

## Nuclear magnetic resonance as a quantitative tool to study interactions in biomacromolecules\*

S. Grzesiek<sup>‡</sup>, M. Allan, F. Cordier, D. Häussinger, P. Jensen, J. Kahmann, S. Meier, and H. J. Sass

*Biozentrum, University of Basel, CH 4056 Basel, Switzerland*

*Abstract:* High-resolution nuclear magnetic resonance (NMR) has emerged as one of the most versatile tools for the quantitative study of structure, kinetics, and thermodynamics of biomolecules and their interactions at atomic resolution. Traditionally, nuclear Overhauser enhancements (NOEs) and chemical shift perturbation methods are used to determine molecular geometries and to identify contact surfaces, but more recently, weak anisotropic orientation, anisotropic diffusion, and scalar couplings across hydrogen bonds provide additional information.

Examples of such technologies are shown as applied to the quantitative characterization of function and thermodynamics of several biomacromolecules. In particular, (1) the structural and dynamical changes of the TipA multidrug resistance protein are followed upon antibiotic binding, (2) the trimer-monomer equilibrium and thermal unfolding of foldon, a small and very efficient trimerization domain of the T4 phagehead, is described in atomic detail, and (3) the changes of individual protein hydrogen bonds during thermal unfolding are quantitatively followed by scalar couplings across hydrogen bonds.

*Keywords:* Biomacromolecules; biomolecular interactions; protein folding; thermal unfolding; hydrogen bonds; scalar coupling; anisotropic orientation; protein dynamics; TipA; foldon.

### INTRODUCTION

Since the first determination of a complete protein structure by solution nuclear magnetic resonance (NMR) in 1985 by Wüthrich and coworkers [1], this technique has evolved tremendously. Most of the advances can be linked to the use of isotope labeling of biomolecules with the stable isotopes <sup>13</sup>C, <sup>15</sup>N, and <sup>2</sup>H [2–4]. As a consequence, ever larger molecular weights have become accessible with the largest complete structure determinations in the range of 40–60 kDa, complete resonance assignments of up to 81 kDa [5], and resolved spectra of proteins as large as 900 kDa [6].

However, the use of isotope labeling and the development of other new techniques have also increased enormously the number and precision of NMR observables that can be used for the description of biomacromolecular structure, dynamics, and function in solution. Most notable are advances in NMR relaxation experiments that yield information on amplitudes and time scales of global and local motions

---

\*Paper based on a presentation at the 18<sup>th</sup> IUPAC International Conference on Chemical Thermodynamics (ICCT-2004), 17–21 August 2004, Beijing, China. Other presentations are published in this issue, pp. 1297–1444.

<sup>‡</sup>Corresponding author: E-mail: stephan.grzesiek@unibas.ch

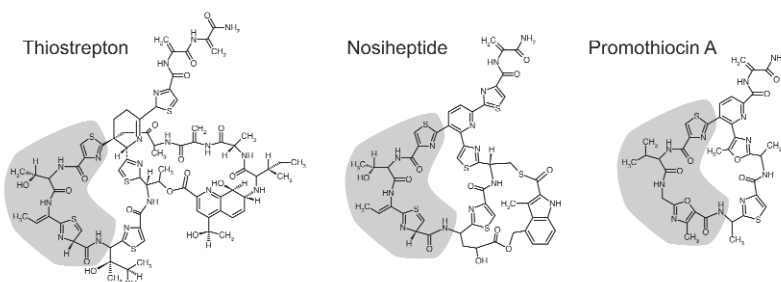
[7], the weak anisotropic orientation of biological macromolecules in magnetic fields that allow the precise determination of the average directions of internuclear vectors relative to an external director [8–10], and the finding that hydrogen bonds can be detected directly and described quantitatively by electron-mediated scalar couplings between magnetic nuclei on both sides of the hydrogen bridge [11–17].

In the following, we show examples of such techniques applied to the description of structural and dynamical changes during antibiotic binding of an antibiotic resistance regulator, the dissociation and thermal unfolding of a homotrimeric protein, and the behavior of individual protein hydrogen bonds during thermal unfolding.

## TRANSITION FROM STRUCTURAL DISORDER TO ORDER DURING THE ANTIBIOTIC BINDING OF AN ANTIBIOTIC RESISTANCE REGULATOR PROTEIN

### TipA proteins

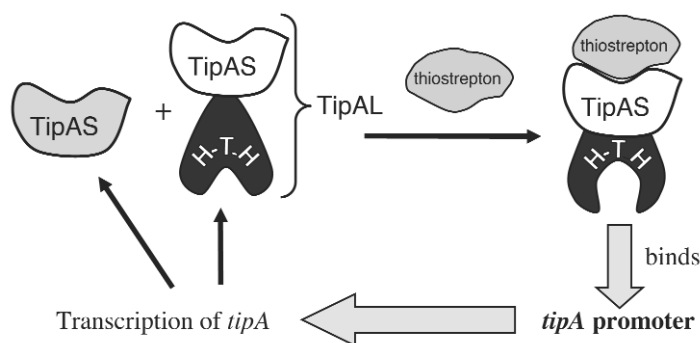
Multidrug resistance (MDR) systems in bacteria often rely on transport proteins or modifying enzymes to exclude or inactivate a large variety of different antimicrobial xenobiotics, and on transcriptional regulatory proteins having a comparable recognition spectrum [18]. One example of such promiscuous transcriptional regulatory proteins is the thiostrepton-induced protein A (TipA) [19], which was first identified in *Streptomyces*, organisms recognized as the producers of most natural antibiotics. The *tipA* gene can confer resistance to the antibiotic thiostrepton in *Streptomyces lividans*. Its transcription is induced by a large family of antibiotics produced by *Streptomyces*, *Bacillus*, and *Micrococcus* including thiostrepton, nosiheptide, and promothiocin (Fig. 1A). These compounds are highly modified, cyclic thiopепptides with an attached linear peptide “tail” containing at least one dehydroalanine.



**Fig. 1** Various antibiotics of the thiostrepton class recognized by TipA proteins.

In response to thiostrepton and other antibiotics of its class, the *tipA* gene is expressed in *S. lividans* as two alternate in-frame translation products (Fig. 2): a long form, TipAL (253 amino acids), and a short form, TipAS (144 amino acids), which constitutes the C-terminal part of TipAL. The N-terminal domain of TipAL is responsible for DNA binding, whereas the C-terminal TipAS forms the antibiotic recognition domain.

Antibiotic binding to TipAS or TipAL is irreversible owing to the formation of a covalent bond between cysteine-214 and a dehydroalanine residue of the antibiotic. Binding of thiostrepton increases the affinity of TipAL to its DNA operator site *ptipA*. This in turn leads to the binding of TipAL to the DNA operator and activates transcription of both TipAL and TipAS. Thus, TipAL is its own transcriptional activator. Other downstream genes have also been identified and are upregulated owing to this activation mechanism.



