds are found in homologous domains in contrast to large variations in non-cysteine residues. A striking example are the about 200 solved structures of different EGF-domains, which exhibit the same fold although sequence identities of non-cysteine residues are difficult to detect (smart, key words EGF and structure). This high conservation of structure leads us to assume, that the related Cys-rich repeats in minicollagen of corals and in NOWA proteins have the same global fold as the MCRD of minicollagen-1 of *Hydra*, which was investigated in the present work.

Conservation of residues is low (Fig. 1) and would not allow the prediction of structure identities for conventional globular proteins (33)

Earlier data on the interaction of minicollagens and NOWA in the nematocyst wall conclusively demonstrated a switch from a soluble state of both proteins with all disulfides internally linked to an insoluble dense polymeric state with intermolecular disulfide bonds (see introduction). Analogous assembly processes involving disulfide reshuffling were reported for virus capsid proteins (19), collagen IV (34) and other systems (17,18). The only cysteines in minicollagen-1 are located in the MCRDs. Consequently a disulfide reshuffling interaction was proposed between these domains. The data also suggested corresponding interactions between the Cys-rich domains of minicollagen and NOWA. *In vivo* the situation may be rather complex because of possible interactions between N- and C-terminal domains of different minicollagens in the same organism. At each end of a minicollagen three MCRDs are present (Fig. 4). This opens the possibility of a simultaneous interaction between minicollagens and NOWA.

Extensive searches of the database failed to reveal MCRD-like domains in non-nematocyst proteins. The domain type seems to be highly specialized for disulfide mediated crosslinking of the nematocyst wall. For several organisms with nematocyst organelles minicollagen and NOWA genes have been found but sequencing of cnidaria genes is still too incomplete to allow a general conclusion. Interestingly in minicollagen of corals (AF507373.1) Cys-residues frequently occur after glycines in the collagenous part of the sequence. Cysteine residues, which interlink collagen chains of the same molecule are located at the ends of a triple helix to form a stable disulfide knot (35,36) or located in interruptions of the regular Gly-X-Y repeat (37). Cys-residues in X-position of collagens are highly unusual and are most likely used to link between collagen molecules. It may be speculated that the nematocyst walls of corals are cross-linked by disulfide bridges between collagen triple helices in addition to those between MCRDs thus providing a higher tensile strength.

Future work will focus on the mechanism of the disulfide reshuffling process using isolated Cys-rich domains of minicollagen and NOWA. A likely possibility is the interaction between N-and C-terminal domains of minicollagens. *In vivo*, the disulfide exchange reaction probably is catalyzed by disulfide

isomerases or other helper proteins, which have to be defined prior to *in vitro* studies. Our working hypothesis is that only one or two out of the three disulfide bonds will be reshuffled from intra to intermolecular. Likely candidates are the bonds Cys2-Cys18 and Cys10-Cys19, which are both surface exposed.

Acknowledgements

We thank Eric Steel for expert technical assistance.

References

1. Vollmer, S. V., and Palumbi, S. R. (2002) *Science* **296**, 2023-2025
2. Wang, W., Omori, M., Hayashibara, T., Shimoike, K., Hatta, M., Sugiyama, T., and Fujisawa, T. (1995) *Gene* **152**, 195-200.
3. Kurz, E. M., Holstein, T. W., Petri, B. M., Engel, J., and David, C. N. (1991) *J Cell Biol* **115**, 1159-1169.
4. Bateman, J. F., Lamande,S.R. and Ramshaw,J.A.M. (1996) *Collagen superfamily*. Extracellular Matrix. Molecular Components and Interactions (Comper, W. D., Ed.), 2, Harwood Academic Publisher, Amsterdam
5. Engel, J. (1997) *Science* **277**, 1785-1786
6. van der Rest, M., and Garrone, R. (1991) *Faseb J* **5**, 2814-2823
7. Ricard-Blum, S., Dublet, B., and van der Rest, M. (2000) *Unconventional Collagens*, Oxford University Press, New York
8. Holstein, T., and Tardent, P. (1984) *Science* **223**, 830-833.
9. Engel, U., Pertz, O., Fauser, C., Engel, J., David, C. N., and Holstein, T. W. (2001) *Embo J* **20**, 3063-3073.
10. Özbek, S., Pertz, O., Schwager, M., Lustig, A., Holstein, T., and Engel, J. (2002) *J Biol Chem* **277**, 49200-49204
11. Engel, U., Özbek, S., Engel, R., Petri, B., Lottspeich, F., and Holstein, T. W. (2002) *J. Cell Sci.* **115**
12. Lenhoff, H. M., Kline, E. S., and Hurley, R. (1957) *Biochim. Biophys. Acta* **26**, 204-205
13. Blanquet, R., and Lenhoff, H. M. (1966) *Science* **154**, 152-153.
14. Özbek, S., Engel, U., and Engel, J. (2002) *J Struct Biol* **137**, 11-14
15. Mullen, G. E., Kennedy, M. N., Visintin, A., Mazzoni, A., Leifer, C. A., Davies, D. R., and Segal, D. M. (2003) *Proc Natl Acad Sci U S A* **100**, 3919-3924
16. Wieringa, R., De Vries, A. A., Post, S. M., and Rottier, P. J. (2003) *J Virol* **77**, 12996-13004
17. Knaus, K. J., Morillas, M., Swietnicki, W., Malone, M., Surewicz, W. K., and Yee, V. C. (2001) *Nat Struct Biol* **8**, 770-774.
18. Wagner, D. D., Lawrence, S. O., Ohlsson-Wilhelm, B. M., Fay, P. J., and Marder, V. J. (1987) *Blood* **69**, 27-32.
19. Li, P. P., Nakanishi, A., Clark, S. W., and Kasamatsu, H. (2002) *Proc Natl Acad Sci U S A* **99**, 1353-1358
20. Engel, J. (1992) *Biochemistry* **31**, 10643-10651
21. Handford, P. A. (2000) *Biochim Biophys Acta* **1498**, 84-90

22. Tamaoki, H., Miura, R., Kusunoki, M., Kyogoku, Y., Kobayashi, Y., and Moroder, L. (1998) *Protein Eng* **11**, 649-659
23. Mandard, N., Bulet, P., Caille, A., Daffre, S., and Vovelle, F. (2002) *Eur J Biochem* **269**, 1190-1198
24. Housset, D., Benabicha, F., Pichon-Pesme, V., Jelsch, C., Maierhofer, A., David, S., Fontecilla-Camps, J. C., and Lecomte, C. (2000) *Acta Crystallogr D Biol Crystallogr* **56** (Pt 2), 151-160
25. Milbradt, A. G., Kerek, F., Moroder, L., and Renner, C. (2003) *Biochemistry* **42**, 2404-2411
26. Fainzilber, M., Nakamura, T., Gaathon, A., Lodder, J. C., Kits, K. S., Burlingame, A. L., and Zlotkin, E. (1995) *Biochemistry* **34**, 8649-8656
27. Laederach, A., Andreotti, A. H., and Fulton, D. B. (2002) *Biochemistry* **41**, 12359-12368
28. Boudko, S. P., and Engel, J. (2004) *J Mol Biol* **335**, 1289-1297
29. Welker, E., Wedemeyer, W. J., Narayan, M., and Scheraga, H. A. (2001) *Biochemistry* **40**, 9059-9064
30. Creighton, T. E. (1997) *Biol Chem* **378**, 731-744
31. Weissman, J. S., and Kim, P. S. (1991) *Science* **253**, 1386-1393
32. Zhang, R. M., and Snyder, G. H. (1989) *J Biol Chem* **264**, 18472-18479
33. Doolittle, R. F. (1995) *Annu Rev Biochem* **64**, 287-314
34. Gunwar, S., Ballester, F., Noelken, M. E., Sado, Y., Ninomiya, Y., and Hudson, B. G. (1998) *J Biol Chem* **273**, 8767-8775
35. Bruckner, P., Bächinger, H. P., Timpl, R., and Engel, J. (1978) *Eur J Biochem* **90**, 595-603
36. Barth, D., Kyrieleis, O., Frank, S., Renner, C., and Moroder, L. (2003) *Chemistry* **9**, 3703-3714
37. Kühn, K. (1995) *Matrix Biol* **14**, 439-445
38. Wüthrich, K. (1986) *NMR of proteins and nucleic acids*, Wiley, New York
39. Jeener, J., Meier, B. H., Bachman, P., and Ernst, R. R. J. (1979) *Chem. Phys.* **71**, 4546-4553
40. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *Journal of Biomolecular NMR* **6**, 277-293
41. Garrett, D. S., Powers, R., Gronenborn, A.M., and Clore, G.M. (1991) *Journal of Magnetic Resonance* 95, 214-220
42. Brunger, A. T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M.,

Simonson, T., and Warren, G.L. (1998) *Acta Crystallographica Section D-Biological Crystallography* **54**, 905-921

43. Koradi, R., Billeter, M., and Wüthrich, K. (1996) *J Mol Graph* **14**, 51-55, 29-32

Figure legends

Figure 1. *A*, minicollagen-1 amino acid sequence and domain structure. Propeptide sequence is in blue, MCRDs in red, polyproline sequences in light green, collagen repeat in dark green. *B*, Alignment of MCRDs in minicollagen molecules from different cnidarians and in NOWA. MCol1h: minicollagen-1 *Hydra*, MCol2h: minicollagen-2 *Hydra* (3), MColad: minicollagen *Acropora donei* (2), MColac: minicollagen *Acropora cervicornus* (1), MColap: minicollagen *Acropora palmate* (1), N, C: N-, C-terminal, The sequence of the Cys-rich region of NOWA in *Hydra* (NWh) starts with repeat 1 and terminates with repeat 8. Residues in MCol1hC are numbered starting at proline preceding the first cysteine and the same numbering was used in the NMR structures. The highly conserved cysteine residues are marked in red. Proline in position 12, which is conserved with two exceptions, is marked in purple. The sequence of the synthesized and investigated peptide is underlined.

Figure 2. NMR structures of the C-terminal MCRD of minicollagen-1. The backbone is displayed in blue, side chains in red except for cysteine side chains in yellow. *Structure 1 (left,* accession number: 1SOP*):* the ten energy lowest structures are superimposed using the backbone atoms of residues 2-19 with an rmsd of 0.6 Å. Disordered residues 21-24 are depicted in grey. *Structure 2 (right*, accession number: 1SP7*),* all-heavy-atom representation of the 20 lowest energy conformers of the minicollagen-1 C-terminal MCRD out of 100 structures calculated in CNS (42), with best fit for backbone heavy atoms of residues 2-21. The charged C-terminal sequence $K^{22}R^{23}K^{24}$ (grey) is disordered at pH 6.5 in the absence of additional salt.

Figure 3. Stereoview of the same conformer as in Figure 2 in ribbon presentation. Only structure 2 is shown. Cysteines and disulfide bonds are shown in yellow and Pro12 is in aquamarine. Disulfide bonds (Cys2-Cys18, Cys6-Cys14 and Cys10-Cys19) as well as the conserved Pro12 are indicated in ball and stick representation. Dashed lines indicate the hydrogen bonds identified by MOLMOL. These are bonds between Pro5 (O) and Gln9 (H^N); Val11 (O) and Cys14 (H^N); Pro15 (O) and Cys18 (H^N). The first two H-

bonds belong to the □–helix, others belong to the three different turns. Pro12 induces a □ I turn from residues 11 to 14. All proline residues in the structure can be shown to be in *trans* conformation from experimental NOE distance information.

Figure 4. Folding kinetics of the MCRD as monitored by NMR. The intensity of the amide resonance signal of Val5 in folded MCRD is shown upon reoxidizing fully reduced MCRD with an excess of oxidized glutathione. The reoxidation reaction is completed within 2 hours at pH 7.5 when catalyzed by the glutathione system. The unusually fast reoxidation is in agreement with data from fluorescence experiments (data not shown).

