Folding of the Transcription Factor Brinker

and

Interactions of the Bacterial Second Messenger c-di-GMP

studied by NMR

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Martin Gentner
aus Rheinfelden (Baden), Deutschland

Basel, 2013

Namensnennung-Keine kommerzielle Nutzung-Keine Bearbeitung 2.5 Schweiz

---

Sie dürfen:

das Werk vervielfältigen, verbreiten und öffentlich zugänglich machen

Zu den folgenden Bedingungen:

Namensnennung. Sie müssen den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen (wodurch aber nicht der Eindruck entstehen darf, Sie oder die Nutzung des Werkes durch Sie würden entlohnt).

Keine kommerzielle Nutzung. Dieses Werk darf nicht für kommerzielle Zwecke verwendet werden.

Keine Bearbeitung. Dieses Werk darf nicht bearbeitet oder in anderer Weise verändert werden.

- Im Falle einer Verbreitung müssen Sie anderen die Lizenzbedingungen, unter welche dieses Werk fällt, mitteilen. Am Einfachsten ist es, einen Link auf diese Seite einzubinden.

- Jede der vorgenannten Bedingungen kann aufgehoben werden, sofern Sie die Einwilligung des Rechteinhabers dazu erhalten.
- Diese Lizenz lässt die Urheberpersönlichkeitsrechte unberührt.

Die gesetzlichen Schranken des Urheberrechts bleiben hiervon unberührt.

Die Commons Deed ist eine Zusammenfassung des Lizenzvertrags in allgemeinverständlicher Sprache: http://creativecommons.org/licenses/by-nc-nd/2.5/ch/legalcode.de

Haftungsausschluss:
Die Commons Deed ist kein Lizenzvertrag. Sie ist lediglich ein Referenztext, der den zugrundeliegenden Lizenzvertrag übersichtlich und in allgemeinverständlicher Sprache wiedergibt. Die Deed selbst entfaltet keine juristische Wirkung und erscheint im eigentlichen Lizenzvertrag nicht. Creative Commons ist keine Rechtsanwaltsgesellschaft und leistet keine Rechtsberatung. Die Weitergabe und Verlinkung des Commons Deeds führt zu keinem Mandatsverhältnis.

Quelle: http://creativecommons.org/licenses/by-nc-nd/2.5/ch/ Datum: 3.4.2009

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von

Prof. Dr. Stephan Grzesiek

PD Dr. Daniel Häußinger

Basel, den 26.06.2012

Prof. Dr. Martin Spiess
(Dekan)

Summary

Nuclear magnetic resonance (NMR) spectroscopy is a technique, which allows the non-invasive investigation of structures, dynamics and interactions of biomolecules. The main goal of this thesis was to elucidate the folding mechanism of the transcription factor Brinker and its implications for DNA recognition as well as the characterization of unfolded protein states by NMR. This constitutes the first part of this thesis. The second part describes the quantitative characterization of the intermolecular interactions between monomers of the bacterial second messenger c-di-GMP at physiologically relevant concentrations.

The *Drosophila melanogaster* transcription factor Brinker is a nuclear repressor, which is involved in cellular growth and differentiation. In the absence of DNA, Brinker is completely disordered. However, in the presence of DNA or at low temperatures, the Brinker DNA binding domain (BrkDBD) adopts a well-folded structure. Thus, BrkDBD represents an extreme case of the coupling between binding and folding phenomenon. We have aimed to elucidate this folding mechanism in order to understand its implications for DNA recognition.

From our data, it is clear that the BrkDBD folding energy landscape sharply depends on buffer anion type and concentration. We show that folded BrkDBD always adopts the same structure irrespective of the conditions. Our data indicate α-helical propensity for 3 of the 4 native helices in unfolded BrkDBD, which may serve as initial contact points for DNA recognition. Resonance broadening due to conformational exchange on the micro- to millisecond time scale between folded and unfolded BrkDBD was analyzed by NMR relaxation dispersion experiments indicating a two-state folding mechanism. Only few residues show a different behavior and are all located at the DNA binding interface. This local conformational heterogeneity may be important for DNA recognition. Based on these findings, we propose a mechanism of DNA recognition by BrkDBD, where the electrostatics-driven folding is a key component, accelerating the recognition process (Gentner *et al.*, in preparation). In addition, we have analyzed the side-chain χ_1 rotamer distribution of urea-denatured ubiquitin and protein G, revealing that individual residues show significant deviations from statistical-coil ensemble averages (Vajpai *et al.*, 2010), indicating local bias towards the folded state.

C-di-GMP is a bacterial second messenger, involved in many signaling events. Its most important effect is to trigger the transition from motile to sessile bacterial lifestyles which plays a major role in biofilm formation. In solution, c-di-GMP has been reported to form several oligomers in the presence of monovalent cations, particularly potassium. However, only monomeric and dimeric c-di-GMP have been observed in complexes with proteins or RNA. We have carried out a detailed kinetic and thermodynamic analysis of c-di-GMP polymorphism in the presence of potassium, which showed that predominantly monomers and only few dimers exist at physiological concentrations (Gentner *et al.*, 2012). Additionally, we present NOE and ROE

structural information on c-di-GMP oligomers, which indicate that these are not entirely all-syn and all-anti as opposed to the literature.

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

1. Gentner, M., Allan, M. G., Zaehringer, F., Schirmer, T., and Grzesiek, S.
   Oligomer Formation of the Bacterial Second Messenger c-di-GMP: Reaction Rates and Equilibrium Constants Indicate a Monomeric State at Physiological Concentrations
   Journal of the American Chemical Society 134 (2012) 1019-1029.

2. Vajpai, N., Gentner, M., Huang, J.-r., Blackledge, M., and Grzesiek, S.
   Side-Chain χ_1 Conformations in Urea-Denatured Ubiquitin and Protein G from ^3J Coupling Constants and Residual Dipolar Couplings.
   Journal of the American Chemical Society 132 (2010) 3196-3203.

3. Gentner, M., Schmidt, T., and Grzesiek, S.
   Folding of the Transcription Factor Brinker studied by NMR
   (2013), *in preparation*.

Abbreviations and Symbols

Å	100 pm
AHD-DNA	antennapedia homeodomain-specific DNA
BMRB	Biological Magnetic Resonance Data Bank, www.bmrb.wisc.edu
Brk	Brinker protein
Brk1	Brinker protein, residues 1-101
BrkDBD	Brinker DNA binding domain, residues 43-101
γ	gyromagnetic ratio
cAMP	cyclic adenosine monophosphate
CASP	critical assessment of techniques for protein structure prediction
CD	circular dichroism
cGMP	cyclic guanosine monophosphate
c-di-AMP	cyclic diadenylic acid
c-di-GMP	cyclic diguanylic acid
CPMG	Carr-Purcell-Meiboom-Gill, referring to 180 degree (spin-echo) pulse trains
Da	Dalton, g mol^{-1}
δ	chemical shift, ppm
$\lvert\Delta\omega\rvert$	chemical shift difference between two states, obtained from relaxation dispersion experiments
DOSY	diffusion-ordered spectroscopy
DTT	dithiothreitol
FID	free induction decay
G-quartet	Guanine quartet / Guanine quadruplex
GMP	guanosine monophosphate
GTP	guanosine triphosphate
h	Planck constant, $6.626 * 10^{-34}$ J s
$\hbar$	$h/(2\pi)$
HSQC	heteronuclear single quantum coherence
IDP	intrinsically disordered protein
IMP	inosine monophosphate
J	scalar coupling