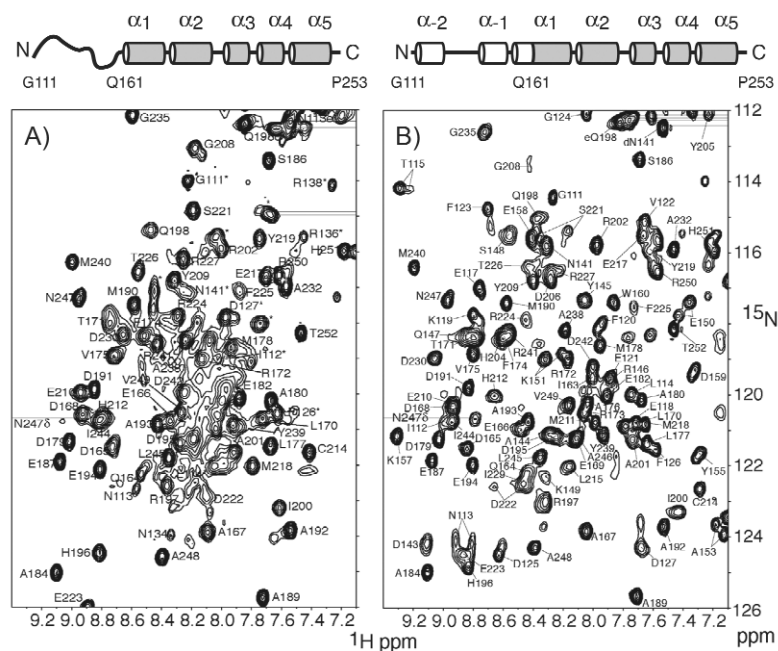
**Fig. 2** Mechanism of *tipA* transcriptional activation by antibiotic binding.

The following questions arise about this mechanism: (1) how are the TipA proteins able to recognize an entire class of structurally different antibiotics and (2) which conformational changes are induced during complex formation that ultimately lead to DNA binding? Parts of these questions can be addressed from the structural and dynamical information derived from solution NMR methods [19].

### Chemical shifts as sensitive probes of structural order in biomolecules

A first glimpse of structural changes induced by the antibiotic binding can be obtained from a two-dimensional  $^1\text{H}$ - $^{15}\text{N}$  COSY spectrum of the TipAS protein (Fig. 3). This spectrum correlates in its two dimensions the amide proton ( $^1\text{H}$ ) and nitrogen  $^{15}\text{N}$  chemical shifts of the isotope labeled protein. In general, protein chemical shifts have a very strong structural dependence, and programs such as TALOS [20] exist that are able to predict secondary structure conformations from backbone chemical shifts alone. In a  $^1\text{H}$ - $^{15}\text{N}$  COSY, one cross-peak should be visible for every non-proline amino acid of a protein. The spectrum of TipAS in the absence of antibiotic (Fig. 3A) shows only about 100 of such resolved amide resonances. Apparently, not all of the 139 non-proline amino acids of TipAS are detectable. Part of this phenomenon is caused by serious overlap in the central part of the spectrum ( $\delta^1\text{H} \sim 8.1$  ppm;  $\delta^{15}\text{N} \sim 120$  ppm). This central part of the  $^1\text{H}$ - $^{15}\text{N}$  COSY usually contains amino acids that are in unstructured, random coil conformations. A second reason for the missing amide resonances is broadening of the lines beyond detection owing to conformational exchange on the micro- to millisecond time scale of chemical shifts. Thus, the pure visual inspection of the  $^1\text{H}$ - $^{15}\text{N}$  COSY in Fig. 3A indicates that about 40 amino acids of the protein are unstructured and/or undergo conformational exchange on the micro- to millisecond time scale. In contrast, when thioestrepton antibiotics such as promothiocin A are added to the protein, the  $^1\text{H}$ - $^{15}\text{N}$  COSY spectrum undergoes a dramatic change (Fig. 3B). The overlap in the central part of the spectrum disappears, many more resonances are detected, and the spectral dispersion is largely increased. Without deeper analysis, this observation unambiguously indicates that binding of the ligand causes large structural changes from an only partially folded to a folded protein.

After a more extended analysis that determines the sequence-specific assignment of the protein backbone resonances by standard heteronuclear NMR methods, the chemical shifts can be used to define the sequence positions of secondary structure elements in a protein. For the TipAS protein, the backbone  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shifts indicate that the secondary structure of free TipAS consists of five C-terminal  $\alpha$ -helices. However, approximately 40 N-terminal amino acids are unstructured (Fig. 3A). In contrast, the binding of promothiocin A (and also of the other thioestrepton-like antibiotics) to TipAS induces the formation of two additional  $\alpha$ -helices in this N-terminal part, whereas the secondary structure of the C-terminal part remains virtually unchanged (Fig. 3B). Thus, the ligand-induced change from the partially folded to the fully folded protein can be identified with the transition of the unfolded N-terminus to an all  $\alpha$ -helical structure.

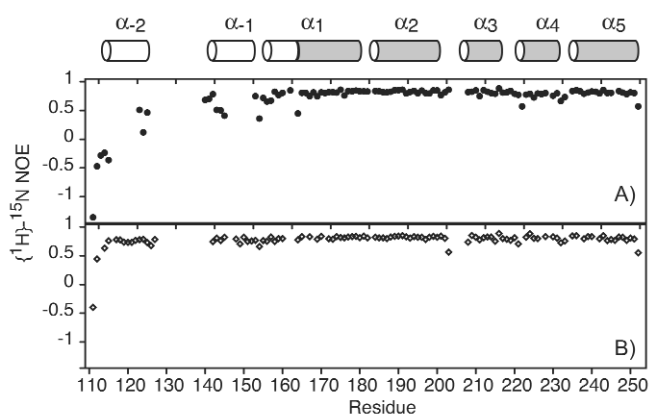


**Fig. 3**  $^1\text{H}$ - $^{15}\text{N}$  COSY spectra of free TipAS (A) and in complex with promothiocin A (B). The secondary structures of the respective states of TipAS are shown on top. Helices only formed after complexation with the antibiotic are depicted as white barrels.

### Detection of ligand-induced changes in protein dynamics by $^{15}\text{N}$ relaxation

This folding transition can also be described by dynamical parameters that are accessible from  $^{15}\text{N}$  nuclear relaxation experiments. Such relaxation studies have emerged as a powerful approach for the determination of the motional properties of molecules in solution [7]. In a  $^{15}\text{N}$  relaxation experiment, one creates nonequilibrium nuclear spin order and records how this relaxes back to thermal equilibrium. The return to equilibrium of the  $^{15}\text{N}$  nuclei is caused by the thermal motions of the molecule. The random motions lead, for example, to random magnetic dipolar interactions between the  $^{15}\text{N}$  nuclei and the nearby amide  $^1\text{H}$  nuclei. These random dipolar interactions couple the nuclear spins to the heat reservoir of the biomolecules and the solution. Thus, nuclear relaxation rates depend on the amplitudes and time scales of the thermal motions, which include the overall and the internal motions of the biomolecule. The effect of the motions on relaxation rates is described by a spectral density function  $J(\omega)$  that gives the relative distribution of the frequencies of the diffusive motions. Usually,  $^{15}\text{N}$  longitudinal ( $T_1$ ), transverse ( $T_2$ ) relaxation times and  $\{^1\text{H}\}$ - $^{15}\text{N}$  nuclear Overhauser enhancements (NOEs) are measured for all the amide groups of a protein. Of these, the NOEs have a very steep dependence on the time scale of local internal motions. NOE values smaller than about 0.6 correspond to large amplitude motions of the N-H vector on time scales of the proton Larmor frequencies (typically 600–800 MHz) or faster, whereas NOEs around 0.8 indicate a very limited amplitude for such fast motions.