Figure 5. Schematic presentation of the minicollagen-1 trimer. Full-length minicollagen-1 (central part of the figure) and connections of polyproline regions to MCRDs (upper and lower parts of the figure) are shown. Yellow circles and ribbon globular structures are MCRDs; blue lines and blue ribbons – polyproline type II helices (PP); red lines – collagen triple helix. Basic tails of the C-terminal MCRDs are represented in faint curves. In the upper part only the MCRD core (2-21 residues) is shown. The connection to the N-terminal 23 residue polyproline helix is indicated by a red arrow.

Figure 1

A

Minicollagen-1

propeptide N-Cys-rich poly-Pro

KTLHEMLKRDANPCGSYCPSVCAPACAPVCCYPPPPPPPPPPPPPPPPPPPPPAPLP

collagenous poly-Pro C-Cys-rich basic tail

GNPGPPGRPGPPGAPGPAGPPGLPGPPGPPGAPGQGGLPGQPAPPPPPCPPVCVAQCVPTCPQYCCPAKRK

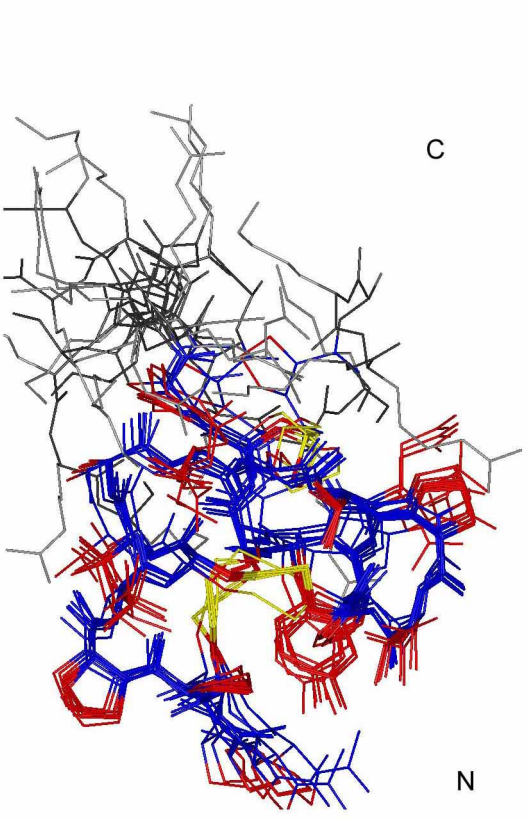
B

N- and C- terminal Cys-rich repeats of minicollagens (MCol) and NOWA (NW)

```
                1          10         20
MCol1hN     DANPCGS    YCPSVCAPACAPVCCYPPPPPPPPPPPPPPPP
MCol2hN     SAQACGY    NCPAICAPACTPICCAPPPPPPPPPPPPPPPP
MColadN        CGY     GCPSMCAPACEPTCCAPPPPPPPPP
MColapN     REASPCGY   GCPSMCAPACEPTCCAPPPPPPPPP
MColacN     REASPCGY   GCPSMCAPACEPTCCAPPPPPPPPP
MCol1hC     PPPPPCPP   VCVAQCVPTCPQYCCPAKRK
MCol2hC     PPPPPCPP   ICPTQCVPYCPQYCCLKGPPGPP
MColadC      PPPPCPP   ICIQHCIRICPQPCCSPPPPP

NWhREP1     QITGTCPS    GCSGDCYPECKPGCCGQVNLNAPVQP
NWhREP2     SGYTACSQ YPNCGLSCQSSCSQSCCQQNPYQPSVMSGTIVIQP
NWhREP3     NEQSVCPQ HPGCSQHCAPRCSPQCCQQSMNSLYQP
NWhREP4     PQMSACPQ FPSCSPTCAPQCSQLCCQQSSMPLQM
NWhREP5     PQMPSCPQ FPSCSASCAPQCSQQCCQQPSMSIQP
NWhREP6     LOISSCPO FPSCSPSCAPOCSOOCCOOPSMPIOL
NWhREP7     PLMGSCSQ MPGCSASCAPLCSQQCCQQQAMLQQSIMQQPMM
NWhREP8     MAQNPCSLQQPGCSSACAPACRLSCCSLGRMNLGR
```

Figure 2



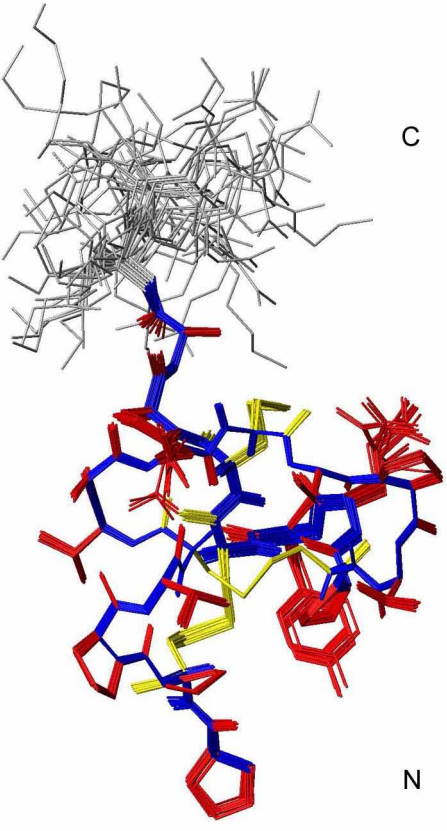


Figure 3

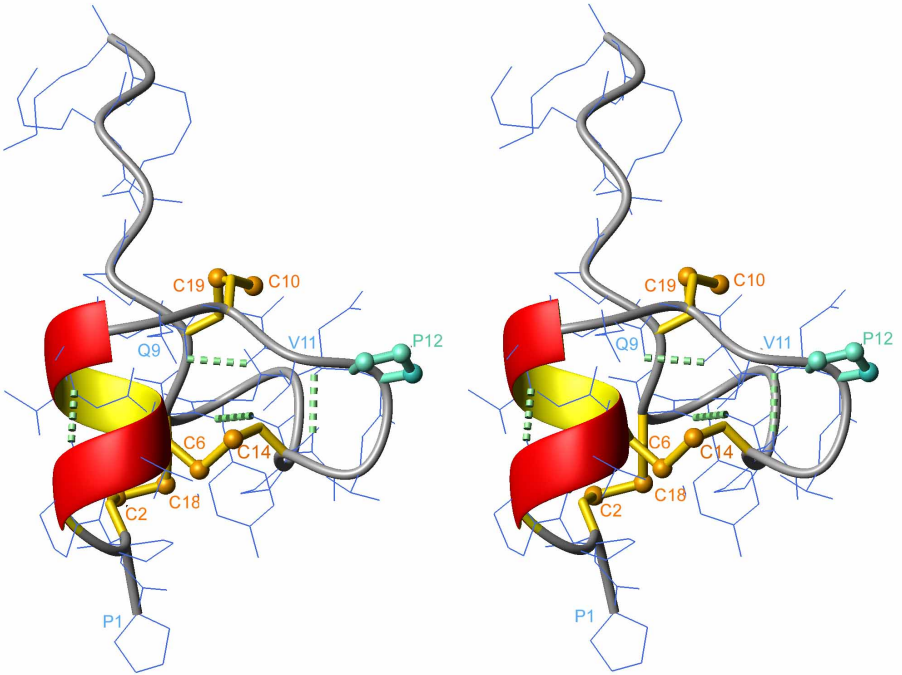

C19
C10
Q9
V11
P12
C6
C14
C2
C18
P1
C19
C10
Q9
V11
P12
C6
C14
C2
C18
P1

Figure 4

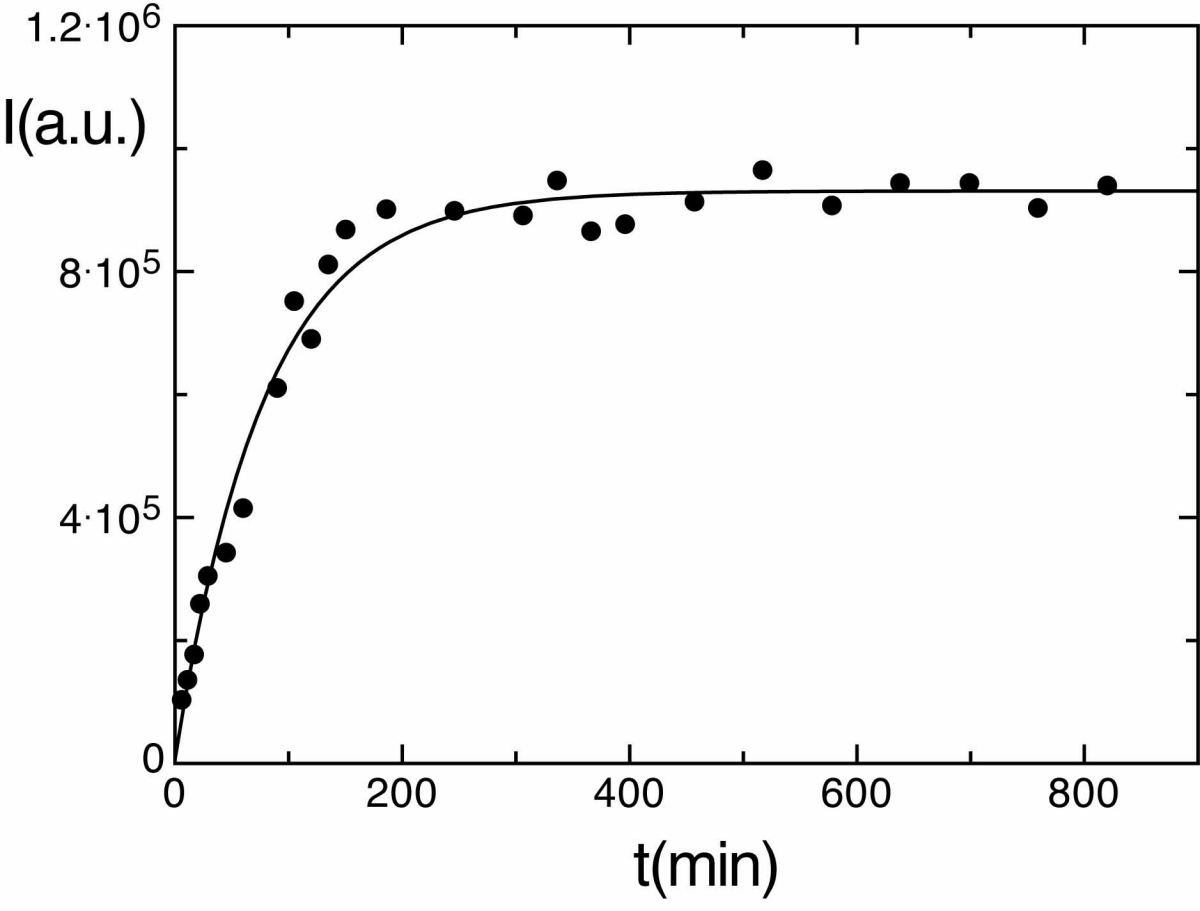
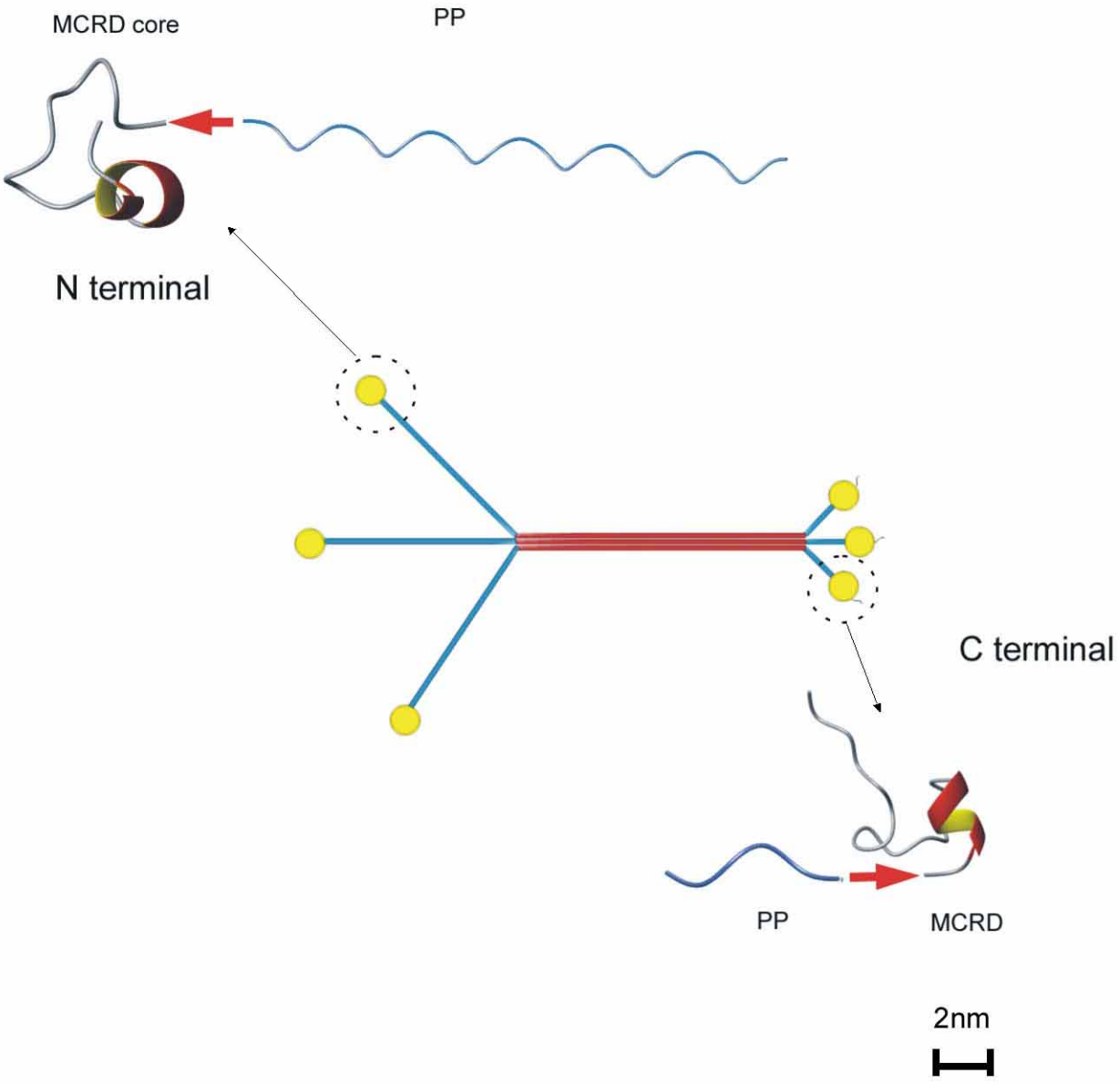