k_B	Boltzmann constant, $1.381*10^{-23}$ J K^{-1}
K_d	dissociation constant
k_{ex}	exchange rate, obtained from relaxation dispersion experiments
M	mol l^{-1}
MD	monomer/dimer fast exchange resonance
MD simulation	molecular dynamics simulation
MTSL	S-(2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
O_a/O_a’	c-di-GMP octamers in anti conformation
O_s	c-di-GMP octamers in syn conformation
PAGE	polyacrylamide gel electrophoresis
p_b	relaxation dispersion-derived excited state population
PDB	RCSB Protein Data Bank, www.rcsb.org/pdb
ppm	parts per million
R_1, R_2	longitudinal and transverse relaxation rate
RDC	residual dipolar coupling
ROE	rotating frame Overhauser effect
ROESY	rotating frame Overhauser effect spectroscopy
τ_{CPMG}	time between consequtive 180 degree CPMG pulses
t_1, t_2, ...	chemical shift evolution times
T_1, T_2	longitudinal and transverse relaxation time
T_a	c-di-GMP tetramers in anti conformation
T_s	c-di-GMP tetramers in syn conformation
TCEP	tris(2-carboxyethyl)phosphine
θ_{res}	mean residue ellipticity, deg cm^2 $dmol^{-1}$
TOCSY	total correlation spectroscopy
TRIS	tris(hydroxymethyl)aminomethane
UniProtKB	UniProt Knowledgebase, www.uniprot.org
WATERGATE	water suppression by gradient tailored excitation
wt	wild type

Atoms and angles are referred to according to IUPAC nomenclature.

Contents

Chapter 1

NMR Analysis of Protein Folding and Unfolded States

1.1 Introduction

Protein folding is the process during which an unfolded polypeptide chain adopts a well-defined, functional, three-dimensional structure, referred to as the native state[53]. Clearly, the folding process must be guided, since random exploration of the whole conformational space would require an astronomic amount of time (Levinthal's paradox[142]). Even though folding of various proteins has been studied in detail, a complete general understanding has not yet been achieved[52]. Particularly with respect to misfolding-related diseases, it has become a very important issue to understand all factors which decide whether a protein folds into the native state or adopts a disease-related structure[53,122].

1.1.1 Current understanding of protein folding

The current view is that protein folding can be described as a diffusion of a polypeptide along a free energy landscape, which eventually converges to the absolute minimum, the native state[53]. This energy landscape is assumed to have a funnel shape[24,184,268]. The folding funnel is typically shown as 2D[122] or 3D[53] plot, depicting the free energy as a function of conformational entropy and the free enthalpy (Figure 1.1). Sometimes, the x-axis depicts residue contacts, which increase towards the native state[53].

The remarkable speed of protein folding can thus be explained by the funnel shape of the free energy surface, which restricts the accessible conformational space. On the atomic level, this has been described as a series of local optimizations followed

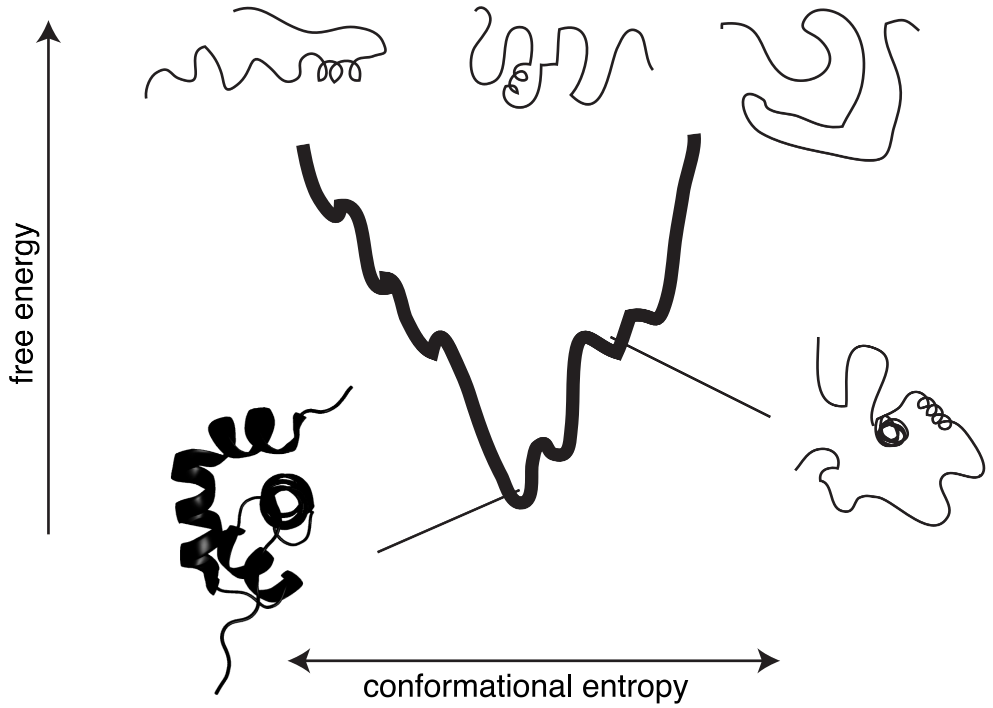


Figure 1.1 Example of a 2D protein folding funnel. The free energy is depicted as a function of the conformational entropy. In 3D funnels, the free enthalpy is on the third axis. The narrow global minimum in this Figure corresponds to the native state, while the broad top corresponds to the unfolded state ensemble, illustrated by the structures. In addition, the structure of a folding intermediate (local minimum) is shown. Folding can start from any point in the unfolded state, leading to different trajectories of single molecules until each reaches the bottom.

by assembly into the folded state rather than a single global search for the folded state[49].

The folding funnel picture takes the multiplicity of states into account, which are sampled during protein folding. While the broad top of the folding funnel corresponds to an ensemble of unfolded structures, the narrow bottom represents the native, folded state[53] (Figure 1.1). Local minima represent folding intermediates (e.g. molten globule structures, which possess secondary, but lack a stable tertiary structure[135]) or misfolded states, while transition states of protein folding are located at saddle points[53]. It is noteworthy that the funnel model implies that various trajectories converge to the native state, indicating alternate folding routes.

The detailed shape of such a free energy surface depends on the individual protein and its folding mechanism as well as on the surrounding conditions. Therefore, the general understanding of protein folding and the design of prediction methods requires elucidation of individual mechanisms for different types of folders (e.g. two-state or multiple-state) and a detailed description of their free energy surface.

The existence of misfolding and metamorphic proteins[23] indicate that the folding funnel may not be rigid, even for a single amino acid sequence. Ambient conditions including temperature, pressure and interactions with other molecules (e.g. chaperones, prosequences or denaturants) can alter the energy landscape. Such events must play a role in misfolding, since for example amyloid fibrils (based on the formation and assembly of intermolecular parallel and/or anti-parallel β-sheet stacks) are formed instead of the native structure under certain conditions[53,181,282].