The folding transition between the free and the antibiotic bound form of TipAS is very clearly visible from such NOE data. For the free TipAS protein (Fig. 4),  $\{^1\text{H}\}$ - $^{15}\text{N}$  NOE values within the C-terminal, folded part of the sequence are close to 0.8 and indicate that the internal mobility on the nanosecond time scale is very limited and comparable to other stable and well-folded proteins. In contrast, for the residues that could be assigned within the N-terminal unfolded region of free TipAS, the  $\{^1\text{H}\}$ - $^{15}\text{N}$  NOE values are significantly below 0.8. This shows the presence of large amplitude motions for this region on the subnanosecond time scale. When the antibiotic is bound to the protein (Fig. 4B),



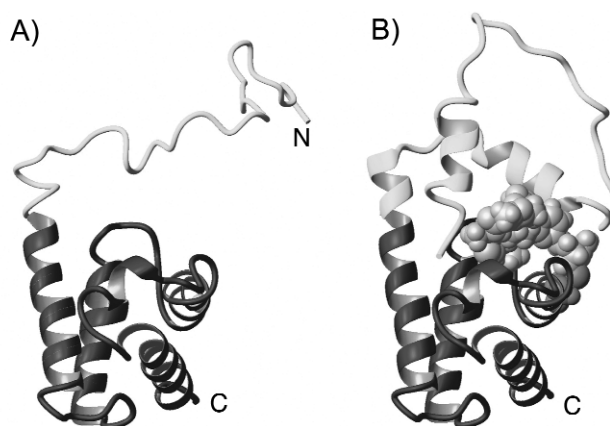
**Fig. 4** Sequence-specific heteronuclear  $\{^1\text{H}\}\text{-}^{15}\text{N}$  NOEs of free TipAS (A) and in complex with promothiocin A (B). The secondary structure elements are shown on the top. Ligand-induced  $\alpha$ -helices are depicted as white barrels.  $\alpha$ -helices present in both forms of the protein are shown as black barrels. Note that due to extensive line broadening from intermediate exchange, no data are available for a number of residues in the N-terminal region.

the  $\{^1\text{H}\}\text{-}^{15}\text{N}$  NOE values within the C-terminal, folded part are not changed, but the NOEs in the N-terminal region also increase to values around 0.8. Thus, the subnanosecond flexibility in this now folded N-terminus is significantly reduced and becomes identical to the rest of the folded protein.

In principle, a further quantitative analysis of the  $^{15}\text{N}$  relaxation data is possible. This analysis then yields order parameters of individual N–H internuclear distance vectors for subnanosecond time scale motions (see, e.g., [21]). Changes in such order parameters reflect altered flexibility and consequent changes in conformational entropy [7]. Thus, the order parameters can be used to estimate an upper bound for the change in the Gibbs free energy,  $\Delta G$ , or the entropy,  $\Delta S$ , resulting from conformational restriction.

### Structures of apo and complexed forms of TipAS: The coupling of binding and folding

After a complete analysis of all NMR data that yield structural information such as  $^1\text{H}\text{-}^1\text{H}$  NOEs, scalar couplings, and residual dipolar couplings (RDCs), the structure of the folded part of free (apo) TipAS [19] and a preliminary structure of the TipAS–Promothiocin A complex could be determined (Fig. 5). The antibiotic binds into a deep cleft in the folded part of apo TipAS. In the complexed form, the newly formed N-terminal helices wrap themselves around the antibiotic, whereas the rest of the structure is very similar to the free form. The preliminary structure of the TipAS complex gives an explanation of the promiscuous binding behavior to the class of thiostrepton antibiotics. Despite large variations in the overall structure of these antibiotics, the first macrocycle (shaded in Fig. 1) is more conserved. In the complex, this conserved motif of the antibiotic is accommodated by the large cleft in the stable part of the apo TipAS structure, whereas the flexible N-terminus of apo TipAS wraps itself around the variable parts of the antibiotic. The binding apparently stabilizes the N-terminus to form the additional  $\alpha$ -helices. Many examples of such coupled binding and folding events of unstructured proteins have been revealed by NMR spectroscopy in recent years. As pointed out by Dyson and Wright in a review on such phenomena [22], the intrinsic lack of structure can confer functional advantages such as the ability to bind to different targets, perhaps in different conformations. In addition, the rate of binding and thus the efficiency of regulation by the ligand could be enhanced owing to the larger “capture radius” of an unfolded structure—a mechanism that has been proposed recently and termed “fly-casting” [23]. TipAS seems to provide an example where the efficient binding of various antibiotics occurs both by



**Fig. 5** NMR structure of free TipAS (A) and preliminary NMR structure of the TipAS-Promotiocin A complex (B). The antibiotic is shown in space-fill representation, the protein as a ribbon diagram. The N-terminal part of TipAS undergoing the ligand-induced structural and dynamical transition is shown in light gray.

the coupling of binding and folding within a flexible part of the protein and by the recognition via a pre-formed binding surface.

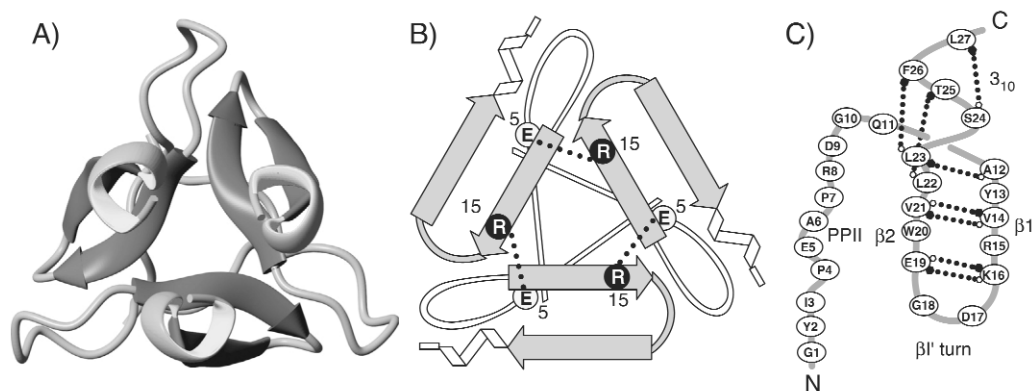
Until the present, the structures of the full TipAL protein and of its complex with DNA have not been solved. Therefore, it is unclear how binding of antibiotics to TipAL induces transcription of the TipA proteins. Our results indicate that only in response to ligand binding, the N-terminal third of TipAS adopts a structure that is stable on the millisecond time scale. Analogous changes as for TipAS are observed in the spectra of the full TipAL protein upon drug binding (M. Allan, unpublished results). Thus, a similar transition from a flexible conformation to an all-helical structure occurs within a part of TipAL that connects the N-terminal DNA-binding domain to the antibiotic recognition domain. It is expected that the structural transition of this linker exerts forces onto the DNA-binding domain that change its structure and that lead to increased affinity for the DNA promoter and ultimately to increased transcription and translation.

In the example of TipAS, much information on the ligand-induced folding is provided by solution NMR. Yet, for a complete understanding of the binding–folding event in terms of energetic and thermodynamic parameters, an even more extensive description would be desirable. Thus, for example, the time scale of the dynamical description by the  $^{15}\text{N}$  relaxation parameters covers only the sub-nanosecond regime, and many of the residues in the unfolded N-terminus of TipAS are not accessible owing to extensive chemical exchange broadening. In the following example of the folding and trimerization of foldon (the natural trimerization domain of T4 fibrin), it was possible to take the description further owing to the smaller size of this model system and the use of novel NMR techniques.