Figure 5



Chapter 5
Assignment of protein tyrosine phosphatase 1B (PTP1B)

Over the last two decades, it has become clear that the reversible phosphorylation of tyrosine residues is central to cellular signaling in response to extracellular stimuli. Responses in the cell include metabolism, cell growth, proliferation, differentiation and migration (Tonks, 2003), both under normal and pathological conditions. Disturbance of the normal homeostasis of tyrosine phosphorylation and dephosphorylation is implicated in several human diseases. The homeostasis is achieved by the concerted action of various protein tyrosine kinases (PTKs) and protein tyrosine phosphatases (PTPs).

With the advent of whole-genome sequence information, PTPs have been recognized as a diverse family (Figure 5–1) with more than 110 identified vertebrate catalytic domains in the GenBank (http://www.ncbi.nlm.nih.gov/Genbank) and 38 localized PTP sequences in the human genome (Andersen et al., 2004) (http://ptp.cshl.edu). Protein tyrosine phosphatases among others include cdc25A (responsible for cell cycle progression), CD45 from B- and T-cells, the viral minimal variant VH1, YOP 2b from the *Yersinia* genus of bacteria (causing e.g. Black Death) (Walton and Dixon, 1993; Zhang, 2002) and PTP1B, which is the prototypic PTP first isolated from human placenta in 1988 (Tonks et al., 1988).

Due to the central role of PTPs in cell homeostasis and metabolism, their transcription and translation are tightly regulated (Andersen et al., 2004). Furthermore, the PTP1B protein activity is reversibly modulated by the phosphorylation of serine, threonine and tyrosine residues (Flint et al., 1993; Tao et al., 2001) and by redox regulation of the cysteine in the active center (Salmeen et al., 2003).

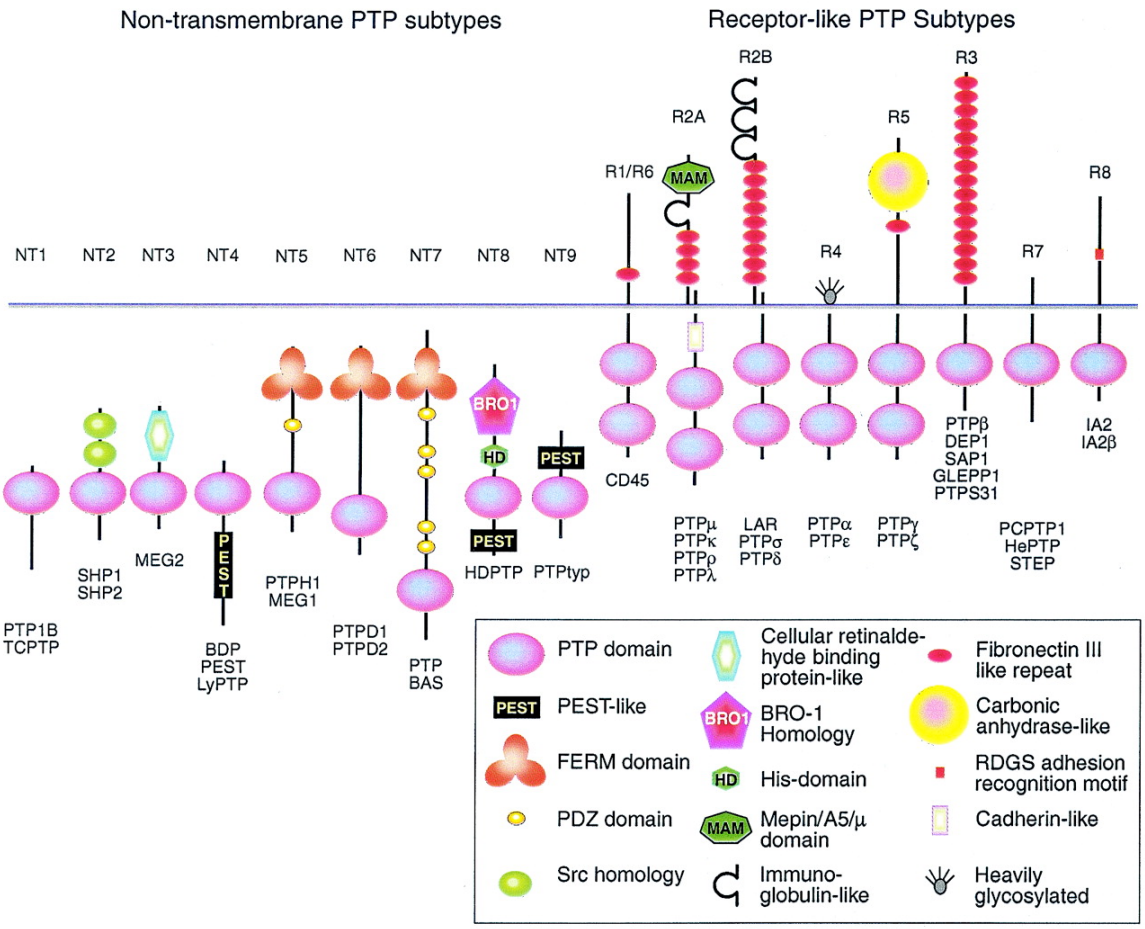


Figure 5–1. Domain organization of PTPs. The central, conserved catalytic domain is termed PTP domain (taken from (Andersen et al., 2001)). PTP1B is shown on the left.

As PTKs exist in mutated oncogenic forms, PTPs as their antagonists are interesting due to their potential capacity as tumor suppressors. PTP1B has been shown to suppress oncogenic transformation and to act on cell adhesion due to the regulation of integrin and cadherin-catenin signaling (Tonks, 2003).

PTP1B dephosphorylates receptor tyrosine kinases. *In vivo* targets of PTP1B are the insulin receptor and growth factor receptors (epidermal growth factor receptor and insulin-like growth factor receptor) (Haj et al., 2002; Tonks, 2003). Ablation of the *PTP1B* gene in mice (Elchebly et al., 1999) leads to increased insulin sensitivity. In result, serum glucose levels are decreased, concurrent with increased insulin-receptor phosphorylation in muscle and liver. Surprisingly, the knockout additionally leads to resistance against obesity in a high fat diet, yielding decreased adipocyte volumes in fat tissue. Very importantly, *PTP1B* knockout mice do not show an increased predisposition to cancer. These data strongly suggest that specific inhibition of

PTP1B could be an effective strategy against obesity and type II diabetes, which are increasing health risks and cost factors at the current lifestyle in developed countries.

All PTPs share a conserved phosphatase domain (PTP domain, see Figure 5–1) with a signature motif [I/V]HCXXGXXR[S/T] around the catalytic cysteine residue 215, which is central to the phosphotyrosine-ester hydrolysis by nucleophilic attack of the sulfhydryl group. The PTP1B catalytic domain consists of 12 β-strands and 8 α-helices (Barford et al., 1994), as determined by X-ray crystallography. A central 10 stranded β-sheets forms the structural scaffold of the protein. The active site around C215 sits in a crevice near the surface between β-12 and α-4 and is highly reactive. Conserved residues in the PTP catalytic domain cluster into 10 motives at the helical face of the protein (Figure 5–2). Resulting from the quest for specific inhibition of PTP1B, there are furthermore more than 40 PTP1B complex structures with ligands in the PDB. Nonetheless, inhibitory drugs against PTP1B have not reached the market. Compounds have been withdrawn from clinical trials due to unsatisfactory results and side effects, or have shown insufficient specificity *in vivo* (Tonks, 2003).

NMR is a widely established method to detect and quantify specific intermolecular interactions, even at weak affinities. This is exploited for example to screen chemical libraries for binding to protein targets in the pharmaceutical industry. The screening depends on the sequence specific backbone assignment to map lead binding onto the protein structure, if available (Hajduk et al., 1999).

We have obtained the backbone ^{15}N, $^{1}H^{N}$, $^{13}C'$ and $^{13}C^{\alpha,\beta}$ assignments of PTP1B (1-298) using standard and TROSY (Pervushin et al., 1997; Salzmann et al., 1999) HNCA, HN(CO)CA, HNCACB, HN(CO)CACB, HNCO and HN(CA)CO (Bax and Grzesiek, 1993) as well as ^{15}N NOESY spectra. The U–^{13}C, U–^{15}N, U–^{2}H sample of the catalytic domain of PTP1B was prepared by collaborators from Novartis (for involved collaborators see (Meier et al., 2002b)). Due to the reactivity of the active site cysteine, spectra were obtained from a flame-sealed sample under nitrogen atmosphere, with the Novartis inhibitor CGS35385 (DADEXLIP-amide, X = phenylalanine p-difluoromethyl phosphonate) as a substrate analogue bound to the protein. In addition, perdeuterated DTT was used to avoid protein association via disulfide bridges formed by the highly reactive cysteine in the active center.