Simpler examples are proteins in the presence of effector molecules which trigger conformational changes such as helix formation[112]. Therefore, modification of the free energy landscape does not necessarily cause misfolding, it can also serve as regulator of function.

The conformational diversity of single polypeptides also explains why computer-based structure prediction algorithms often fail to predict the native three-dimensional structure from the amino acid sequence[49]. To improve prediction algorithms and gain further understanding, biennial CASP (Critical Assessment of Techniques for Protein Structure Prediction) events have been organized, where in a global blind test, the sequences of already solved, but not yet released structures are predicted[171]. From these events, the most significant successes have been the detection of distant homologs and the deduction of models for difficult targets[49].

Even though computational approaches have improved over time[49] and for some proteins the predictions nearly perfectly match the experimental structures[145,220], multidomain proteins or very large proteins without known homologs still remain very challenging for sequence-based structure prediction.

1.1.2 Intrinsically disordered proteins and unfolded state ensembles

Intrinsically disordered proteins (IDPs) are proteins which are completely or largely unfolded in the absence of a specific binding partner[60,62]. Thus, either a broad, shallow region of the energy landscape or a rather rugged energy surface without deep minima[249] is occupied, which is in contrast to the conventional picture, where a single correctly folded structure is assumed to be a prerequisite for functionality (structure-function paradigm). Binding of a specific partner molecule to the IDP modulates the free energy surface, often leading to the formation of a folded conformation, which corresponds to a rather narrow free energy minimum in the folding funnel.

The physiological role of this mechanism is under debate, but it has been proposed that the unfolded polypeptide has a higher capture radius to detect the binding partner ('fly-casting mechanism'[223]), which accelerates the search process. However, a recent coarse-grained molecular dynamics (MD) simulation study concluded that, for one particular complex of an IDP with its interaction partner (pKID-KIX), rather low free energy barriers than an increased capture radius lead to faster recognition accompanied by folding[98]. In this case, the effects of the increased capture radius and the slower diffusion due to the extended conformation nearly compensated.

Further features of IDPs are the ability to interact with several targets and the accessibility of posttranslational modification sites[84]. In addition, unfolded polypeptide chains are subject to faster proteolytic degradation[84,248]. This has been proposed to enable tight control of the abundance of IDPs both *in vitro* and in cells[80,84,248].

Interestingly, IDPs are very abundant in eukaryotes compared to other organisms[248]. An analysis of the various types of proteins, which are partially or fully disordered revealed that the majority is involved in interactions with other molecules[56,57,93,248]. Virtually no enzymes are found in this category of proteins, which can be rationalized by the requirement of a certain geometry for enzymatic functionality.

Even though the detailed mechanisms of coupling between binding and folding may not be identical for different IDPs, the existence and particularly the involvement of IDPs in interactions with binding partners[56,57,93,248] indicate that the modification of the folding funnel upon binding has been established as a tool in many organisms. Two prominent mechanisms are induced fit (folding after binding) and conformational selection (binding after folding)[128]. While the first mechanism implies that conformational changes are a consequence of the binding event, conformational selection means that binding occurs only to the correctly folded partner. Even though induced fit has been described to be faster and therefore more common[128], a combination of both mechanisms is also possible.

It is important to investigate the detailed interplay between certain proteins (e.g. IDPs) and their cognate binding partners, since this contributes to the general understanding of the coupling between binding and folding phenomenon[128,271,272]. Complete analysis of the folding of a protein includes elucidation of the energy landscape at all levels, ideally under one particular set of conditions. These levels are the folded and the unfolded states, as well as transition states and intermediates, if present. This is extremely challenging, since folding intermediates and transition states are often only transiently and sparsely populated and unfolded states are very heterogeneous due to the large size of the structural ensemble.

1.1.3 NMR methods to study protein folding and unfolded states

Many fast experimental techniques such as fluorescence, fluorescence energy transfer, circular dichroism or small-angle X-ray scattering can be used to study folding and unfolding events below the time scale of seconds[196,205]. Since individual measurements of optical techniques can in principle be achieved in femtoseconds, the time-resolution of these analytical methods is often only limited by the dead time of the experimental setup[205]. The dead time has been reduced to only few tens of microseconds in modern mixing devices[205]. This allowed for example to study early folding and unfolding events in apomyoglobin[274], which are beyond the reach of conventional stopped-flow or quenched-flow techniques. However, these techniques allow only limited structural resolution.
Their time-resolution is a significant advantage compared to the intrinsically slow high-resolution (atomic scale) methods such as X-ray cristallography or NMR spectroscopy. The latter can be applied to study slow folding (e.g. due to proline isomerization) in real-time when combined with rapid mixing techniques[10–13]. Although the required experimental time for multidimensional NMR spectroscopy has been reduced in recent years and could be successfully applied to study folding intermediates within several minutes[202], it is still much slower than the aforementioned low-resolution techniques.

The following paragraphs give a short overview of NMR methods that can be used to characterize various parts of the folding funnel at atomic resolution. Several NMR methods even allow to investigate low-populated folding intermediates and unfolded state ensembles.

NMR chemical shift analysis
All nuclei, except those with both even atomic mass and even atomic number, possess a nuclear spin I different from 0, which gives rise to a magnetic moment. The size of this magnetic moment is given by $\gamma \cdot \mathrm{I}$, where the proportionality constant γ is the nucleus-dependent gyromagnetic ratio. The possible spin states (2I+1) are determined by the magnetic quantum number (I). These states are degenerate (of equal energy) in the absence of a magnetic field. However, the possible spin states are of different energy in the presence of a magnetic field, which can be visualized as preferred orientations of the magnetic moments in the view of classical mechanics. For example spin-1/2 nuclei can align parallel and antiparallel (which are the

two possible spin states) to the external magnetic field (B_0), which is defined to point in positive z-direction. No preference for certain spin orientations is present in the transverse plane, which are thus randomly distributed to give a zero vector sum. The magnetic field imposes a torque on individual nuclear magnetic moments, which causes precession around the magnetic field direction, with a frequency referred to as Larmor frequency[38]:

$$\nu = \frac{\gamma \cdot B_0}{2\pi}$$

Furthermore, the energy difference between parallel and antiparallel spin orientations for spin-half nuclei in an external magnetic field depends on the field strength and the gyromagnetic ratio:

$$\Delta E = h \cdot \nu = \frac{h}{2\pi} \cdot \gamma \cdot B_0 = \hbar \cdot \gamma \cdot B_0$$

where h is the Planck constant. For spin-half nuclei, the parallel orientation is slightly energetically favored over the antiparallel orientation, provided γ is positive. The resonance condition for NMR is the transition of spin states, which happens if the frequency of an external source matches the Larmor frequency.

The chemical shift (δ, in ppm) of a nucleus depends on the difference between its actual Larmor frequency (ν) and a reference resonance frequency (ν_0) and is defined as:

$$\delta = \frac{\nu - \nu_0}{\nu_0} \cdot 10^6 \text{ppm}$$

The factor of 10^6 is introduced, since the frequency differences are on the order of several hundred to few thousand Hertz, while the absolute frequency is on the order of several hundred Megahertz for protons. Such resonance frequency differences are caused by electron motions, which lead to different magnetic shielding (induction of a magnetic field, opposed to the external magnetic field) of individual nuclei depending on the surrounding electron density[31].