### ATOMIC DETAILS OF DISSOCIATION AND THERMAL UNFOLDING OF FIBRITIN FOLDON: A HOMOTRIMERIC PROTEIN FROM T4 PHAGES\*

Fibrin is an elongated, trimeric, 486 amino acid protein of the bacteriophage T4 head with structural, chaperone, as well as sensory functions. The C-terminal part of the fibrin trimer is formed from a trimeric coiled-coil domain, which is terminated by a C-terminal trimeric  $\beta$ -sheet propeller consisting of monomeric  $\beta$ -hairpin segments (Fig. 6). This  $\beta$ -propeller is necessary for the correct folding of the holoprotein [25] and hence has been termed foldon. Residues G457 to L483 (renumbered G1 to L27 for convenience in the present work) are sufficient to form an isolated trimeric foldon structure [26]. Owing

\*This part has been adapted and shortened from Meier et al. [24].



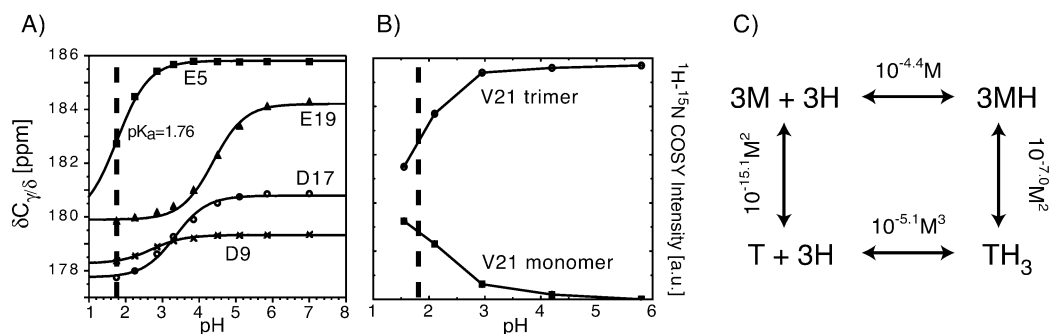
**Fig. 6** The structure of T4 fibrin foldon. (A) NMR structure of the trimer. (B) Scheme of the trimer  $\beta$ -hairpin topology. (C) Monomer H-bond topology.

to the extreme stability of the trimer, foldon can be used as an artificial trimerization inducer or enhancer. Thus, collagen and HIV gp140 trimers have been stabilized via the attachment of the foldon domain.

By applying quantitative solution NMR techniques to this model system, the following questions could be answered: Which forces keep the trimer together? Can the monomer exist as a separate structured entity? If so, what are its structure and thermodynamic properties? Can we learn something from thermal unfolding of the foldon monomer about its structural determinants?

### Foldon trimer-monomer equilibrium

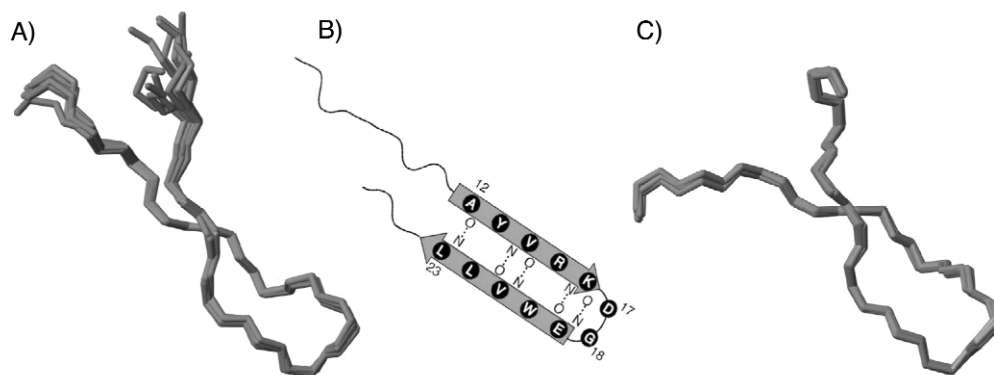
At low pH, the foldon trimer disintegrates into a monomeric (A-state) form that has similar properties as an early intermediate of the trimer folding pathway detected in time-resolved stopped-flow experiments [26]. The formation of this A-state monomer from the trimer can be directly followed by solution NMR [24]. From the pH and concentration dependence of the NMR resonance intensities and positions, the reason for the dissociation of the trimer into the monomer can be traced back to the titration of an intermonomer salt bridge between residues E5 and R15 (Figs. 6B; 7A,B). Based on these data, a detailed quantitative description of the chemical equilibrium between protonated monomeric MH, deprotonated monomeric M, protonated trimeric  $\text{TH}_3$ , and deprotonated trimeric forms T could be derived (Fig. 7C). 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 (16 kJ per mole monomer) at physiological pH and 25 °C.



**Fig. 7** NMR quantification of monomer-trimer equilibrium. (A) pH titration of side-chain carboxylate  $^{13}\text{C}$  resonances showing an extreme pK shift of residue E5 due to its intermolecular salt bridge with residue R15 (see Fig. 6). (B) Intensities of monomer and trimer resonances in  $^1\text{H}$ - $^{15}\text{N}$  COSY spectra as a function of pH. The transition midpoint ( $\sim$  pH 1.8) is close to the pK of the E5 side chain. (C) Chemical equilibrium and respective constants between protonated/deprotonated monomer and trimer species derived from intensity and chemical shift titrations.

### Foldon A-state monomer structure

Despite 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 (Fig. 8A) from a limited set of distance, dihedral angle, and residual dipolar coupling restraints [24]. The structure is well defined for residues G10–L23 representing the hairpin and a small, kinked N-terminal extension (G10, Q11). Outside of the hairpin, i.e., in the N- and C-terminal parts, the structure is largely disordered (Fig. 8). Apparently, this is caused by the lack of stabilizing secondary, tertiary, and quaternary contacts.



**Fig. 8** Solution structure of the A-state foldon monomer. (A) Backbone representation of the 10 lowest energy structures of the foldon monomer (PDB-code 1U0P). (B) H-bond topology of the A-state monomer. For comparison, the hairpin part of the 10 lowest energy structures of foldon within the native trimer (1RFO) is shown in (C).

Within the hairpin of the foldon monomer structure, the  $\beta$ -sheet H-bonds show regular in-line geometry (Fig. 8B) that is very similar to the hairpin conformation within the intact trimer (Fig. 8C). Apparently, no further tertiary monomer or quaternary trimer interactions are required to form the hairpin. Thus, foldon has an autonomously folding  $\beta$ -hairpin motif. Most likely, the foldon A-state structure resembles the monomeric on-path folding intermediate that can be detected in time-resolved



stopped-flow fluorescence experiments [26]. Apparently, the monomeric intermediate is stabilized at low pH by weakening the association reaction. Thus, the isolated foldon monomer with the intact hairpin appears as a stable, low-energy intermediate in the folding pathway of the trimer.