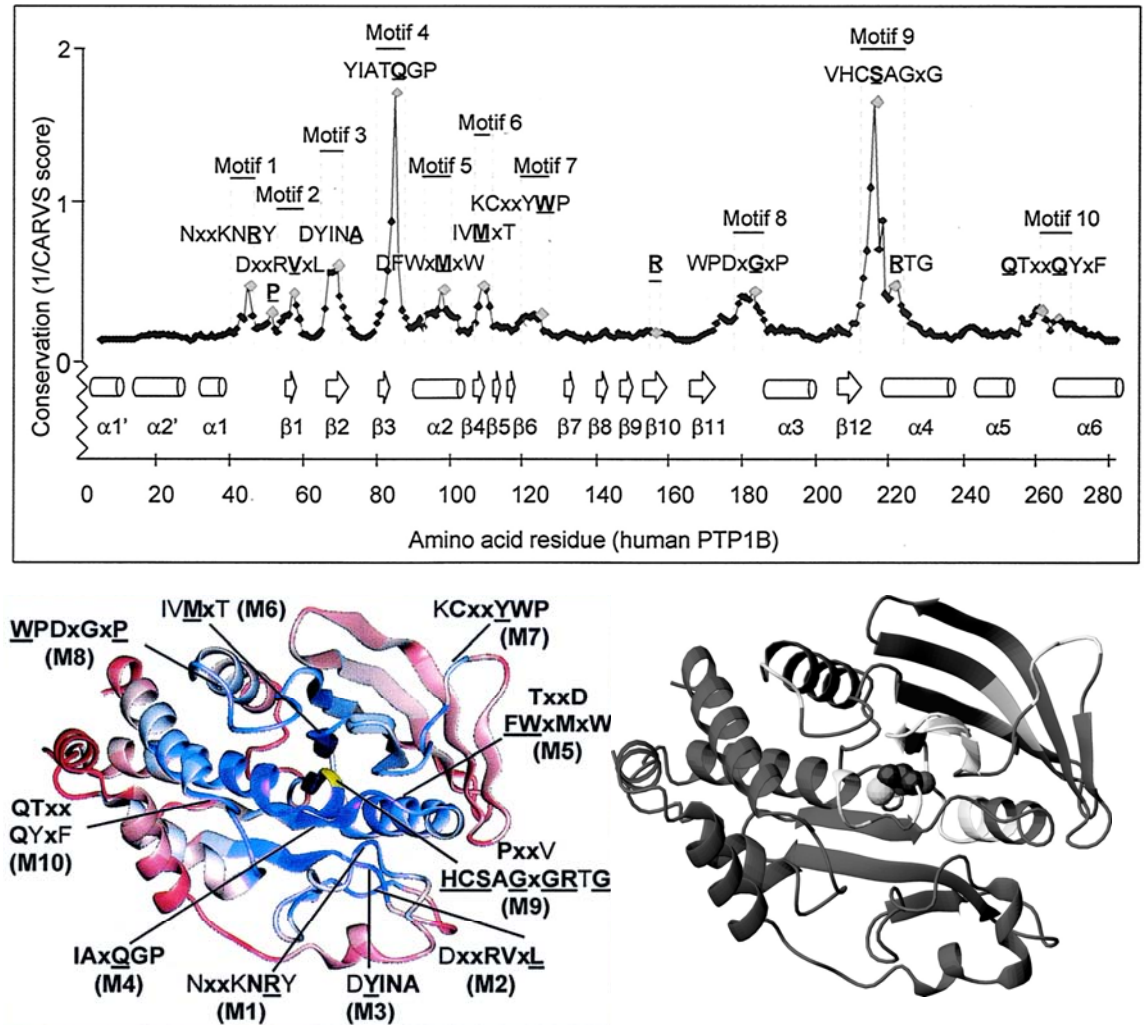


Figure 5–2. Top: Conservation of residues in the PTP catalytic domain (from (Andersen et al., 2001)). Bottom: Conserved motives (left) and resonance assignments (right) mapped onto the PTP crystal structure (PDB code 2HNP) (Barford et al., 1994). Unassigned residues are painted white in the ribbon and cluster in loops and β-sheet regions near the active site (C215, shown as spacefill).

Secondary chemical shifts agree with the secondary structure pattern in the crystal structure and thus support the assignment (Figure 5–3). Missing assignments are due to line broadening beyond detection in the NMR spectra and cluster in the structure near and in the active site, protruding to the β-sheet face of the protein. Nearly all conserved residues, in contrast, lie on the α-helical face (compare Figure 5–2). Almost all of the residues relevant for substrate binding and turnover could thus be assigned (Y46, R47, D48, D181, F182, R254, R257, Q262 and Q266).

Line broadening beyond detection indicates μs- to ms-dynamics of the unassigned residues. PTP1B has been known to undergo large conformational changes due to an

induced fit mechanism of substrate binding and in the course of redox regulation. The dynamics observed in the NMR spectra reflects this mobility of the active site.

The recorded spectra partially suffered from the relatively poor deuteration (~60 %) of the PTP1B sample. Further investigations on the disordering of non-assigned residues, e.g. with recently devised NMR spectra incorporating CPMG transfer methods (Loria et al., 1999; Wang et al., 2001), were therefore not pursued. The assignments are deposited in the BMRB under accession number 5474 and have been published in the Journal of Biomolecular NMR (Meier et al., 2002b).

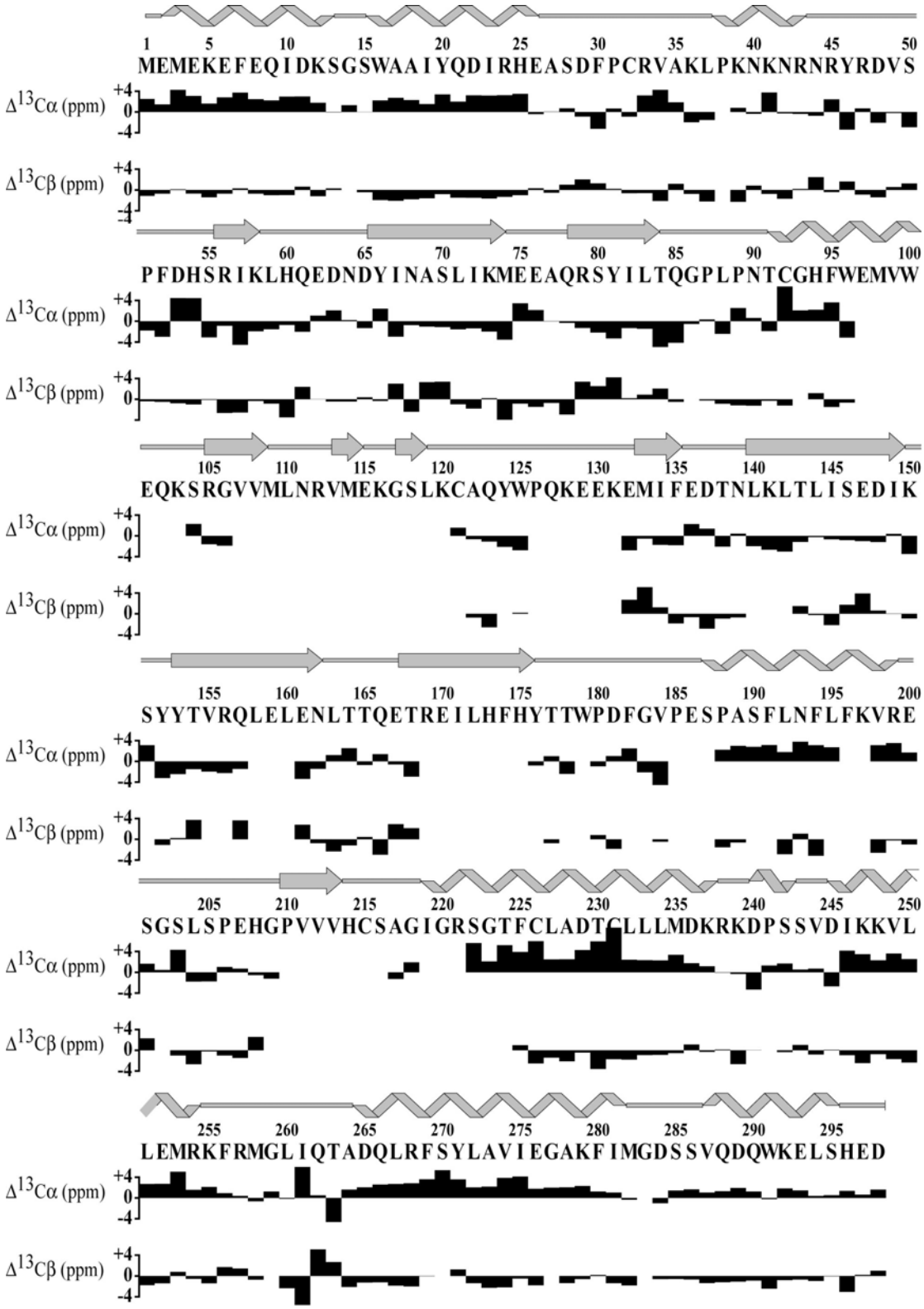


Figure 5–3. Secondary structure and secondary chemical shifts match in the PTP1B backbone assignment. Random coil shifts are taken from (Spera and Bax, 1991).

Original Publication

Meier S., Li Y.C., Koehn J., Vlattas I., Wareing J., Jahnke W.G., Wennogle L.P., Grzesiek S.

Letter to the Editor: Backbone resonance assignment of the 298 amino acid catalytic domain of protein tyrosine phosphatase 1B (PTP1B).

Journal of Biomolecular NMR **24**, 165-166 (2002).

Journal of Biomolecular NMR, **24:** 165–166, 2002.
KLUWER/ESCOM

Letter to the Editor: Backbone resonance assignment of the 298 amino acid catalytic domain of protein tyrosine phosphatase 1B (PTP1B)

Sebastian Meier[a], Yu-Chin Li[b], James Koehn[b], Isidoros Vlattas[c], James Wareing[c], Wolfgang Jahnke[d], Lawrence P. Wennogle[c] & Stephan Grzesiek[a,*]
[a]*Division of Structural Biology and Division of Biophysical Chemistry, Biozentrum, University of Basel, 4056 Basel, Switzerland;* [b]*Core Technology Department and* [c]*Novartis Institute for Biomedical Research, Novartis Pharmaceuticals Corporation, 556 Morris Avenue, Summit, NJ 07901-1398, U.S.A.;* [d]*Core Technologies, Novartis Pharma AG, CH-4002 Basel, Switzerland*

Received 26 July 2002; Accepted 13 August 2002

Key words: diabetes, heteronuclear NMR, obesity

Biological context

Protein activation via tyrosine phosphorylation by tyrosine kinases and the concomitant deactivation/dephosphorylation by protein tyrosine phosphatases (PTPs) plays a central role in cellular signaling. Tyrosine phosphorylation regulates cell homeostasis by the control of gene-transcription, cell-cycle progression, cell growth, cell metabolism, immune response as well as programmed cell death (Walton et al., 1993; Zhang, 2002). Due to the low level of cellular phosphotyrosine these pathways and the respective enzymes were identified only rather recently: the first tyrosine kinase src by Hunter et al. (1980), the first protein tyrosine phosphatases by Nelson and Branton (1984). Receptor tyrosine kinases include the insulin receptor and a large variety of other growth factor receptors. Examples of PTPs include cdc25A responsible for cell cycle progression, CD45 activating B- and T-cells, the Vaccinia virus minimal variant VH1, YOP 2b required for the virulence of the pathogenic *Yersinia* genus of bacteria, and PTP1B, which is a potential target for diabetes and obesity treatment.