Therefore, chemical shifts are an excellent probe of the local, structure-dependent environment of nuclei. The resonance dispersion indicates whether a protein is folded (broad dispersion) or unfolded (narrow dispersion). For unfolded proteins, the lack of dispersion of ^{1}H nuclei can usually be overcome by the application of multidimensional NMR, which exploits the broader resonance dispersion of the heteronuclei (e.g. ^{13}C and ^{15}N)[59].

Chemical shift deviations from random coil shifts, which are typically recorded for single amino acids in the center of denatured tetra-, penta- or hexa-peptides, indicate secondary structure propensities[7,18,21,97,116,125,163,197,203,219,233,240,267]. Such secondary chemical shifts can be obtained for both folded and unfolded states. Typically, the secondary chemical shifts have a much smaller amplitude for unfolded states due to conformational averaging[59]. Due to the more pronounced sequence dependence of the chemical shifts in unfolded states, correction factors for nearest neighbor effects have been introduced[218,238].

While large positive ^{13}Cα and ^{13}CO secondary chemical shifts indicate α-helical structure in folded proteins, local structural propensities may still be indicated in partially or completely unfolded proteins, despite the smaller magnitude. Averaging of secondary shifts over a sequence of residues allows to estimate the percentage of α-helical conformers[63]. Therefore, it is important to obtain random coil shifts which have been determined at conditions that are very similar to the actual experimental conditions[59]. Additionally, modifications of the conformational energy landscpape can be monitored by chemical shift changes depending on the surrounding conditions such as pressure[3], heating[41], or denaturant concentration[207].

Recently, significant progress in computational structure predictions, solely based on chemical shifts, has been achieved[30,221,266]. Chemical shift-based structure prediction programs such as CHESHIRE[30], CS-Rosetta[221] or CS23D[266] rely on matching experimental chemical shifts of various nuclei (typically ^{1}HN, ^{15}N, ^{13}CO, ^{13}Cα, and ^{1}Hα) with database protein fragments of similar chemical shifts. This approach is particularly powerful for low-populated protein states, which cannot be investigated by other techniques at atomic resolution (see below).

Nuclear Overhauser effect

Nuclear Overhauser effects (NOEs) allow to detect nuclei within a 5 Å sphere of the observed nucleus due to their distance dependence (r^{-6})[58]. Typically, long-range NOEs, which correlate residues which are far apart in the sequence but close in space, are indispensable for the determination of three-dimensional protein structures[59]. However, for unfolded states, such long-range NOEs are usually not detected due to the conformational heterogeneity as well as too low populations of transiently structured forms. Local, short-range and medium-range NOEs can indicate α-helical or turn-like conformers also for unfolded states[59,182].

Hydrogen exchange experiments

The measurement of amide proton hydrogen exchange allows to assess the integrity of structural domains under various conditions at atomic resolution[58]. Amide protons in flexible parts of the protein exchange significantly faster with solvent deuterons than folded structures in which hydrogen bonds and the solvent-inaccessibility protect these protons. A series of measurements under various denaturing conditions enables to study unfolding events and may even indicate which parts of the native state unfold first[204].

Scalar couplings

Scalar couplings, particularly over three bonds (^{3}J) provide access to dihedral angles due to the Karplus relation[120,121]:

$$^{3}J_{\mathrm{PQ}}(\phi) = \mathrm{A} + \mathrm{B}\, cos\phi + \mathrm{C}\, cos2\phi$$

where ϕ is the torsion angle between the two coupled nuclei P and Q, and A, B and C are nucleus-dependent Karplus coefficients.

For unfolded proteins, deviations from 'random coil' distributions of dihedral angles can reveal similarities to the native state[67,265]. This allows for example to determine side-chain rotamer preferences in denatured proteins for individual residues[250].
Electron-mediated scalar couplings across hydrogen bonds have allowed to directly determine hydrogen bond donors and acceptors[83,192,263]. This is important, since hydrogen bonds, apart from their presence in secondary structures, can form contacts between residues, which are distant in the amino acid sequence and are thus crucial for the tertiary structure of folded or partially folded protein states. Such long-range contacts have been observed in urea-denatured ubiquitin, indicating native-like structure in the first β-hairpin[159] .

Paramagnetic relaxation enhancement
Long-range distance information can be obtained from paramagnetic relaxation enhancements (PREs), caused by radical spin labels[58,59]. The unpaired electron of the paramagnetic moiety possesses a high gyromagnetic ratio leading to a distance-dependent (r^{-6}) enhancement of the relaxation rates, leading to significantly lower intensities for residues, which are in the vicinity of the spin label. The effective radius of the resonance bleaching depends both on the dynamics of the system and on the paramagnetic compound.
In most cases, spin labels (e.g. MTSL) are covalently attached to a cysteine side chain. It has to be ensured that the spin label does not perturb the investigated system by altering the structure or the structural ensemble[59]. Experiments recorded on the paramagnetic sample are compared to a diamagnetic reference (usually an identical sample for which the spin label has been reduced by the addition of e.g. ascorbic acid) and reveal regions of enhanced relaxation. Besides the distance information for folded proteins, this method allows also to detect transient long-range contacts in unfolded state ensembles[59].

Residual dipolar couplings[143]
The dipole-dipole interaction energy experienced by a magnetic nucleus S with spin $\vec{S}$ in the presence of another magnetic nucleus I with spin $\vec{I}$ is given by:

$$H_d = \frac{\gamma_S \gamma_I \hbar^2 \mu_0}{4\pi r^3}\left\{\vec{S}\cdot\vec{I} - 3\frac{(\vec{S}\cdot\vec{r})(\vec{I}\cdot\vec{r})}{r^2}\right\}$$

where γ is the gyromagnetic ratio, $\vec{r}$ is the distance vector between the two nuclei, $\hbar$ is the Planck constant divided by 2π and μ_0 is the vacuum permeability.
In the case of heteronuclear dipole-dipole interactions, the transverse components average to zero due to their rapid oscillations and can be neglected. Only the z-component (parallel to the external magnetic field $\vec{B}_0$) remains. The orientation of the internuclear vector $\vec{r}$ changes due to rotational diffusion and therefore, the heteronuclear dipolar Hamiltonian becomes a time and structural ensemble average:

$$\langle H_d^{hetero}\rangle = -h\frac{\gamma_S \gamma_I \hbar \mu_0}{4\pi^2 r^3}\left\langle \frac{3cos^2\theta - 1}{2}\right\rangle S_z I_z = hD_{IS}\cdot S_z I_z \tag{1.1}$$

where D_{IS} is the residual dipolar coupling (RDC) between spins $\vec{I}$ and $\vec{S}$, θ is the angle between the internuclear vector $\vec{r}$ (which is assumed to be of constant length r) and the external magnetic field and the brackets $\langle\rangle$ indicate the time and ensemble averaging.