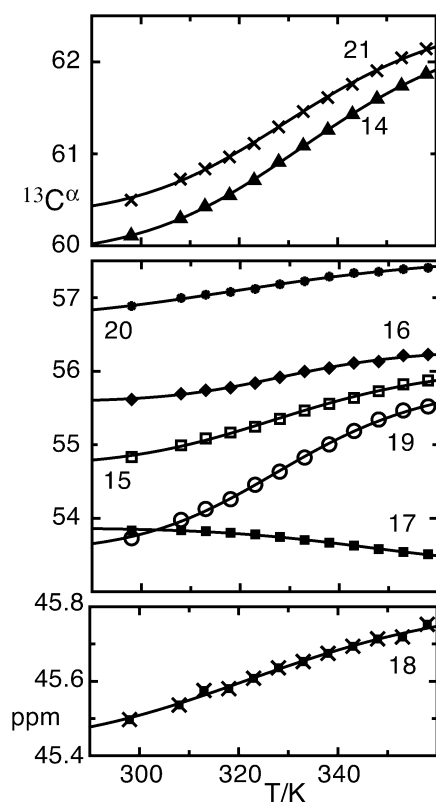
### Quantitative analysis of thermal $\beta$ -hairpin unfolding

In order to learn more about the particular tendency of foldon to form a stable  $\beta$ -hairpin, the characterization of the A-state monomer was taken further by analyzing its thermal stability in a residue-specific manner both by chemical shifts and a new approach using RDCs [24].

#### *Thermal $\beta$ -hairpin unfolding characterized by chemical shifts*

The thermal equilibrium unfolding of the foldon monomer was followed simultaneously by  $\delta^1\text{H}^\alpha$ ,  $\delta^{13}\text{C}^\alpha$ ,  $\delta^{15}\text{N}$ , and  $\delta^1\text{H}^{\text{N}}$  chemical shifts. Whereas  $\delta^1\text{H}^\alpha$  and  $\delta^{13}\text{C}^\alpha$  primarily report on backbone and to a lesser extent on side-chain conformations,  $\delta^1\text{H}^{\text{N}}$  and  $\delta^{15}\text{N}$  are also very strongly influenced by the formation of H-bonds. All of the monitored chemical shifts within the hairpin show a cooperative two-state transition behavior upon thermal unfolding (Fig. 9). As the two states are in fast chemical exchange, observed shifts are a weighted average of both, folded and unfolded states. This allows the determination of their populations. The temperature dependence of the populations can then be fitted to populations obtained from the Gibbs–Helmholtz equation (Fig. 9, solid lines). Since the heat capacity difference  $\Delta C_p$  is usually very small for the thermal denaturation of  $\beta$ -hairpins [27], this parameter was approximated as  $\Delta C_p = 0$  to minimize errors in fitting of the other thermodynamic parameters. The fit of the thermal  $^1\text{H}^\alpha$  and  $^{13}\text{C}^\alpha$  chemical shift transitions yields apparent average folding enthalpies  $\Delta H^0 = -50.7 \pm 5.2$  kJ/mol and folding entropies  $\Delta S^0 = -152 \pm 16.4$  J/(mol·K) for the hairpin strand residues Y13–K16 and E19–L22. This indicates that the foldon hairpin is populated to almost 90 % at room temperature in the A-state. Thus, the hairpin is a highly stable structure with thermodynamic properties in aqueous solution comparable to designed hairpins stabilized by the addition of 50 % methanol or 30 % trifluoroethanol.

Melting temperatures calculated as  $T_m = \Delta H^0/\Delta S^0$  have values of  $T_m = 332.6 \pm 2.7$  K for strand residues Y13–K16 and E19–L22. In contrast, the melting temperatures for the  $\beta$ -turn resonances of D17 and G18 are increased by about 10 K, as apparent from the behavior of the  $^{13}\text{C}^\alpha$  shift of D17 (Fig. 9) and the  $^1\text{H}^\alpha$  shifts of D17 and G18 (not shown). This finding is interesting, since it hints at a particular stability of the turn configuration. Hairpin nucleation could start from such a stabilized turn as predicted by kinetic [28] and thermodynamic [29] zipper models of  $\beta$ -hairpin folding. Further evidence for the particular stability of the turn was found from the following RDC data.



**Fig. 9** Residue-specific thermal transitions in the hairpin as monitored by backbone chemical shifts of  $^{13}\text{C}^\alpha$  nuclei. Solid lines are fits to the Gibbs–Helmholtz equation assuming two-state exchange.

#### *Thermal $\beta$ -hairpin unfolding characterized by residual dipolar couplings*

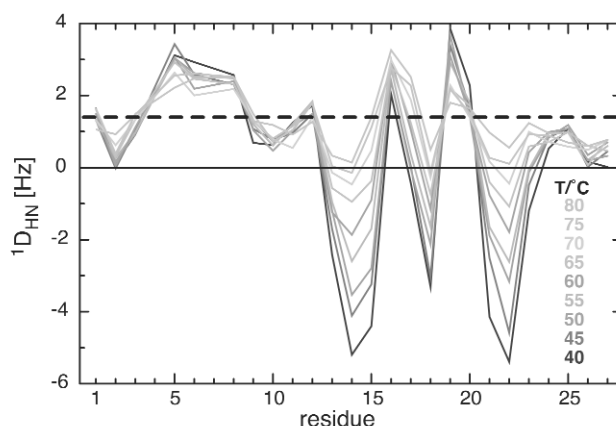
Chemical shifts and NOEs are widely used to characterize changes in local structures during folding transitions of polypeptides. However, owing to the complicated structural dependencies of chemical shifts and the influence of local dynamics and spin diffusion on NOEs, the variations in these NMR parameters are in general difficult to relate to precise changes of geometric parameters. More recently, the weak alignment of proteins in liquid solution has become a tool to directly monitor average net orientations and order parameters of individual bonds by residual dipolar couplings (RDCs) [9]. Albeit biomolecules orient already to some extent in magnetic fields owing to their intrinsic magnetic susceptibility, by far stronger alignments, and, thus, larger RDCs can be obtained by dissolving the biomolecules in anisotropic liquid phases. The introduction of inert orienting media, such as strained polyacrylamide gels [30,31], 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.

A characterization of folding phenomena based on RDCs offers various advantages: first, the measured RDC  $D_{IS}$  between two nuclei  $I$  and  $S$  represents a straightforward ensemble and time average over the relative nuclear conformation in the solution:

$$D_{IS} = \frac{-\gamma_I \gamma_S \hbar \mu_0}{4\pi^2 r^3} \left\langle \frac{3 \cos^2 \theta - 1}{2} \right\rangle = D_{IS}^{\max} \langle P_2(\cos \theta) \rangle \quad (1)$$

where  $\theta$  represents the angle between the internuclear distance vector and the external magnetic field,  $r$  is the internuclear distance (assumed as constant, e.g., within a stable chemical bond), and  $P_2$  is the second-order Legendre polynomial. Thus, the RDCs correspond to well-defined order parameters. [Note that  $D_{IS}$  vanishes in a isotropic solution, since the isotropic time and ensemble average  $\langle P_2(\cos\theta) \rangle$  vanishes. Hence, the necessity to use orienting media for making RDCs observable.] Second, long-range information on the relative orientation of different internuclear vectors is contained in the RDC data, because the angle  $\theta$  is determined relative to a common external direction, i.e., the magnetic field [9]. Thus, non-local order can also be monitored during the folding transition. Third, RDCs can be measured with high sensitivity and accuracy for a large number of sites. Therefore, a detailed description of the folding transition should be possible.