PTP1B can dephosphorylate a number of tyrosine kinases *in vitro* and in cell culture at the endoplasmic reticulum (ER) (Haj et al., 2002). *In vivo* specificity of this phosphatase for insulin receptor dephosphorylation was suggested by the finding that the active site mutant C215S copurifies with insulin receptors from eucaryotic cell lines (Seely et al., 1996; Bandyopadhyay et al., 1997). Association with the activated insulin receptor is mediated by three tyrosine residues in the catalytic domain that become phosphorylated upon insulin treatment (Bandyopadhyay et al., 1997). The importance of PTP1B in the regulation of glucose metabolism was underlined by the finding that homozygous PTP1B knockout in mice leads to strongly increased blood glucose clearance and hypersensitivity to insulin (Elchebly et al., 1999). Additionally, homo- and heterozygous knockouts proved to decrease weight gain by 50% on high fat diet. For these reasons, PTP1B appears as an attractive target for specific inhibitors that would counteract type II diabetes and obesity.

Full length human PTP1B consists of 435 residues. It encompasses the highly conserved 298 amino acid, N-terminal catalytic domain ($PTP1B^{cat}$), a proline rich domain and a C-terminal hydrophobic anchor that targets the protein to the membrane of the ER (Frangioni et al., 1992; Haj et al., 2002). The crystal structure of the $PTP1B^{cat}$ construct 1-321 (Barford et al., 1994) shows that this domain consists of 12 β-strands and 8 α-helices. The catalytic residue C215 is located near the surface in a crevice between β12 and α4. Here we report the NMR resonance assignments of the conserved 298 amino acid $PTP1B^{cat}$ domain in complex with the peptidic Novartis inhibitor CGS35385.

Methods and experiments

Human PTP-1B (1-298) was cloned from a human hippocampal cDNA library (Clonetech) using PCR and placed into a pET 19-b vector (Novagen). Re-

*To whom correspondence should be addressed. E-mail: Stephan.Grzesiek@unibas.ch

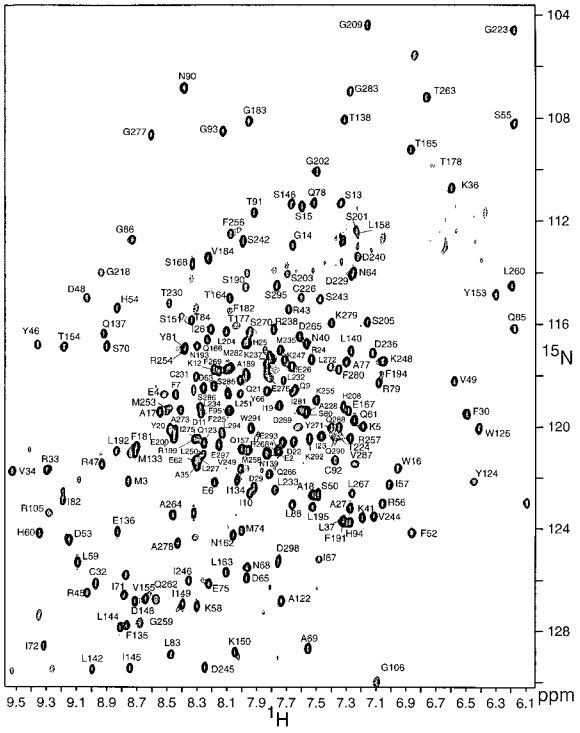


Figure 1. ^{1}H-^{15}N TROSY spectrum of PTP1Bcat at 293 K.

combinant, uniformly ^{2}H (∼60%), ^{15}N (>95%), ^{13}C (>95%) labeled PTP1B^{cat} was expressed in *E. coli* strain BL21 (DE3), and the soluble protein was purified by cation followed by anion exchange chromatography. A 500 µl NMR sample was prepared containing 0.5 mM PTP1B^{cat}, 1 mM peptidic inhibitor CGS35385 (DADEXLIP-amide, X = phenylalanine-p-difluoromethyl phosphonate), 25 mM TRIS-d11, 10 mM DL-1,4-DTT-d10, 0.02% sodium azide, 50 mM NaCl, 95% H_2O/5% D_2O at pH 7.5. The NMR tube was flame-sealed under nitrogen. PTP1Bcat tends to aggregate at concentrations higher than 0.5 mM and temperatures above 293 K. Peak dispersion was improved at pH 7.5 relative to pH 6.5. All spectra were recorded on Bruker DRX 600 and 800 spectrometers at 293 K. Assignments were derived from standard and TROSY versions of HN(CO)CA, HN(CA)CO, HN(CO)CACB, HNCACB, HNCA, HNCO, and ^{15}N-edited NOESY experiments. ^{1}H, ^{15}N, and ^{13}C chemical shifts were referenced relative to the frequency of the ^{2}H lock resonance of water.

Extent of assignment and data deposition

The assignments comprise 78% of all ^{1}HN, ^{15}N, ^{13}C′, $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ resonances covering 244 of the 285 non-proline residues. Due to intermediate conformational exchange, almost all of the resonances for the remaining residues were broadened beyond detection in the three-dimensional spectra. A map of the missing assignments onto the crystal structure reveals that most of these residues surround the active site whereas a few others are located in turn and loop regions. Apparently, the active site undergoes conformational exchange despite the presence of the high affinity inhibitor CGS35385 (IC_{50} = 100 nM). Very likely, this apparent plasticity of the active site is linked to the ability of PTP1B to accommodate a large variety of pTyr-containing substrates and the large structural rearrangements of the active site observed upon ligand binding (Zhang, 2002). Despite the difficulty in observing the active site, the assignments include at least 9 key residues (Y46, R47, D48, D181, F182, R254, R257, Q262, Q266) involved in substrate binding and turnover (Zhang, 2002). Chemical shifts have been deposited in the BMR data bank under accession number 5474.

Acknowledgments

We would like to thank Iou-Iou Sytwu and James Dellureficio for the synthesis of the CGS35385 inhibitor. This work was supported by SNF grant 3100-061757.00.

References

Bandyopadhyay, D., Kusari, A., Kenner, K.A., Liu, F., Chernoff, J., Gustafson, T.A. and Kusari, J. (1997) *J. Biol. Chem.*, **272**, 1639–1645.

Barford, D., Flint, A.J. and Tonks, N.K. (1994) *Science*, **263**, 1397–1404.

Elchebly, M., Payette, P., Michaliszyn, E., Cromlish, W., Collins, S., Loy, A.L., Normandin, D., Cheng, A., Himms-Hagen, J., Chan, C.C., Ramachandran, C., Gresser, M.J., Tremblay, M.L. and Kennedy, B.P. (1999) *Science*, **283**, 1544–1548.

Frangioni, J.V., Beahm, P.H., Shifrin, V., Jost, C.A. and Neel, B.G. (1992) *Cell*, **68**, 545–560.

Haj, F.G., Verveer, P.J., Squire, A., Neel, B.G. and Bastiaens, P.I. (2002) *Science*, **295**, 1708–1711.

Hunter, T. and Sefton, B.M. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 1311–1315.

Nelson, R.L. and Branton, P.E. (1984) *Mol. Cell Biol.*, **4**, 1003–1012.

Seely, B.L., Staubs, P.A., Reichart, D.R., Berhanu, P., Milarski, K.L., Saltiel, A.R., Kusari, J. and Olefsky, J.M. (1996) *Diabetes*, **45**, 1379–1385.

Walton, K.M. and Dixon, J.E. (1993) *Annu. Rev. Biochem.*, **62**, 101–120.

Zhang, Z.Y. (2002) *Annu. Rev. Pharmacol. Toxicol.*, **42**, 209–234.

Chapter 6
Conclusions and context

In this thesis, we have made use of the distinctive features of residual dipolar couplings to gain insights into relative orientations of protons and heavy nuclei, into local structures of unfolded states and during folding and into order and dynamics on a large time scale, which are all difficult to obtain with other structural methods.

The thesis describes new alignment media and new NMR techniques for the measurement of residual dipolar couplings (I). Weak alignment techniques and other NMR methods have been applied in studying the structure, stability and secondary structure formation of the fibritin folding nucleus (II), the oxidative folding and structural biology of a minicollagen cysteine rich domain (III) and the human protein tyrosine phosphatase PTP1B (IV).

I.) The measurement of RDCs has recently been introduced into the field of protein folding to monitor the long-range order in denatured proteins induced by steric alignment in polyacrylamide gels (Shortle and Ackerman, 2001). As protein unfolding often requires hostile environments, stable inert matrices are necessary for the measurement of RDCs under these extreme conditions. We describe the development of a negatively charged polyacrylate/acrylamide copolymer fulfilling these requirements. Other polymers have meanwhile been devised for the introduction of positive charges (Ulmer et al., 2003) or for the alignment of organic molecules in a crosslinked polystyrene matrix (Luy et al., 2004).

Residual dipolar couplings have been recognized to increase the quality of structural models. The measurement of long-range RDCs in perdeuterated proteins in this thesis has greatly increased the number of experimental data available and has a potential to efficiently define the molecular geometry. Dipolar couplings over distances up to 7.7 Å, including RDCs across hydrogen bonds, have been observed at high experimental

precision. In a first application, we have shown that the positions of protons, which are not fixed in secondary structures, are incorrectly predicted from a high-resolution X-ray conformer of protein G.
The measurement of long-range, as well as one- and two-bond RDCs in perdeuterated protein G permits structural and dynamic studies with more than 16 RDCs per residue from only two alignment media. Structural models from RDCs in the complete absence of NOEs have been reported with roughly half the number of RDCs per residue (Hus et al., 2001). In addition, due to the high structural qualities obtained upon RDC refinement of model proteins (Ulmer et al., 2003), the use of RDCs may foster the understanding and prediction e.g. of hydrogen bond energies and of chemical shifts.

II.) We have shown that the measurement of RDCs can be used to study secondary structure formation in the β-hairpin of the fibritin foldon A-state, which may be regarded as a stabilized folding intermediate. This folding nucleus contains a single native β-hairpin. RDCs of amide groups indicate that the β-hairpin has stable hydrogen bonds, whereas the sidechain packing is more flexible as judged from C^{α}–H^{α} RDCs. Overall, the hairpin is highly stable in aqueous solution, underlining that foldon is a protein domain, which is evolutionarily optimized for fast and stable structure formation.

Increasing disorder during thermal unfolding of the foldon A-state has been monitored by the use of RDCs measured upon weak alignment in covalent hydrogels. Local structural stabilities of the turn and of large hydrophobic residues in the β-hairpin have been observed. Furthermore NMR relaxation data have shown that the native structure of the foldon hairpin is formed in ~6 μs at room temperature. Due to the fast folding of secondary structure elements in less than a few microseconds, these folding events have become computationally amenable to molecular dynamics simulations. Kinetic and thermodynamic "zipper models" of β-hairpin formation from such simulations (Dill et al., 1993; Munoz et al., 1998; Klimov and Thirumalai, 2000) corroborate the finding that residues near the turn are stabilized at earlier folding times and at higher temperatures than the strand residues.

III.) The measurement of RDCs at natural isotope enrichment has been used to obtain a high quality structure of a cysteine rich minicollagen domain, which exhibits a new fold. The cysteine rich domain forms the native structure and the correct disulfide bond pattern between six cysteines in high yield during oxidative folding. When these disulfide bridges are reduced, local turn structures are abolished, underlining the importance of disulfide bonds for the stabilization of structures in small polypeptide chains.

The domain is crosslinked in its natural environment to resist an osmotic pressure of roughly 150 bar, resulting in an explosive discharge of nematocysts. Crosslinking occurs via disulfide reshuffling from intra- to intermolecular disulfide bonds. Indeed, equilibrium unfolding indicates the presence of one vulnerable disulfide bond in the domain, which would be susceptible to reshuffling. Structures with reorganized disulfide bonds have not been obtained for structural characterization, yet, as the reshuffling is presumably catalyzed *in vivo*. Thus, an atomic model for the nematocyst cell wall structure and the assembly into a highly stable matrix is still to be developed.