The angular dependent part of equation 1.2 corresponds to the second Legendre polynomial $P_2(cos\theta)$.

$$D_{IS} = -\frac{\gamma_I \gamma_S \hbar \mu_0}{4\pi^2 r^3} \left\langle \frac{3cos^2\theta - 1}{2} \right\rangle = -\frac{\gamma_I \gamma_S \hbar \mu_0}{4\pi^2 r^3} \langle P_2 cos\theta \rangle \tag{1.2}$$

Dipolar couplings cannot be observed in the isotropic case, since the integral of $P_2(cos\theta)$ over the unit sphere is zero, whereas it does not vanish in anisotropic media and thus contributes to the splitting of NMR resonances. The dipolar and the scalar coupling contributions add up to the observed resonance splitting, since the spin part $S_z I_z$ of $\langle H_d^{hetero} \rangle$ in equation 1.1 is identical to that of the heteronuclear J-coupling:

$$H_J^{hetero} = hJ_{IS} \cdot S_z I_z$$

Experimentally, dipolar couplings are therefore determined from the differences between isotropic and anisotropic samples. While in an anisotropic environment, the observed resonance splitting corresponds to the scalar coupling plus the dipolar coupling, only the J-coupling remains in the isotropic case. However, strongly anisotropic environments e.g. in the solid state complicate the spectra due to the strong coupling effects of many spins. Therefore, only weak alignment leading to $\langle P_2(cos\theta) \rangle$ values around 10^{-3} to 10^{-4} are desired[157]. Common media for weak alignment of proteins are filamentous phages[40] or compressed polyacrylamide gels[215,247].

RDCs of weakly aligned proteins are excellent probes of long-range interactions in macromolecules[241,243], reporting on individual bond vector orientations relative to the common frame of the external magnetic field[16] according to equation 1.2. In contrast to PRE experiments, no chemical modification of the protein is required for the measurement of RDCs[16]. Nevertheless, it must be ensured that the alignment medium does not interact with the investigated protein and only provides a frame for weak alignment. Polyacrylamide gels are also well-suited for RDC investigations of denatured proteins under harsh conditions such as 8 M urea[158,224,250].

Dipolar couplings have mostly been used for structure validation or as additional source of restraints for the 3D structure determination of proteins[16]. However, applications of RDCs to determine the conformational landscape and to assess residual structure in unfolded proteins have been described as well[158,224]. Recently, RDC experiments have also been used together with relaxation dispersion methods[88,105,251,252] (see below).

Nuclear relaxation

In NMR spectroscopy, relaxation describes the return of the macroscopic magnetization to its Boltzmann equilibrium. While longitudinal (T_1) relaxation describes the process of restoring the initial, unperturbed +z magnetization, transverse (T_2) relaxation describes the loss of the x-y component, created by the initial perturbation pulse. In contrast to T_1 relaxation, which requires a population change of the nuclear spin states and thus spin transitions, only the phase coherence of the spins is affected during T_2 relaxation. T_2 relaxation is directly visible in the decreasing amplitude of the free induction decay (FID) of the time domain NMR signal, leading to a broader NMR resonance in the frequency domain[38].

Both types of relaxation require local magnetic field fluctuations[38]. For nuclear relaxation, the local fields arise from dipole-dipole interactions and chemical shift anisotropy, which is caused by the unsymmetrical electron distribution in chemical bonds. Additional mechanisms are rapid rotations of molecules or atomic groups within a molecule (which create a magnetic moment) or nuclei which possess a quadrupolar moment. The time-dependence of the magnetic field fluctuations, which lead to relaxation, is caused by molecular motion. The frequency distribution of the fluctuating magnetic fields caused by molecular tumbling is the spectral density[38].

Since the motion and hence the field fluctuations depend on both size and dynamics of a protein, relaxation experiments address both aspects. Backbone and side-chain motions on the time of the inverse of the Larmor frequency can be obtained by ^{15}N T_1, T_2 and heteronuclear NOE (nuclear Overhauser effect) experiments[59]. These data are typically analyzed with the Lipari-Szabo 'model-free' formalism[146,147].

In this approach, overall and internal (local) motions are assumed to be independent. A generalized order parameter reporting on the spatial restriction of the local motion (1 means completely restricted, while 0 indicates fully isotropic local motion) and an effective correlation time, which depends on diffusion and the nature of the motion, can be obtained from nuclear relaxation data.

This approach works well for folded proteins, but the assumption of independent overall tumbling and internal local motions is not correct for unfolded proteins. For the investigation of partially or completely denatured states, either corrections can be included, which take a distribution of correlation times into account[25,26] or reduced spectral density mapping may be applied[64,190].

Relaxation dispersion

Among the multitude of techniques to study folding pathways, NMR relaxation dispersion is becoming a more and more powerful tool, since it allows to study low-populated 'invisible' states[187]. Conformational exchange on the microsecond to millisecond time scale leads to a broadening of NMR resonances, which can be analyzed by relaxation dispersion methods and reveals several important thermodynamic and kinetic parameters of the underlying equilibria.

For quantitative analysis, the exchange contribution to the full line-width (i.e. the effective transverse relaxation rate $R_{2,\,\mathrm{eff}}$) needs to be separated from the intrinsic transverse relaxation ($R_{2,0}$) described above. This separation can be achieved by the application of rapid 180 degree CPMG (Carr-Purcell-Meiboom-Gill[28,156]) pulse trains (spin-echo elements) in the NMR experiment[150,164,172]. The effective transverse relaxation rate of a series of 2D experiments with different numbers of CPMG repeats within a fixed time frame (e.g. 80 ms) can be analyzed as a function of the CPMG pulse repetition rate (ν_{CPMG}) for each residue. Fitting of this function, which can also be expressed in terms of the time spacing between consecutive 180 degree pulses (τ_{CPMG}, shown in eq. 1.3[242], which is derived from the Carver-Richards equation[29]) yields several kinetic parameters, provided enough experimental data are available, ideally by recording spectra at multiple magnetic fields.

$$\mathrm{R}_{2,\,\mathrm{eff}} = \frac{1}{2}\Big(2\,\mathrm{R}_{2,0} + \mathrm{k}_{\mathrm{ex}} - \frac{1}{\tau_{\text{CPMG}}}\,cosh^{-1}(D_{+}cosh(\eta_{+}) - D_{-}cos(\eta_{-}))\Big) \tag{1.3}$$

using the definitions:

$$D_{\pm} = \frac{1}{2}\left(\pm 1 + \frac{\psi + 2\Delta\omega^2}{\sqrt{\psi^2 + \zeta^2}}\right)$$

$$\eta_{\pm} = \frac{\tau_{\text{CPMG}}}{\sqrt{2}}\sqrt{(\pm\psi + \sqrt{\psi^2 + \zeta^2})}$$

$$\psi = \mathrm{k}_{\mathrm{ex}}^2 - \Delta\omega^2$$

$$\zeta = 2\Delta\omega(2\cdot \mathrm{p}_{\mathrm{b}} - 1)\mathrm{k}_{\mathrm{ex}}$$

$$\tau_{\text{CPMG}} = 1/(2\nu_{\text{CPMG}})$$

In a two-state conformational exchange model, the fit parameters are the exchange rate (k_{ex}), the population of the 'invisible' state (p_{b}, which allows calculation of the population of the directly observed state), the relaxation rates in the absence of conformational exchange and, very importantly, the absolute chemical shift difference between the exchanging conformations ($\Delta\omega$). The latter parameter is extremely valuable, since it provides access to chemical shift information of an invisible state for each individual residue.