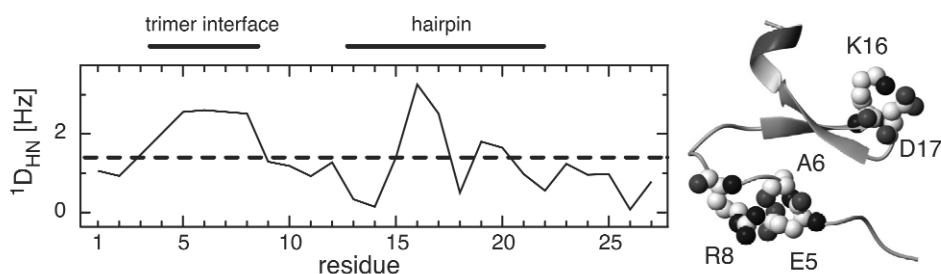
In order to characterize the thermal unfolding of the foldon A-state monomer, the temperature dependence of  $^1D_{NH}$  and  $^1D_{C\alpha H\alpha}$  RDCs was determined in strained polyacrylamide gels. Figure 10 shows the  $^1D_{NH}$  data for temperatures between 40 and 80 °C. At low temperatures, large RDCs are observed within the hairpin region, corresponding to a well-ordered structure. At temperatures above the melting transition, these RDCs converge to small, but non-zero values. The unfolded N- and C-terminal residues outside of the hairpin region all have small, non-zero RDC values indicative of random structures across the entire temperature range. The temperature dependence of the RDCs in the hairpin is quite similar to the chemical shift behavior. Fitted to two-state transitions, the average RDC transition temperatures are close to the chemical shift melting temperatures. However, lower transition temperatures are evident for  $^1D_{NH}$  values from the donor and acceptor peptides planes of the two strong H-bonds between residues V14 and V21. This could be an indication that these H-bonds are destabilized before larger conformational changes of the backbone become detectable in the  $^1H^\alpha$  and  $^{13}C^\alpha$  chemical shifts.



**Fig. 10** Temperature dependence of  $^1D_{NH}$  RDCs in the foldon A-state between 40 and 80 °C. The RDCs are a measure of residue-specific order at the various temperatures. The average of  $^1D_{NH}$  at 80 °C is indicated as a dashed line.

Above the melting transition at 80 °C, the all-residue RDC average deviates from zero. The average amounts to 1.3 Hz for  $^1D_{NH}$  (Fig. 10) and  $-5.0$  Hz for  $^1D_{C\alpha H\alpha}$  (not shown), respectively. Apparently, the orientational distribution of the internuclear N-H<sup>N</sup> or C<sup>α</sup>-H<sup>α</sup> vectors along the polypeptide chain is nonisotropic even above the melting transition, such that  $\langle P_2(\cos\theta) \rangle$  does not vanish. This behavior has been explained recently for weakly aligned, nonspherical random polymer chains [32]. The restrictions of bond movements owing to the covalent geometry of the oriented random polymer lead to such orientational preferences and nonvanishing RDCs for individual bond vectors even when the torsion angles along the polymer chain adopt random conformations.

At 80 °C, individual  $^1D_{\text{NH}}$  (Fig. 11) and  $^1D_{\text{C}\alpha\text{H}\alpha}$  (not shown) RDCs show considerable variations around the respective average values, which are significantly larger than the reproducibility error of the experiments (e.g., <0.3 Hz for  $^1D_{\text{NH}}$ ). These deviations reveal residue-specific orientational order within the molten peptide chain. Thus, for example,  $^1D_{\text{NH}}$  RDC values for residues E5 to R8 and for K16 and D17 are considerably larger than the average (Fig. 11). This directly indicates that these residues have orientational distributions which are different from the other parts of the molten hairpin structure. The deviations for residues E5 to R8 can be explained by their proximity to the two proline residues P4 and P7, which restrict the flexibility of the polypeptide backbone at these positions. On the other hand, the particular thermal stability of the conformation of near-turn residues K16 and D17, which was already detected in the chemical shift data, gives very strong support for the idea that the turns of  $\beta$ -hairpin structures are particularly stabilized such that hairpin formation could start from the turn as a nucleation site.



**Fig. 11** Deviations of  $^1D_{\text{NH}}$  RDCs of the foldon A-state monomer from their average (dashed line) at 80 °C. The deviations indicate residue-specific order above the melting transition (see text). Residues that show particularly strong deviations are indicated in the structure of the monomer (right) in space-fill representation.

A more detailed analysis of the temperature dependence of RDCs can be found in [24]. This analysis shows that local-order parameters can be derived from the RDCs in eq. 1 at the onset of the melting transition. These local-order parameters decrease uniformly for the hairpin strand residues. In contrast, much smaller changes are observed at the turn in agreement with the other observations of its particular thermal stability. The RDC-derived order parameters are ensemble and time averages up to the inverse of the RDC constant,  $D_{\text{IS}}^{-1}$ , i.e., tens of milliseconds. Thus, they cover a much larger time range than the  $\{^1\text{H}\}$ - $^{15}\text{N}$  NOE-derived order parameters. In principle, it should be possible to derive such local-order parameters from the RDCs for a large number of internuclear vectors during the entire melting transition. This would make it possible to determine exactly the entropic changes for individual protein sites during the melting. In practice, the problem is underdetermined from the number of presently determined  $^1D_{\text{NH}}$  and  $^1D_{\text{C}\alpha\text{H}\alpha}$  RDCs. However, this deficit can potentially be overcome by the detection of additional RDCs between other backbone and side-chain nuclei or by using different alignment conditions. This may ultimately lead to a complete quantification of local order in folding/unfolding transitions.

### SCALAR COUPLINGS ACROSS H-BONDS AS A PRECISE MONITOR OF THE THERMAL STABILITY OF INDIVIDUAL PROTEIN H-BONDS

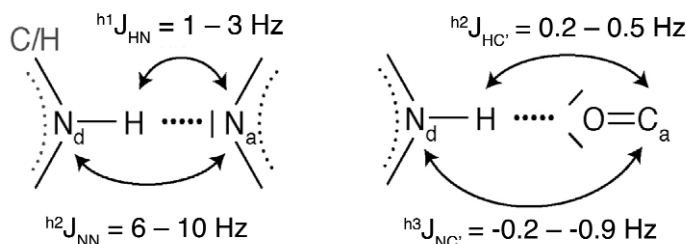
As a final example of modern NMR techniques applied to the biological interactions, we show the determination of the temperature stability of individual protein H-bonds by the detection of electron-mediated scalar couplings. NMR is widely used to probe biomolecular H-bonds because of its ability to resolve individual H-bonding groups and because of the large effects of H-bonding on NMR observables such as chemical shifts, coupling constants, hydrogen exchange rates, and fractionation factors. The discovery that scalar couplings can be commonly observed between magnetic nuclei on both sides

of a H-bond in regular secondary structure elements of nucleic acids [11,12] and proteins [14,15], as well as in certain other H-bonds of biomolecules and smaller chemical compounds [13,33] has added another, very direct effect to this arsenal of individually observable H-bond parameters. The subject has been reviewed in [16,17]. H-bond couplings can be used to identify donor and acceptor groups in individual H-bonds from COSY experiments. Because of their strong dependence on H-bond distances and angles, the sizes of the couplings provide very sensitive measures of the H-bond geometries.