IV.) Chemical shift assignments for the human PTP1B protein have been obtained. The protein is an attractive target for medication of type II diabetes and obesity. Successful drugs acting on PTP1B, however, have not been developed, yet. Collaborators in Novartis Pharmaceuticals Corporation are currently using the chemical shift assignments for lead screening.

Bibliography

Abragam, A. 1961. The principles of nuclear magnetism. Clarendon Press, Oxford,. 599 pp.

Al-Hashimi, H.M., Valafar, H., Terrell, M., Zartler, E.R., Eidsness, M.K., and Prestegard, J.H. 2000. Variation of molecular alignment as a means of resolving orientational ambiguities in protein structures from dipolar couplings. *Journal of Magnetic Resonance*. 143:402-6.

Andersen, J.N., Jansen, P.G., Echwald, S.M., Mortensen, O.H., Fukada, T., Del Vecchio, R., Tonks, N.K., and Moller, N.P. 2004. A genomic perspective on protein tyrosine phosphatases: gene structure, pseudogenes, and genetic disease linkage. *Faseb J*. 18:8-30.

Andersen, J.N., Mortensen, O.H., Peters, G.H., Drake, P.G., Iversen, L.F., Olsen, O.H., Jansen, P.G., Andersen, H.S., Tonks, N.K., and Moller, N.P. 2001. Structural and evolutionary relationships among protein tyrosine phosphatase domains. *Mol Cell Biol*. 21:7117-36.

Annila, A., Aitio, H., Thulin, E., and Drakenberg, T. 1999. Recognition of protein folds via dipolar couplings. *Journal of Biomolecular NMR*. 14:223-230.

Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*. 289:905-20.

Barchi, J.J., Jr., Grasberger, B., Gronenborn, A.M., and Clore, G.M. 1994. Investigation of the backbone dynamics of the IgG-binding domain of streptococcal protein G by heteronuclear two-dimensional ^{1}H-^{15}N nuclear magnetic resonance spectroscopy. *Protein Sci*. 3:15-21.

Barfield, M. 2002. Structural dependencies of interresidue scalar coupling $^{h3}J_{NC'}$ and donor ^{1}H chemical shifts in the hydrogen bonding regions of proteins. *Journal of the American Chemical Society*. 124:4158-68.

Barford, D., Flint, A.J., and Tonks, N.K. 1994. Crystal-Structure of Human Protein-Tyrosine-Phosphatase 1B. *Science*. 263:1397-1404.

Baumeister, W. 2002. Electron tomography: towards visualizing the molecular organization of the cytoplasm. *Current Opinion in Structural Biology*. 12:679-84.

Bax, A. 2003. Weak alignment offers new NMR opportunities to study protein structure and dynamics. *Protein Science*. 12:1-16.

Bax, A., and Grzesiek, S. 1993. Methodological Advances in Protein NMR. *Accounts of Chemical Research*. 26:131-138.

Benton, M.J., and Ayala, F.J. 2003. Dating the tree of life. *Science*. 300:1698-700.

Blanco, F., Ramirez-Alvarado, M., and Serrano, L. 1998. Formation and stability of beta-hairpin structures in polypeptides. *Current Opinion in Structural Biology*. 8:107-11.

Blanco, F.J., Rivas, G., and Serrano, L. 1994. A short linear peptide that folds into a native stable beta-hairpin in aqueous solution. *Nature Structural Biology*. 1:584-90.

Blanquet, R., and Lenhoff, H.M. 1966. A disulfide-linked collagenous protein of nematocyst capsules. *Science*. 154:152-3.

Bothnerby, A.A., Domaille, P.J., and Gayathri, C. 1981. Ultra-High-Field NMR-Spectroscopy - Observation of Proton-Proton Dipolar Coupling in Paramagnetic Bis[Tolyltris(Pyrazolyl)Borato]Cobalt(II). *Journal of the American Chemical Society*. 103:5602-5603.

Boudko, S.P., and Engel, J. 2004. Structure formation in the C terminus of type III collagen guides disulfide cross-linking. *Journal of Molecular Biology*. 335:1289-97.

Boudko, S.P., Londer, Y.Y., Letarov, A.V., Sernova, N.V., Engel, J., and Mesyanzhinov, V.V. 2002. Domain organization, folding and stability of bacteriophage T4 fibritin, a segmented coiled-coil protein. *European Journal of Biochemistry*. 269:833-841.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T., and Warren, G.L. 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallographica Section D-Biological Crystallography*. 54:905-921.

Brüschweiler, R. 2003. New approaches to the dynamic interpretation and prediction of NMR relaxation data from proteins. *Current Opinion in Structural Biology*. 13:175-83.

Castellani, F., van Rossum, B., Diehl, A., Schubert, M., Rehbein, K., and Oschkinat, H. 2002. Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature*. 420:98-102.

Castellani, F., van Rossum, B.J., Diehl, A., Rehbein, K., and Oschkinat, H. 2003. Determination of solid-state NMR structures of proteins by means of three-dimensional 15N-13C-13C dipolar correlation spectroscopy and chemical shift analysis. *Biochemistry*. 42:11476-83.

Chothia, C. 1992. Proteins. One thousand families for the molecular biologist. *Nature*. 357:543-4.

Christendat, D., Yee, A., Dharamsi, A., Kluger, Y., Gerstein, M., Arrowsmith, C.H., and Edwards, A.M. 2000. Structural proteomics: prospects for high throughput sample preparation. *Prog Biophys Mol Biol*. 73:339-45.

Clore, G.M., and Schwieters, C.D. 2004. How much backbone motion in ubiquitin is required to account for dipolar coupling data measured in multiple alignment media as assessed by independent cross-validation? *Journal of the American Chemical Society*. 126:2923-38.

Clore, G.M., Starich, M.R., and Gronenborn, A.M. 1998. Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses. *Journal of the American Chemical Society*. 120:10571-10572.

Cohen, J.S., Jaroszewski, J.W., Kaplan, O., RuizCabello, J., and Collier, S.W. 1995. A history of biological applications of NMR spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy*. 28:53-85.

Cordier, F., Barfield, M., and Grzesiek, S. 2003. Direct observation of C^{α}-H^{α}...O=C hydrogen bonds in proteins by interresidue $^{h3}J_{C\alpha C'}$ scalar couplings. *Journal of the American Chemical Society*. 125:15750-1.

Cordier, F., Dingley, A.J., and Grzesiek, S. 1999. A doublet-separated sensitivity-enhanced HSQC for the determination of scalar and dipolar one-bond J-couplings. *Journal of Biomolecular NMR*. 13:175-180.

Cordier, F., and Grzesiek, S. 1999. Direct observation of hydrogen bonds in proteins by interresidue $^{3h}J_{NC'}$ scalar couplings. *Journal of the American Chemical Society*. 121:1601-1602.

Cornilescu, G., Marquardt, J.L., Ottiger, M., and Bax, A. 1998. Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *Journal of the American Chemical Society*. 120:6836-6837.

Creighton, T.E. 1997. Protein folding coupled to disulphide bond formation. *Biol Chem*. 378:731-44.

de Alba, E., and Tjandra, N. 2002. NMR dipolar couplings for the structure determination of biopolymers in solution. *Progress in Nuclear Magnetic Resonance Spectroscopy*. 40:175-197.

de Groot, H.J. 2000. Solid-state NMR spectroscopy applied to membrane proteins. *Current Opinion in Structural Biology*. 10:593-600.

Delaglio, F., Kontaxis, G., and Bax, A. 2000. Protein structure determination using molecular fragment replacement and NMR dipolar couplings. *Journal of the American Chemical Society*. 122:2142-2143.

Derrick, J.P., and Wigley, D.B. 1994. The third IgG-binding domain from streptococcal protein G. An analysis by X-ray crystallography of the structure alone and in a complex with Fab.PG. *Journal of Molecular Biology*. 243:906-18.

Dill, K.A., Fiebig, K.M., and Chan, H.S. 1993. Cooperativity in protein-folding kinetics. *Proceedings of the National Academy of Sciences of the United States of America*. 90:1942-6.

Dingley, A.J., and Grzesiek, S. 1998. Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide $^{2}J_{NN}$ couplings. *Journal of the American Chemical Society*. 120:8293-8297.

Dosset, P., Hus, J.C., Blackledge, M., and Marion, D. 2000. Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *Journal of Biomolecular NMR*. 16:23-28.

Elchebly, M., Payette, P., Michaliszyn, E., Cromlish, W., Collins, S., Loy, A.L., Normandin, D., Cheng, A., Himms-Hagen, J., Chan, C.C., Ramachandran, C., Gresser, M.J., Tremblay, M.L., and Kennedy, B.P. 1999. Increased insulin sensitivity and obesity resistance in mice lacking the protein tyrosine phosphatase-1B gene. *Science*. 283:1544-8.

Engel, U., Ozbek, S., Streitwolf-Engel, R., Petri, B., Lottspeich, F., Holstein, T.W., Oezbek, S., and Engel, R. 2002. Nowa, a novel protein with minicollagen Cys-rich domains, is involved in nematocyst formation in Hydra. *J Cell Sci*. 115:3923-34.

Engel, U., Pertz, O., Fauser, C., Engel, J., David, C.N., and Holstein, T.W. 2001. A switch in disulfide linkage during minicollagen assembly in Hydra nematocysts. *Embo Journal*. 20:3063-3073.

Engler, N., Ostermann, A., Niimura, N., and Parak, F.G. 2003. Hydrogen atoms in proteins: positions and dynamics. *Proceedings of the National Academy of Sciences of the United States of America*. 100:10243-8.

Ernst, R.R., Bodenhausen, G., and Wokaun, A. 1987. Principles of nuclear magnetic resonance in one and two dimensions. Clarendon Press ;
Oxford University Press, Oxford [Oxfordshire], New York. xxiv, 610 pp.

Fiaux, J., Bertelsen, E.B., Horwich, A.L., and Wüthrich, K. 2002. NMR analysis of a 900K GroEL GroES complex. *Nature*. 418:207-11.

Flint, A.J., Gebbink, M.F., Franza, B.R., Jr., Hill, D.E., and Tonks, N.K. 1993. Multi-site phosphorylation of the protein tyrosine phosphatase, PTP1B: identification of cell cycle regulated and phorbol ester stimulated sites of phosphorylation. *Embo Journal*. 12:1937-46.

Flory, P.J. 1953. Principles of polymer chemistry. Cornell University Press, Ithaca ; London. xvi, 672 pp.

Frank, S., Kammerer, R.A., Mechling, D., Schulthess, T., Landwehr, R., Bann, J., Guo, Y., Lustig, A., Bächinger, H.P., and Engel, J. 2001. Stabilization of short collagen-like triple helices by protein engineering. *Journal of Molecular Biology*. 308:1081-9.

Goldman, M. 1988. Quantum description of high-resolution NMR in liquids. Clarendon Press ;
Oxford University Press, Oxford [Oxfordshire], New York. xiv, 268 pp.

Griesinger, C., Sorensen, O.W., and Ernst, R.R. 1987. Practical Aspects of the E-COSY Technique - Measurement of Scalar Spin Spin Coupling-Constants in Peptides. *Journal of Magnetic Resonance*. 75:474-492.

Gronenborn, A.M. 2003. Protein Structure Refinement using Residual Dipolar Couplings. *Biological Magnetic Resonance*. 20.

Güthe, S., Kapinos, L., Möglich, A., Meier, S., Grzesiek, S., and Kiefhaber, T. 2004. Very fast folding and association of a trimerization domain from bacteriophage T4 fibritin. *Journal of Molecular Biology*. 337:905-15.

Haj, F.G., Verveer, P.J., Squire, A., Neel, B.G., and Bastiaens, P.I. 2002. Imaging sites of receptor dephosphorylation by PTP1B on the surface of the endoplasmic reticulum. *Science*. 295:1708-11.

Hajduk, P.J., Meadows, R.P., and Fesik, S.W. 1999. NMR-based screening in drug discovery. *Q Rev Biophys*. 32:211-40.