Additional experiments can yield the sign of $\Delta\omega$, which enables the exact location of each resonance of a certain nucleus[185,226]. If this is known for several nuclei (e.g. ^{1}HN, ^{15}NH, ^{13}Cα, ^{13}CO, ^{1}Hα) of each residue, this information can be subject to chemical-shift-based structure prediction (e.g. CS-Rosetta[221]) and thus reveal the atomic resolution structure of a state which may not have been observed directly[130]. Additionally, a modified CPMG experiment allows measurement of RDCs of the low-populated state, which yield bond vector orientations[88,105,251,252] and can be combined with chemical-shift-based structure prediction[130]. Such an approach has successfully been applied to solve the structure of the 'invisible' folding intermediate of the FF-domain of human HYPA/FBP11, a small (71 residues) 4-helix bundle[130]. Structure determination of a FF-domain mutant, which showed inversion of the populations confirmed the predicted structure of the folding intermediate[15]. Generally, relaxation dispersion methods can detect populations as low as 0.5%[187].

Furthermore, relaxation dispersion experiments have been shown to work for the investigation of conformational exchange in backbones of folded states[106,130,131], unfolded states[242] and also for side chains[172,173,227]. Therefore, NMR relaxation dispersion provides insight into several aspects of the folding funnel, even into 'invisible' intermediates.

The multitude of techniques and NMR experiments which allow to study various parts of the folding funnel has been used to shed light on the detailed folding mechanism of many proteins. Here, we have used this methodology to understand the unusually strong coupled binding and folding mechanism of Brinker.

Organization of the chapter

The folding of the transcription factor Brinker and the influence of various anions on the populations as well as the conformational exchange between folded and unfolded states is described in chapter 1.2. In the following section (chapter 1.3), side-chain χ_1 angular distribution are determined for urea-denatured ubiquitin and protein G and their deviations from the statistical coil model are discussed.

1.2 Folding of the Transcription Factor Brinker studied by NMR

Abstract

The *Drosophila melanogaster* transcription factor Brinker (Brk) has been reported to represent an extreme case of coupled binding and folding[41]. The N-terminal part of Brk binds to DNA as strongly as the full length protein, identifying it as DNA binding domain[213]. While completely unstructured in the absence of DNA, residues R46 to R95 adopt a well-folded helix-turn-helix motif after addition of cognate DNA. The NMR structure of a truncated variant comprising residues G43 to N101 (BrkDBD) in complex with specific DNA has been solved previously[41] showing four α-helices, where helix 3 makes the majority of specific contacts to the major groove of the DNA, while several N-terminal residues bind to the minor groove. In the absence of DNA (apo BrkDBD), BrkDBD can also fold into the same structure under certain buffer and temperature conditions. So far, the folding mechanism of BrkDBD has not been analyzed experimentally.

Here, we describe the detailed investigation of folding behavior of apo BrkDBD in the presence of various buffers by solution NMR methods and circular dichroism. The presented data indicate α-helical propensity for 3 of the 4 native helices even in unfolded BrkDBD, which may serve as initial contact points for binding partners. These comprise the recognition helix 3, helix 4 and an extended helix 2, which are all populated up to about 20%. Folded and unfolded states of BrkDBD exchange on the microsecond to millisecond time scale, enabling characterization by relaxation dispersion experiments. These experiments indicate that apo BrkDBD represents a two-state folder based on chemical shift differences, even though a limited number of residues deviate from this model. The populations of the respective states, as well as the folded-unfolded state transition (melting) temperatures show a strong dependence on anion concentrations and type. BrkDBD is unfolded at low salt concentrations and its folding propensity increases with higher charge density (shorter distance between adjacent negative charges) of anions. A correlation of the melting temperatures and the distance between adjacent negative charges in dicarboxylic anions reveals that the melting temperature increases with shorter spatial charge separations.

Thus, electrostatic interactions play a major role for folding of BrkDBD. The arising hypothesis, that BrkDBD folds also in the presence of unspecific DNA due to the negatively charged phosphate backbone could be confirmed experimentally and secondary chemical shift analysis indicated virtually identical α-helices for the folded state at all tested conditions. Based on ^{1}H-^{15}N chemical shifts, significant differences are observed for residues located in the binding interface (H80-Q83) between all folded forms of BrkDBD. These residues are more susceptible to anions, indicating a higher local conformational heterogeneity which may be a feature to probe for specific contacts.

In addition, we find that C66 is a crucial residue for folding of BrkDBD. Chemical modification by MTSL labeling as well as mutation of this residue prevents folding of the entire protein domain. MD simulations indicate that a hydrogen bond network connecting helices 1 and 2 may be less stable for the C66S and the C66N mutants than for wild type and the C66A mutant.

A possible mechanism of DNA recognition by BrkDBD is presented, which is compatible with our experimental data. Unfolded BrkDBD has a high capture radius due to its extended structure. The preformed helix 3 may form initial interactions with DNA and as the protein approaches the DNA, complete folding is induced by electrostatic interactions. The N-terminal end of the elongated helix 2 in the unfolded state needs to unwind in order to prevent steric clashes with the DNA double strand. In its bound and folded state, BrkDBD can perform all modes of translocation along a DNA strand (sliding and hopping[108,110]) and if partially unfolded or unbound, it can translocate *via* 3D diffusion and the intermolecular 'monkey-bar' mechanism[256]. In this mechanism, the link between binding and folding is utilized to accelerate the target recognition process, which may be representative for other single-domain transcription factors.

1.2.1 Introduction

Protein folding is still not completely understood, even after attempting for many decades to deduce patterns which allow to predict protein structures from the corresponding amino acid sequences. Essentially, folding of a protein describes the diffusion of a polypeptide chain through the conformational space, also referred to as an 'energy landscape', until it finally reaches the (global) minimum which corresponds to the native, folded state[53]. The famous Levinthal paradox states that a polypeptide chain cannot sample the full conformational space, since probing each conformation would require an astronomical amount of time[142]. Thus, mechanisms must exist which significantly reduce the size of the required search space to achieve folding in a reasonable time. The detailed folding mechanism may vary from one protein to another, since fast two-state folders as well as folding routes with one or more, often low-populated intermediates have been reported[130].

Advances in experimental techniques have contributed significantly to address some of the most important questions and computational structure prediction routines succeed increasingly well, particularly if similar structures are already known[49]. Therefore, experimental structure determination as well as elucidation of folding mechanisms are important.

Generally, the structure of a polypeptide depends strongly on the surrounding conditions. Already in the 19th century, the famous Hofmeister series of anions[96] described the salting-in and salting-out of proteins, which later has been disccused in the context of protein folding[188,189]. Protein denaturation can be induced by several parameters such as addition of denaturing agents (e.g. urea or guanidinium hydrochloride, heating, cooling (cold-denaturation) or pressure. Many folded proteins can change conformations in the presence of an effector molecule (e.g. c-di-GMP-binding PilZ domains[86]) or even change their complete structure into another fold (e.g. misfolding, leading to formation of amyloid fibrils[53]).

The interest in gaining a complete understanding of protein folding is also driven by the existence of misfolding-related diseases such as Alzheimer's or Parkinson's. Misfolding shows the dilemma of the attempt to predict structures from the amino acid sequence, since it is a consequence of the ability of an amino acid sequence to fold into different three dimensional structures, depending on local conditions. A misfolded structure likely represents a local minimum in the conformational energy landscape at native conditions, but may become a global minimum when conditions are altered for example due to disease[39]. Thus, in order to achieve accurate structure prediction, the influence of parameters which lead to changes in the adopted fold needs to be determined experimentally and included in folding models.