Figure 12 shows the most useful H-bond couplings in the canonical H-bonds of nucleic acids and proteins. Note that the symbol  ${}^h\text{J}_{\text{AB}}$  is used for trans H-bond scalar couplings between nuclei A and B in order to emphasize that one of the  $n$  bonds connecting the two nuclei in the chemical structure is actually a H-bond. The H-bond couplings arise from electronic overlap between the hydrogen and the acceptor orbitals. To a first approximation, the coupling size is proportional to the square of the electronic overlap integral [34,35]. This implicates an approximately exponential dependence of the coupling size on the donor–acceptor distance and also certain dependencies on the angles that describe the H-bond geometry [35]. A salient feature is that straight H-bonds give the largest couplings. Neglecting such angular dependencies, an exponential distance dependence on the nitrogen–oxygen distance  $d_{\text{NO}}$  was determined in proteins by Bax and coworkers [36] based on experimental data for  ${}^{\text{h}^3}\text{J}_{\text{NC}'}$  couplings and the coordinates of an X-ray structure:

$${}^{\text{h}^3}\text{J}_{\text{NC}'} = -5.9 \cdot 10^4 \text{ Hz exp}(-4d_{\text{NO}}/\text{\AA}) \quad (2)$$

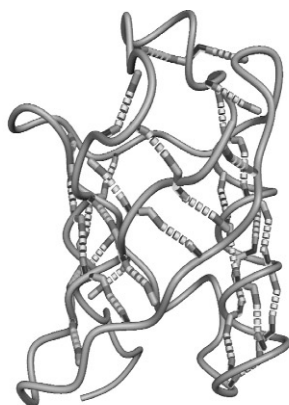
By their strong dependence on the H-bond geometry, the H-bond scalar couplings are one of the most sensitive parameters available to characterize subtle structural rearrangements, which occur under changed conditions in biomacromolecules. In particular, we have used this parameter to follow changes in H-bond geometry during peptide folding [37], ligand binding [38], chemical destabilization [39], hydrogen/deuterium exchange [40], and thermal unfolding [41].



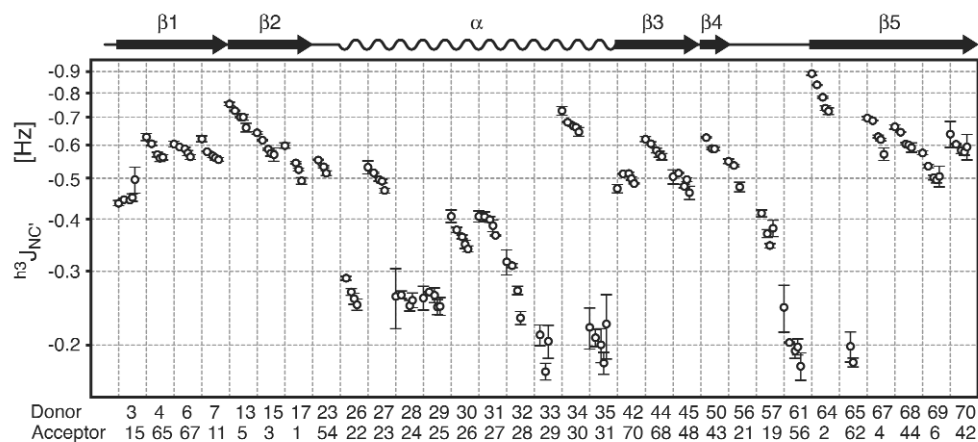
**Fig. 12** Scalar couplings across typical H-bonds in nucleic acids (left) and proteins (proteins). The size of the couplings refers to  ${}^{15}\text{N}$  and  ${}^{13}\text{C}$  isotope labeled donor and acceptor groups.

In the following, we discuss some of the results of the thermal unfolding of the small protein human ubiquitin [41]. Figure 13 shows the ubiquitin H-bond network as detected by COSY experiments using the  ${}^{\text{h}^3}\text{J}_{\text{NC}'}$  couplings. A plot of the size of  ${}^{\text{h}^3}\text{J}_{\text{NC}'}$  scalar couplings measured for temperatures from 5 to 65 °C is shown in Fig. 14. The temperature increase induces a decrease in the absolute size of H-bond coupling constants. This experimental finding directly shows that the H-bond network expands at higher temperature. Assuming a sole exponential dependence on the H-bond distance according to eq. 2, the derivative of  $\ln|{}^{\text{h}^3}\text{J}_{\text{NC}'}|$  with respect to temperature should be a measure of the linear thermal expansion coefficient ( $\alpha_{\text{L}} = r_{\text{NO}}^{-1} \partial r_{\text{NO}}/\partial T$ ) of the hydrogen bond. Assuming an average  $r_{\text{NO}}$  value of 3.0 Å,  $\alpha_{\text{L}}$  can be estimated as

$$\alpha_{\text{L}} \approx -\frac{1}{12} \frac{\partial \ln |{}^{\text{h}^3}\text{J}_{\text{NC}'}|}{\partial T} \quad (3)$$



**Fig. 13** The structure of human ubiquitin shown as a ribbon diagram. H-bonds directly detected by  $^3J_{NC}$ -scalar couplings are shown as broken cylinders.



**Fig. 14** Temperature dependence of  $^3J_{NC}$  coupling constants in human ubiquitin vs. amino acid sequence. For each H-bond,  $^3J_{NC}$  couplings are depicted in equidistant spacing from left to right for the five temperatures 5, 25, 45, 55, and 65 °C. Data are plotted on a logarithmic scale. Corresponding secondary structure elements are shown at the top. Reprinted with permission from [41].

Using this expression and slopes  $\partial \ln |^3J_{NC}| / \partial T$  calculated from a linear regression of the data in Fig. 14, the average linear expansion coefficient  $\alpha_L$  amounts to  $(1.7 \pm 0.2) \cdot 10^{-4}/K$  for all backbone H-bonds observed in ubiquitin. No significant differences in  $\alpha_L$  were found for  $\alpha$ -helical or  $\beta$ -sheet conformations. The linear thermal expansion coefficient for H-bonds agrees remarkably well with values of  $(5.2 \pm 0.5) \cdot 10^{-4}/K$  for the volume expansion coefficient  $\alpha_V = V^{-1} \partial V / \partial T \approx 3\alpha_L$  for staphylococcal nuclease [42] and  $\alpha_V$  values in the range of  $5\text{--}6 \cdot 10^{-4}/K$  for five other proteins [43]. Thus, the entire protein and the H-bonds expand at the same rate, whereas thermal expansion coefficients for covalent bonds are orders of magnitude smaller, e.g.,  $\alpha_L(\text{diamond}) = 1.2 \cdot 10^{-6}/K$  [44]. This is consistent with the notion that the thermal expansion of proteins is mostly due to the weakening of noncovalent chain–chain contacts such as H-bond and side chain to side-chain interactions. It is also interesting to note that the temperature-induced relative changes of  $^2J_{NN}$  couplings in nucleic acids are very similar to the values for  $^3J_{NC}$  couplings in proteins [17]. As both couplings show very similar exponential distance dependencies, this indicates that H-bonds in nucleic acids and proteins have a very similar thermal expansion behavior.

A residue-specific analysis of the  $^3\text{J}_{\text{NC}}$  couplings in ubiquitin reveals that not all hydrogen bonds are affected to the same extent by the thermal expansion. Thus, specific regions can be identified that are the least thermostable. This is the case for the N-terminal region in the vicinity of the H-bond E64 → Q2 (Fig. 14), which is located at the beginning of ubiquitin's parallel  $\beta$ -sheet consisting of strands  $\beta 5$  and  $\beta 1$ . The particular thermal instability can be rationalized from the fact that strand  $\beta 5$  is only H-bonded on one side, whereas its opposite side is partially exposed to water. A more detailed discussion is given in [39,41].