Hansen, M.R., Mueller, L., and Pardi, A. 1998a. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nature Structural Biology*. 5:1065-1074.

Hansen, M.R., Rance, M., and Pardi, A. 1998b. Observation of long-range ^{1}H–^{1}H distances in solution by dipolar coupling interactions. *Journal of the American Chemical Society*. 120:11210-11211.

Hessinger, D.A., Lenhoff, H.M., Loma Linda University. Department of Physiology and Pharmacology., and University of California Irvine. Department of Developmental and Cell Biology. 1988. The biology of nematocysts. Academic Press, San Diego ; London. xii, 600 pp.

Hobmayer, B., Rentzsch, F., Kuhn, K., Happel, C.M., von Laue, C.C., Snyder, P., Rothbacher, U., and Holstein, T.W. 2000. WNT signalling molecules act in axis formation in the diploblastic metazoan Hydra. *Nature*. 407:186-9.

Hodgkinson, P., and Emsley, L. 1999. The accuracy of distance measurements in solid-state NMR. *Journal of Magnetic Resonance*. 139:46-59.

Holm, L., and Sander, C. 1994. Searching protein structure databases has come of age. *Proteins*. 19:165-73.

Holstein, T., and Tardent, P. 1984. An Ultrahigh-Speed Analysis of Exocytosis - Nematocyst Discharge. *Science*. 223:830-833.

Horkay, F., Tasaki, I., and Basser, P.J. 2000. Osmotic swelling of polyacrylate hydrogels in physiological salt solutions. *Biomacromolecules*. 1:84-90.

Hus, J.C., Marion, D., and Blackledge, M. 2001. Determination of protein backbone structure using only residual dipolar couplings. *Journal of the American Chemical Society*. 123:1541-1542.

Jaroniec, C.P., Tounge, B.A., Herzfeld, J., and Griffin, R.G. 2001. Frequency selective heteronuclear dipolar recoupling in rotating solids: accurate ^{13}C-^{15}N distance measurements in uniformly ^{13}C,^{15}N-labeled peptides. *Journal of the American Chemical Society*. 123:3507-19.

Jelsch, C., Teeter, M.M., Lamzin, V., Pichon-Pesme, V., Blessing, R.H., and Lecomte, C. 2000. Accurate protein crystallography at ultra-high resolution: valence electron distribution in crambin. *Proceedings of the National Academy of Sciences of the United States of America*. 97:3171-6.

Kahmann, J.D., Sass, H.J., Allan, M.G., Seto, H., Thompson, C.J., and Grzesiek, S. 2003. Structural basis for antibiotic recognition by the TipA class of multidrug-resistance transcriptional regulators. *Embo Journal*. 22:1824-34.

Karplus, M. 1959. Contact Electron-Spin Coupling of Nuclear Magnetic Moments. *Journal of Chemical Physics*. 30:11-15.

Kay, L.E., Torchia, D.A., and Bax, A. 1989. Backbone Dynamics of Proteins as Studied by N-15 Inverse Detected Heteronuclear NMR-Spectroscopy - Application to Staphylococcal Nuclease. *Biochemistry*. 28:8972-8979.

Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H., and Phillips, D.C. 1958. 3-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature*. 181:662-666.

Klimov, D.K., and Thirumalai, D. 2000. Mechanisms and kinetics of beta-hairpin formation. *Proceedings of the National Academy of Sciences of the United States of America*. 97:2544-9.

Korzhnev, D.M., Kloiber, K., Kanelis, V., Tugarinov, V., and Kay, L.E. 2004. Probing slow dynamics in high molecular weight proteins by methyl-TROSY NMR spectroscopy: application to a 723-residue enzyme. *Journal of the American Chemical Society*. 126:3964-73.

Kurz, E.M., Holstein, T.W., Petri, B.M., Engel, J., and David, C.N. 1991. Mini-collagens in hydra nematocytes. *J Cell Biol*. 115:1159-69.

Lauterbur, P.C. 1973. Image Formation by Induced Local Interactions - Examples Employing Nuclear Magnetic-Resonance. *Nature*. 242:190-191.

LeMaster, D.M., and Richards, F.M. 1988. NMR sequential assignment of Escherichia coli thioredoxin utilizing random fractional deuteriation.PG - 142-50. *Biochemistry*. 27:142-50.

Letarov, A.V., Londer, Y.Y., Boudko, S.P., and Mesyanzhinov, V.V. 1999. The carboxy-terminal domain initiates trimerization of bacteriophage T4 fibritin. *Biochemistry-Moscow*. 64:817-823.

Lifson, S. 1961. Theory of Helix-Coil Transition in Polypeptides. *Journal of Chemical Physics*. 34:1963-&.

Lipari, G., and Szabo, A. 1982. Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules. 2. Analysis of Experimental Results. *Journal of the American Chemical Society*. 104:4559-4570.

Loria, J.P., Rance, M., and Palmer, A.G. 1999. A TROSY CPMG sequence for characterizing chemical exchange in large proteins. *Journal of Biomolecular NMR*. 15:151-155.

Luca, S., White, J.F., Sohal, A.K., Filippov, D.V., van Boom, J.H., Grisshammer, R., and Baldus, M. 2003. The conformation of neurotensin bound to its G protein-coupled receptor. *Proceedings of the National Academy of Sciences of the United States of America*. 100:10706-11.

Luy, B., Kokar, K., and Kessler, H. 2004. An easy and scalable method for the partial alignment of organic molecules for measuring residual dipolar couplings. *Angewandte Chemie-International Edition*. 43:1092-1094.

Meier, S., Güthe, S., Kiefhaber, T., and Grzesiek, S. 2004a. Structure and dynamics in the fibritin folding nucleus containing a stable β-hairpin. *in preparation*.

Meier, S., Häussinger, D., and Grzesiek, S. 2002a. Charged acrylamide copolymer gels as media for weak alignment. *Journal of Biomolecular NMR*. 24:351-356.

Meier, S., Häussinger, D., Jensen, P., Rogowski, M., and Grzesiek, S. 2003. High-accuracy residual $^{1}H^{N}$-^{13}C and $^{1}H^{N}$-$^{1}H^{N}$ dipolar couplings in perdeuterated proteins. *Journal of the American Chemical Society*. 125:44-45.

Meier, S., Häussinger, D., Pokidysheva, E., Bächinger, H.P., and Grzesiek, S. 2004b. Determination of a high-precision NMR structure of the minicollagen cysteine rich domain from Hydra and characterization of its disulfide bond formation. *FEBS Letters*. *in press*.

Meier, S., Li, Y.C., Koehn, J., Vlattas, I., Wareing, J., Jahnke, W.G., Wennogle, L.P., and Grzesiek, S. 2002b. Letter to the Editor: Backbone resonance assignment of the 298 amino acid catalytic domain of protein tyrosine phosphatase 1B (PTP1B). *Journal of Biomolecular NMR*. 24:165-166.

Meier, S., Pokidysheva, E., Milbradt, A., Renner, C., Häussinger, D., Bächinger, H.P., Moroder, L., Grzesiek, S., Holstein, T., Ozbek, S., and Engel, J. 2004c. The structure of the Cys-rich terminal domain of Hydra minicollagen, which is involved in disulfide networks of the nematocyst wall. *Journal of Biological Chemistry*. *in press*.

Meiler, J., Prompers, J.J., Peti, W., Griesinger, C., and Brüschweiler, R. 2001. Model-free approach to the dynamic interpretation of residual dipolar couplings in globular proteins. *Journal of the American Chemical Society*. 123:6098-6107.

Meissner, A., Duus, J.O., and Sorensen, O.W. 1997. Integration of spin-state-selective excitation into 2D NMR correlation experiments with heteronuclear ZQ/2Q pi rotations for (1)J(XH)-resolved E.COSY-type measurement of heteronuclear coupling constants in proteins. *Journal of Biomolecular NMR*. 10:89-94.

Munoz, V., Henry, E.R., Hofrichter, J., and Eaton, W.A. 1998. A statistical mechanical model for beta-hairpin kinetics. *Proceedings of the National Academy of Sciences of the United States of America*. 95:5872-9.

Munoz, V., Thompson, P.A., Hofrichter, J., and Eaton, W.A. 1997. Folding dynamics and mechanism of beta-hairpin formation. *Nature*. 390:196-9.

O'Neil-Cabello, E., Bryce, D.L., Nikonowicz, E.P., and Bax, A. 2004. Measurement of five dipolar couplings from a single 3D NMR multiplet applied to the study of RNA dynamics. *Journal of the American Chemical Society*. 126:66-7.

Orengo, C.A., Pearl, F.M.G., Bray, J.E., Todd, A.E., Martin, A.C., Lo Conte, L., and Thornton, J.M. 1999. The CATH database provides insights into protein structure/function relationships. *Nucleic Acids Research*. 27:275-279.

Ottiger, M., Delaglio, F., and Bax, A. 1998. Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *Journal of Magnetic Resonance*. 131:373-378.

Overhauser, A.W. 1953. Polarization of Nuclei in Metals. *Physical Review*. 92:411-415.

Ozbek, S., Engel, U., and Engel, J. 2002a. A switch in disulfide linkage during minicollagen assembly in hydra nematocysts or how to assemble a 150-bar-resistant structure. *Journal of Structural Biology*. 137:11-14.

Ozbek, S., Pertz, O., Schwager, M., Lustig, A., Holstein, T., and Engel, J. 2002b. Structure/function relationships in the minicollagen of hydra nematocysts. *Journal of Biological Chemistry*. 277:49200-49204.

Pan, K.M., Baldwin, M., Nguyen, J., Gasset, M., Serban, A., Groth, D., Mehlhorn, I., Huang, Z.W., Fletterick, R.J., Cohen, F.E., and Prusiner, S.B. 1993. Conversion of Alpha-Helices into Beta-Sheets Features in the Formation of the Scrapie Prion Proteins. *Proceedings of the National Academy of Sciences of the United States of America*. 90:10962-10966.

Perutz, M.F., Rossmann, M.G., Cullis, A.F., Muirhead, H., Will, G., and North, A.C.T. 1960. Structure of Haemoglobin - 3-Dimensional Fourier Synthesis at 5.5-Å Resolution, Obtained by X-Ray Analysis. *Nature*. 185:416-422.

Pervushin, K., Ono, A., Fernandez, C., Szyperski, T., Kainosho, M., and Wüthrich, K. 1998. NMR scaler couplings across Watson-Crick base pair hydrogen bonds in DNA observed by transverse relaxation optimized spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America*. 95:14147-14151.

Pervushin, K., Riek, R., Wider, G., and Wüthrich, K. 1997. Attenuated T-2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences of the United States of America*. 94:12366-12371.

Petkova, A.T., Ishii, Y., Balbach, J.J., Antzutkin, O.N., Leapman, R.D., Delaglio, F., and Tycko, R. 2002. A structural model for Alzheimer's beta-amyloid fibrils based on experimental constraints from solid state NMR. *Proceedings of the National Academy of Sciences of the United States of America*. 99:16742-16747.

Prestegard, J.H., al-Hashimi, H.M., and Tolman, J.R. 2000. NMR structures of biomolecules using field oriented media and residual dipolar couplings. *Q Rev Biophys*. 33:371-424.

Proctor, W.G., and Yu, F.C. 1950. The Dependence of a Nuclear Magnetic Resonance Frequency Upon Chemical Compound. *Physical Review*. 77:717-717.

Prosser, R.S., Losonczi, J.A., and Shiyanovskaya, I.V. 1998. Use of a novel aqueous liquid crystalline medium for high- resolution NMR of macromolecules in solution. *Journal of the American Chemical Society*. 120:11010-11011.

Ramirez, B.E., and Bax, A. 1998. Modulation of the alignment tensor of macromolecules dissolved in a dilute liquid crystalline medium. *Journal of the American Chemical Society*. 120:9106-9107.