Transient DNA-protein interactions

Recently, low-populated transient intermediates of a homeodomain (HOXD9), interacting with DNA have been inferred from intermolecular paramagnetic relaxation enhancement (PRE) experiments[108,110]. In such experiments, a single spin-label (a stable radical moiety) is attached to a specific site of one of the involved macromolecules and the enhancement of the relaxation rate (due to the large magnetic moment of the unpaired electron) is quantified. Due to the r^{-6} distance dependence of the PREs, spatial information can be obtained for both specific and unspecific complexes of macromolecules. For a system where low-populated transient intermediates are in fast exchange with the dominant state, both structural and mechanistic features can be studied.

In the case of HOXD9 interactions with specific and unspecific DNA, sliding and hopping on a single DNA molecule as well as intermolecular translocation have been reported as mechanisms, which significantly enhance the speed of target recognition[108–110]. These mechanisms have been investigated for three homeodomains by molecular dynamics (MD) simulations, where solely intermolecular electrostatic interactions were taken into account[256]. Both intramolecular sliding and hopping were observed (1D translocation) as well as 3D translocation and a 'monkey-bar' mechanism named after its resemblance to children swinging along monkey bars[256]. In this mechanism, the recognition helix is bound to one DNA molecule and the charged N-terminus of the DNA binding domain interacts with another DNA molecule. The different mechanisms were shown to be salt-dependent in the MD simulations. Changes of ionic strength can therefore alter the contributions of individual mechanisms to the speed of DNA recognition. Additionally, the charge pattern of the unfolded N- or C-terminal tails has been investigated computationally and has been interpreted in the context of charge conservation in transcription factors, revealing that the experimentally observed sequences are evolutionally optimized for efficient DNA search[257].

Such mechanisms in general may be particularly attractive for intrinsically disordered proteins (IDPs), which require a specific binding partner for folding and which are often found to exhibit protein-DNA or protein-protein interactions[93,248]. For IDPs, the combination of these mechanisms with their high capture radius (due to an elongated chain conformation[223]), would not only accelerate the search for cognate binding partners, but also influence their folding behavior due to the coupling between binding and folding[128,271,272].

In a comparison of predicted naturally disordered regions in proteins, it has been found that more than 82% of transcription factors posses extended disordered regions[148]. In contrast, such regions are found in less than 55% and 19% of proteins of two control datasets, devoid of transcription factors[148].

The transcription factor Brinker

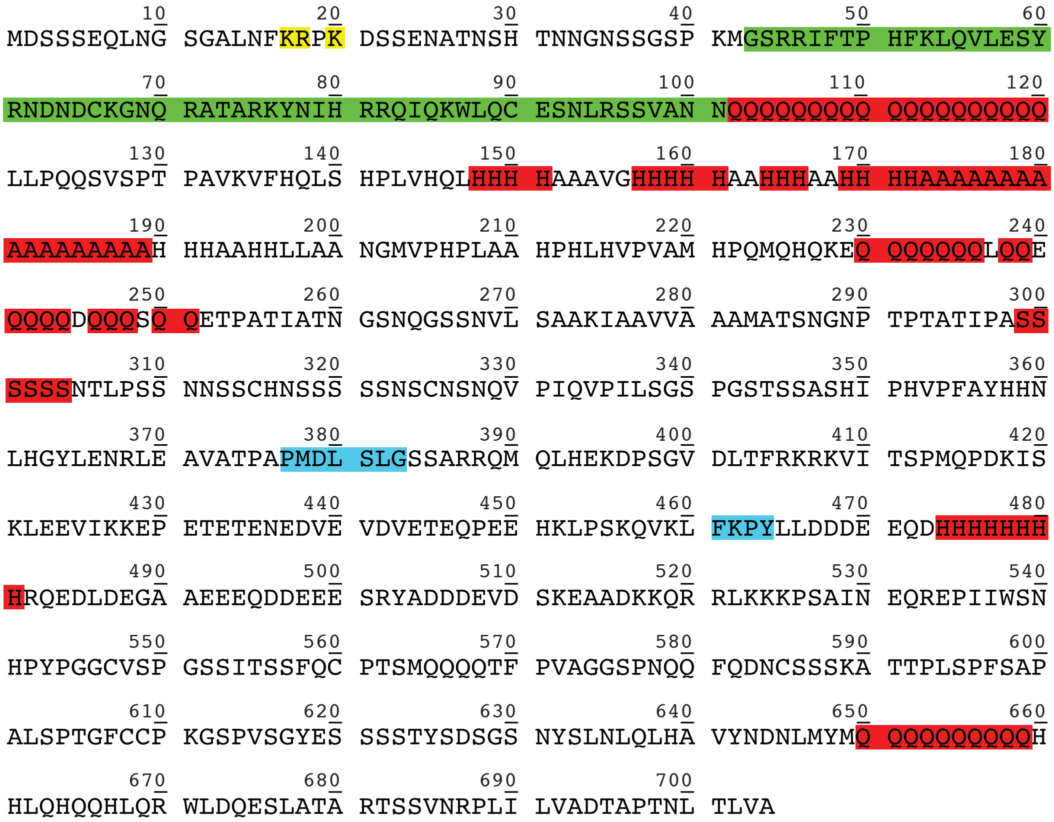
The Brinker protein (Brk) is a transcription factor in *Drosophila melanogaster*, which represses the target genes of Decapentaplegic Dpp[114]. Dpp is a member of the transforming growth factor-β superfamily[2] responsible for the development of limbs, wings and organs and it is the *Drosophila* homolog of vertebrate Bmp2 and Bmp4. The repression is achieved by using a complex molecular mechanism including several proteins[155,174,200]. Target genes of Brk, such as *optomoter blind* (*omb*) are repressed in a concentration-dependent manner[170,174]. Therefore, Brk is an important component in the Dpp morphogen signaling pathway.

Brk has a length of 704 amino acids, comprising several domains, mostly of unknown function as well as several stretches of repeated amino acids comprising three poly-Q, two poly-H, one poly-A and one poly-S, with lengths of 4 up to 19 residues (Figure 1.2). Disordered, repetitive amino acid stretches are often found in the sequences of transcription factors[65,100]. In addition to the N-terminal DNA binding domain (residues 1-101), two corepressor binding site motifs have been reported[27,113] and binding has been confirmed experimentally[90,278].

The N-terminal domain of Brinker (residues 1-101, Brk1) has been described as disordered in the absence of DNA, but residues 43-101 (BrkDBD) form a well defined helix-turn-helix motif comprising 4 α-helices in the presence of its cognate DNA as revealed by solution NMR spectroscopy (Figure 1.3[41]). Structurally, this fold belongs to the DNA-binding homeodomain-like family, characterized by the helix-turn-helix motif. However the arrangement of the helices differs considerably from their relative orientations in homeodomains[41]. The main differences are shorter lengths of helices 2 and 3 in BrkDBD. In BrkDBD, mainly the N-terminal part of helix 3 is responsible for the GGCGTC motif recognition while the entire longer helix 3 in homeodomains is involved in the recognition of AT-rich DNA. Consistent with the folded BrkDBD in complex with cognate DNA, a construct comprising residues S44-A99 binds equally strong to DNA as full-length Brinker[213].