## CONCLUDING REMARKS

Progress in modern biology hinges on the quantitative understanding of the biomolecular interactions that make up the complex network of living organisms. The given examples should illustrate how NMR in solution can be used beyond a pure structural description to obtain a more quantitative picture of biomolecular function in terms of free energies, conformational entropies, and other parameters. It is hoped that ultimately such a precise description of single model systems can help to solve the puzzle how all biological phenomena are determined from the linear information contained in the genomes.

## ACKNOWLEDGMENTS

This work was supported by SNF grant 31-43'091.95 to S. G.

## REFERENCES

1. M. P. Williamson, T. F. Havel, K. Wüthrich. *J. Mol. Biol.* **182**, 295–315 (1985).
2. S. W. Fesik and E. R. P. Zuiderweg. *J. Magn. Reson.* **78**, 588–593 (1988).
3. D. Marion, L. E. Kay, S. W. Sparks, D. A. Torchia, A. Bax. *J. Am. Chem. Soc.* **111**, 1515–1517 (1989).
4. S. Grzesiek, J. Anglister, H. Ren, A. Bax. *J. Am. Chem. Soc.* **115**, 4369–4370 (1993).
5. V. Tugarinov, R. Muhandiram, A. Ayed, L. E. Kay. *J. Am. Chem. Soc.* **124**, 10025–10035 (2002).
6. J. Fiaux, E. B. Bertelsen, A. L. Horwich, K. Wüthrich. *Nature* **418**, 207–211 (2002).
7. A. G. Palmer, 3<sup>rd</sup>. *Chem. Rev.* **104**, 3623–3640 (2004).
8. J. R. Tolman, J. M. Flanagan, M. A. Kennedy, J. H. Prestegard. *Proc. Natl. Acad. Sci. USA* **92**, 9279–9283 (1995).
9. N. Tjandra and A. Bax. *Science* **278**, 1111–1114 (1997).
10. A. Bax. *Protein Sci.* **12**, 1–16 (2003).
11. A. Dingley and S. Grzesiek. *J. Am. Chem. Soc.* **120**, 8293–8297 (1998).
12. K. Pervushin, A. Ono, C. Fernandez, T. Szyperski, M. Kainosho, K. Wüthrich. *Proc. Natl. Acad. Sci. USA* **95**, 14147–14151 (1998).
13. I. G. Shenderovich, S. N. Smirnov, G. S. Denisov, V. A. Gindin, N. S. Golubev, A. Dunger, R. Reibke, S. Kirpekar, O. L. Malkina, H.-H. Limbach. *Ber. Bunsenges. Phys. Chem.* **102**, 422–428 (1998).
14. F. Cordier and S. Grzesiek. *J. Am. Chem. Soc.* **121**, 1601–1602 (1999).
15. G. Cornilescu, J.-S. Hu, A. Bax. *J. Am. Chem. Soc.* **121**, 2949–2950 (1999).
16. S. Grzesiek, F. Cordier, A. J. Dingley. *Methods Enzymol.* **338**, 111–133 (2001).
17. S. Grzesiek, F. Cordier, V. Jaravine, M. Barfield. *Prog. NMR Spectr.* **45**, 275–300 (2004).
18. S. Grkovic, M. H. Brown, R. A. Skurray. *Semin. Cell Dev. Biol.* **12**, 225–237 (2001).
19. J. D. Kahmann, H. J. Sass, M. G. Allan, H. Seto, C. J. Thompson, S. Grzesiek. *EMBO J.* **22**, 1824–1834 (2003).
20. G. Cornilescu, F. Delaglio, A. Bax. *J. Biomol. NMR* **13**, 289–302 (1999).

21. L. K. Nicholson, T. Yamazaki, D. A. Torchia, S. Grzesiek, A. Bax, S. J. Stahl, J. D. Kaufman, P. T. Wingfield, P. Y. S. Lam, P. K. Jadhav, C. N. Hodge, P. J. Dommelle, C. H. Chang. *Nat. Struct. Biol.* **2**, 274–280 (1995).
22. H. J. Dyson and P. E. Wright. *Curr. Opin. Struct. Biol.* **12**, 54–60 (2002).
23. B. A. Shoemaker, J. J. Portman, P. G. Wolynes. *Proc. Natl. Acad. Sci. USA* **97**, 8868–8873 (2000).
24. S. Meier, S. Guthe, T. Kiefhaber, S. Grzesiek. *J. Mol. Biol.* **344**, 1051–1069 (2004).
25. A. V. Letarov, Y. Y. Londer, S. P. Boudko, V. V. Mesyanzhinov. *Biochemistry-Moscow* **64**, 817–823 (1999).
26. S. Güthe, L. Kapinos, A. Möglich, S. Meier, S. Grzesiek, T. Kiefhaber. *J. Mol. Biol.* **337**, 905–915 (2004).
27. C. M. Santiveri, J. Santoro, M. Rico, M. A. Jimenez. *J. Am. Chem. Soc.* **124**, 14903–14909 (2002).
28. K. A. Dill, K. M. Fiebig, H. S. Chan. *Proc. Natl. Acad. Sci. USA* **90**, 1942–1946 (1993).
29. D. K. Klimov and D. Thirumalai. *Proc. Natl. Acad. Sci. USA* **97**, 2544–2549 (2000).
30. R. Tycko, F. J. Blanco, Y. Ishii. *J. Am. Chem. Soc.* **122**, 9340–9341 (2000).
31. H. J. Sass, G. Musco, S. J. Stahl, P. T. Wingfield, S. Grzesiek. *J. Biomol. NMR* **18**, 303–309 (2000).
32. M. Louhivuori, K. Fredriksson, K. Paakkonen, P. Permi, A. Annala. *J. Biomol. NMR* **29**, 517–524 (2004).
33. P. R. Blake, J.-B. Park, M. W. W. Adams, M. F. Summers. *J. Am. Chem. Soc.* **114**, 4931–4933 (1992).
34. M. Barfield, A. J. Dingley, J. Feigon, S. Grzesiek. *J. Am. Chem. Soc.* **123**, 4014–4022 (2001).
35. M. Barfield. *J. Am. Chem. Soc.* **124**, 4158–4168 (2002).
36. G. Cornilescu, B. E. Ramirez, M. K. Frank, G. M. Clore, A. M. Gronenborn, A. Bax. *J. Am. Chem. Soc.* **121**, 6275–6279 (1999).
37. V. A. Jaravine, A. T. Alexandrescu, S. Grzesiek. *Protein Sci.* **10**, 943–950 (2001).
38. F. Cordier, C. Wang, S. Grzesiek, L. K. Nicholson. *J. Mol. Biol.* **304**, 497–505 (2000).
39. F. Cordier and S. Grzesiek. *Biochemistry* **43**, 11295–11301 (2004).
40. V. A. Jaravine, F. Cordier, S. Grzesiek. *J. Biomol. NMR* **29**, 309–318 (2004).
41. F. Cordier and S. Grzesiek. *J. Mol. Biol.* **317**, 739–752 (2002).
42. H. Seemann, R. Winter, C. A. Royer. *J. Mol. Biol.* **307**, 1091–1102 (2001).
43. H. J. Hinz, T. Vogl, R. Meyer. *Biophys. Chem.* **52**, 275–285 (1994).
44. D. R. Lide. *CRC Handbook of Chemistry and Physics*, 80<sup>th</sup> ed., CRC Press, Boca Raton (1999).