Ramirez, B.E., Voloshin, O.N., Camerini-Otero, R.D., and Bax, A. 2000. Solution structure of DinI provides insight into its mode of RecA inactivation. *Protein Science*. 9:2161-9.

Richardson, J.S. 1981. The anatomy and taxonomy of protein structure. *Adv Protein Chem*. 34:167-339.

Ruckert, M., and Otting, G. 2000. Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *Journal of the American Chemical Society*. 122:7793-7797.

Salmeen, A., Andersen, J.N., Myers, M.P., Meng, T.C., Hinks, J.A., Tonks, N.K., and Barford, D. 2003. Redox regulation of protein tyrosine phosphatase 1B involves a sulphenyl-amide intermediate. *Nature*. 423:769-73.

Salzmann, M., Wider, G., Pervushin, K., Senn, H., and Wüthrich, K. 1999. TROSY-type triple-resonance experiments for sequential NMR assignments of large proteins. *Journal of the American Chemical Society*. 121:844-848.

Sanders, C.R., 2nd, and Schwonek, J.P. 1992. Characterization of magnetically orientable bilayers in mixtures of dihexanoylphosphatidylcholine and dimyristoylphosphatidylcholine by solid-state NMR. *Biochemistry*. 31:8898-905.

Sass, H.J., Musco, G., Stahl, S.J., Wingfield, P.T., and Grzesiek, S. 2000. Solution NMR of proteins within polyacrylamide gels: Diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes. *Journal of Biomolecular NMR*. 18:303-309.

Sass, J., Cordier, F., Hoffmann, A., Cousin, A., Omichinski, J.G., Lowen, H., and Grzesiek, S. 1999. Purple membrane induced alignment of biological macromolecules in the magnetic field. *Journal of the American Chemical Society*. 121:2047-2055.

Saupe, A., and Englert, G. 1963. High-Resolution Nuclear Magnetic Resonance Spectra of Orientated Molecules. *Physical Review Letters*. 11:462-&.

Schoenborn, B.P. 1969. Neutron diffraction analysis of myoglobin. *Nature*. 224:143-6.

Shenderovich, I.G., Smirnov, S.N., Denisov, G.S., Gindin, V.A., Golubev, N.S., Dunger, A., Reibke, R., Kirpekar, S., Malkina, O.L., and Limbach, H.H. 1998. Nuclear magnetic resonance of hydrogen bonded clusters between F- and (HF)(n): Experiment and theory. *Berichte Der Bunsen-Gesellschaft-Physical Chemistry Chemical Physics*. 102:422-428.

Shortle, D., and Ackerman, M.S. 2001. Persistence of native-like topology in a denatured protein in 8 M urea. *Science*. 293:487-489.

Spera, S., and Bax, A. 1991. Empirical Correlation between Protein Backbone Conformation and C-Alpha and C-Beta C-13 Nuclear-Magnetic-Resonance Chemical-Shifts. *Journal of the American Chemical Society*. 113:5490-5492.

Stehling, M.K., Turner, R., and Mansfield, P. 1991. Echo-planar imaging: magnetic resonance imaging in a fraction of a second. *Science*. 254:43-50.

Strelkov, S.V., Tao, Y., Shneider, M.M., Mesyanzhinov, V.V., and Rossmann, M.G. 1998. Structure of bacteriophage T4 fibritin M: a troublesome packing arrangement. *Acta Crystallographica Section D-Biological Crystallography*. 54:805-816.

Sun, H.M., Rodin, A., Zhou, Y.H., Dickinson, D.P., Harper, D.E., HewettEmmett, D., and Li, W.H. 1997. Evolution of paired domains: Isolation and sequencing of jellyfish and hydra Pax genes related to Pax-5 and Pax-6. *Proceedings of the National Academy of Sciences of the United States of America*. 94:5156-5161.

Tao, J., Malbon, C.C., and Wang, H.Y. 2001. Insulin stimulates tyrosine phosphorylation and inactivation of protein-tyrosine phosphatase 1B in vivo. *Journal of Biological Chemistry*. 276:29520-5.

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

Tian, F., Fowler, C.A., Zartler, E.R., Jenney, F.A., Jr., Adams, M.W., and Prestegard, J.H. 2000. Direct measurement of ^{1}H-^{1}H dipolar couplings in proteins: a complement to traditional NOE measurements. *Journal of Biomolecular NMR*. 18:23-31.

Tjandra, N., and Bax, A. 1997. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science*. 278:1111-1114.

Tjandra, N., Marquardt, J., and Clore, G.M. 2000. Direct refinement against proton-proton dipolar couplings in NMR structure determination of macromolecules. *Journal of Magnetic Resonance*. 142:393-396.

Tollinger, M., Crowhurst, K.A., Kay, L.E., and Forman-Kay, J.D. 2003. Site-specific contributions to the pH dependence of protein stability. *Proceedings of the National Academy of Sciences of the United States of America*. 100:4545-50.

Tolman, J.R. 2001. Dipolar couplings as a probe of molecular dynamics and structure in solution.PG - 532-9. *Current Opinion in Structural Biology*. 11:532-9.

Tolman, J.R., Al-Hashimi, H.M., Kay, L.E., and Prestegard, J.H. 2001. Structural and dynamic analysis of residual dipolar coupling data for proteins. *Journal of the American Chemical Society*. 123:1416-1424.

Tolman, J.R., Flanagan, J.M., Kennedy, M.A., and Prestegard, J.H. 1995. Nuclear Magnetic Dipole Interactions in Field-Oriented Proteins - Information for Structure Determination in Solution. *Proceedings of the National Academy of Sciences of the United States of America*. 92:9279-9283.

Tolman, J.R., Flanagan, J.M., Kennedy, M.A., and Prestegard, J.H. 1997. NMR evidence for slow collective motions in cyanometmyoglobin. *Nature Structural Biology*. 4:292-7.

Tonks, N.K. 2003. PTP1B: from the sidelines to the front lines! *FEBS Letters*. 546:140-8.

Tonks, N.K., Diltz, C.D., and Fischer, E.H. 1988. Purification of the Major Protein-Tyrosine-Phosphatases of Human-Placenta. *Journal of Biological Chemistry*. 263:6722-6730.

Torchia, D.A., Sparks, S.W., and Bax, A. 1988. Delineation of Alpha-Helical Domains in Deuteriated Staphylococcal Nuclease By 2D NOE NMR-Spectroscopy. *Journal of the American Chemical Society*. 110:2320-2321.

Trempe, J.F., Morin, F.G., Xia, Z.C., Marchessault, R.H., and Gehring, K. 2002. Characterization of polyacrylamide-stabilized Pf1 phage liquid crystals for protein NMR spectroscopy. *Journal of Biomolecular NMR*. 22:83-87.

Tugarinov, V., and Kay, L.E. 2003. Quantitative NMR studies of high molecular weight proteins: application to domain orientation and ligand binding in the 723 residue enzyme malate synthase G. *Journal of Molecular Biology*. 327:1121-33.

Tugarinov, V., Muhandiram, R., Ayed, A., and Kay, L.E. 2002. Four-dimensional NMR spectroscopy of a 723-residue protein: chemical shift assignments and secondary structure of malate synthase g. *Journal of the American Chemical Society*. 124:10025-35.

Tugarinov, V., Sprangers, R., and Kay, L.E. 2004. Line narrowing in methyl-TROSY using zero-quantum ^{1}H-^{13}C NMR spectroscopy. *Journal of the American Chemical Society*. 126:4921-5.

Tycko, R., Blanco, F.J., and Ishii, Y. 2000. Alignment of biopolymers in strained gels: A new way to create detectable dipole-dipole couplings in high-resolution biomolecular NMR. *Journal of the American Chemical Society*. 122:9340-9341.

Ulmer, T.S., Ramirez, B.E., Delaglio, F., and Bax, A. 2003. Evaluation of Backbone Proton Positions and Dynamics in a Small Protein by Liquid Crystal NMR Spectroscopy. *Journal of the American Chemical Society*. 125:9179-9191.

Vijay-Kumar, S., Bugg, C.E., and Cook, W.J. 1987. Structure of ubiquitin refined at 1.8 Å resolution. *Journal of Molecular Biology*. 194:531-44.

Vuister, G.W., and Bax, A. 1993. Quantitative J Correlation - a New Approach for Measuring Homonuclear 3-Bond J(H(N)H(Alpha) Coupling-Constants in N-15-Enriched Proteins. *Journal of the American Chemical Society*. 115:7772-7777.

Wagner, G., Pardi, A., and Wüthrich, K. 1983. Hydrogen-Bond Length and H-1-NMR Chemical-Shifts in Proteins. *Journal of the American Chemical Society*. 105:5948-5949.

Walton, K.M., and Dixon, J.E. 1993. Protein-Tyrosine Phosphatases. *Annual Review of Biochemistry*. 62:101-120.

Wang, C., Grey, M.J., and Palmer, A.G., 3rd. 2001. CPMG sequences with enhanced sensitivity to chemical exchange. *Journal of Biomolecular NMR*. 21:361-6.

Weber, J. 1990. Poly(gamma-glutamic acid)s are the major constituents of nematocysts in Hydra (Hydrozoa, Cnidaria). *J Biol Chem*. 265:9664-9.

Wlodawer, A., Walter, J., Huber, R., and Sjolin, L. 1984. Structure of bovine pancreatic trypsin inhibitor. Results of joint neutron and X-ray refinement of crystal form II. *Journal of Molecular Biology*. 180:301-29.

Wright, P.E., and Dyson, H.J. 1999. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *Journal of Molecular Biology*. 293:321-31.

Wu, Z., and Bax, A. 2002a. Measurement of long-range ^{1}H-^{1}H dipolar couplings in weakly aligned proteins. *Journal of the American Chemical Society*. 124:9672-3.

Wu, Z.R., and Bax, A. 2002b. Measurement of long-range ^{1}H-^{1}H dipolar couplings in weakly aligned proteins. *Journal of the American Chemical Society*. 124:9672-9673.

Wüthrich, K. 1986. NMR of proteins and nucleic acids. Wiley, New York. xv, 292 pp.

Xu, Y., Oyola, R., and Gai, F. 2003. Infrared study of the stability and folding kinetics of a 15-residue beta-hairpin. *Journal of the American Chemical Society*. 125:15388-94.

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

Zerella, R., Evans, P.A., Ionides, J.M., Packman, L.C., Trotter, B.W., Mackay, J.P., and Williams, D.H. 1999. Autonomous folding of a peptide corresponding to the N-terminal beta-hairpin from ubiquitin. *Protein Science*. 8:1320-31.

Zhang, R.M., and Snyder, G.H. 1989. Dependence of formation of small disulfide loops in two-cysteine peptides on the number and types of intervening amino acids. *Journal of Biological Chemistry*. 264:18472-9.

Zhang, Z.Y. 2002. Protein tyrosine phosphatases: structure and function, substrate specificity, and inhibitor development. *Annu Rev Pharmacol Toxicol*. 42:209-34.

Zimm, B.H., and Bragg, J.K. 1959. Theory of the Phase Transition between Helix and Random Coil in Polypeptide Chains. *Journal of Chemical Physics*. 31:526-535.

Zimm, B.H., Doty, P., and Iso, K. 1959. Determination of the Parameters for Helix Formation in Poly-Gamma-Benzyl-L-Glutamate. *Proceedings of the National Academy of Sciences of the United States of America*. 45:1601-1604.

Zwahlen, C., Vincent, S.J.F., Gardner, K.H., and Kay, L.E. 1998. Significantly improved resolution for NOE correlations from valine and isoleucine (C-gamma 2) methyl groups in N-15,C-13- and N-15,C-13,H-2-labeled proteins. *Journal of the American Chemical Society*. 120:4825-4831.

Zweckstetter, M., and Bax, A. 2000. Prediction of sterically induced alignment in a dilute liquid crystalline phase: Aid to protein structure determination by NMR. *Journal of the American Chemical Society*. 122:3791-3792.