Initial studies of BrkDBD heat denaturation

Whereas BrkDBD is unfolded in the absence of DNA above 20 °C temperature, it folds below 15 °C, based on a NMR ^{1}H-^{15}N HSQC data[41]. To assess the structural differences of the folded state in the absence and the presence of its specific DNA, the short apo BrkDBD construct has also been investigated by NMR spectroscopy[41]. It is nearly identical to the DNA-bound form based on ^{13}Cα and ^{13}CO



```
        10         20         30         40         50         60
MDSSSEQLNG SGALNFKRPK DSSENATNSH TNNGNSSGSP KMGSRRIFTP HFKLQVLESY

        70         80         90        100        110        120
RNDNDCKGNQ RATARKYNIH RRQIQKWLQC ESNLRSSVAN NQQQQQQQQQ QQQQQQQQQQ

       130        140        150        160        170        180
LLPQQSVSPT PAVKVFHQLS HPLVHQLHHH HAAAVGHHHH HAAHHHAAHH HHAAAAAAAA

       190        200        210        220        230        240
AAAAAAAAAH HHAAHHLLAA NGMVPHPLAA HPHLHVPVAM HPQMQHQKEQ QQQQQQLQQE

       250        260        270        280        290        300
QQQQDQQQSQ QETPATIATN GSNQGSSNVL SAAKIAAVVA AAMATSNGNP TPTATIPASS

       310        320        330        340        350        360
SSSSNTLPSS NNSSCHNSSS SSNSCNSNQV PIQVPILSGS PGSTSSASHI PHVPFAYHHN

       370        380        390        400        410        420
LHGYLENRLE AVATPAPMDL SLGSSARRQM QLHEKDPSGV DLTFRKRKVI TSPMQPDKIS

       430        440        450        460        470        480
KLEEVIKKEP ETETENEDVE VDVETEQPEE HKLPSKQVKL FKPYLLDDDE EQDHHHHHHH

       490        500        510        520        530        540
HRQEDLDEGA AEEEQDDEEE SRYADDDEVD SKEAADKKQR RLKKKPSAIN EQREPIIWSN

       550        560        570        580        590        600
HPYPGGCVSP GSSITSSFQC PTSMQQQQTF PVAGGSPNQQ FQDNCSSSKA TTPLSPFSAP

       610        620        630        640        650        660
ALSPTGFCCP KGSPVSGYES SSSTYSDSGS NYSLNLQLHA VYNDNLMYMQ QQQQQQQQQH

       670        680        690        700
HLQHQQHLQR WLDQESLATA RTSSVNRPLI LVADTAPTNL TLVA
```

Figure 1.2 Amino acid sequence of full-length Brinker (UniProtKB: Q9XTN4). The short 59-residue construct of the N-terminal DNA binding domain (BrkDBD) is highlighted in green. Positively charged residues, commonly found in the middle of N-terminal tails[257] of transcription factors, are shown in yellow. Stretches of repetitive amino acids which are predicted to be unstructured are shown in red. Corepressor binding motifs are highlighted in blue.

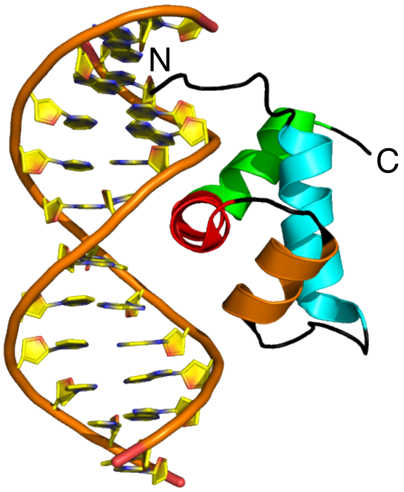


Figure 1.3 NMR structure of BrkDBD from the protein data bank (pdb code 2GLO[41]) of BrkDBD in complex with GC-rich omb12T5 DNA. Helices are colored: helix 1 (cyan), helix 2 (orange), helix 3 (red) and helix 4 (green). Capital letters denote N- and C-terminus, respectively. The recognition helix 3 binds directly to the GGCGTC motif in the major groove. Additional contacts to the phosphate backbone were reported for Q70 and R71 at the N-terminus of helix 2. The N-terminus of BrkDBD interacts with the minor groove.

secondary chemical shifts. The slight differences between resonances of DNA-free and bound form could be attributed to small local differences in the DNA binding interface (residues at the N-terminal end of helix 3, which bind to the major groove and residues at the N-terminal part of BrkDBD, which interact with the minor groove). Below 20 °C, the ratio of resonance intensities between folded and unfolded state changed in a temperature-dependent manner, reflecting variations in the populations of these two states.

A fast technique to study the thermal folding-unfolding transition is circular dichroism (CD). In CD, typically the residue mean ellipticity at 222 nm, indicative of α-helical secondary structure propensity, as a function of the temperature can be monitored to study denaturation of helices[79,231,244,270]. Ideally, these data yield a sigmoidal curve and the midpoint of the transition is defined as the melting temperature.

Comparison of CD data on BrkDBD dissolved in either 20 mM potassium phosphate or 20 mM Tris buffer indicated an increase of the melting temperature of about 5 °C in phosphate buffer[41]. A pH change from 5.6 to 7.0 had a similar effect based on CD data (C. Bodenreider and T. Kiefhaber, unpublished). The ability to populate the folded and the unfolded state depending on the conditions makes this short construct of BrkDBD a promising system to study protein folding both in the absence and in the presence of DNA.

Here, we describe in detail effects of salt on the folding energy landscape of the short 59 residue Brinker BrkDBD in the absence of DNA by comparing heat denaturation at various salt conditions by CD and NMR. NMR data of the folded states of apo BrkDBD are compared for a series of buffer conditions including the presence of specific and unspecific DNA. The propensity to form α-helical secondary structure is deduced from ^{13}Cα secondary chemical shifts under various conditions. ^{15}N relaxation dispersion experiments are applied to study the low-populated state which is in exchange with unfolded BrkDBD.

Several attempts have been made to attach the radical spin label MTSL to the cysteine side chains of wild type BrkDBD and the C90S and C66S mutants in order to detect transient DNA-protein encounter complexes by PREs[108]. However, BrkDBD did neither fold upon MTSL modification at C66 nor upon mutation of C66. MD simulations were performed to address the reason for the folding inability of MTSL-modified C66 and C66 mutants of BrkDBD.

A potential mechanism of DNA recognition by BrkDBD is proposed, based on experimental results and theoretical considerations. This mechanism may be representative for DNA recognition by intrinsically disordered transctiption factors.

1.2.2 Results and Discussion

Heat denaturation of BrkDBD studied by circular dichroism

Initial NMR experiments at various buffer conditions indicated that the populations of folded and unfolded BrkDBD depend on the type and concentration of buffer anions. The folded state of BrkDBD was significantly higher populated in the presence of buffer anions with multiple negative charges. Since BrkDBD is folded in the complex with specific DNA, the initial working hypothesis was that buffers, which mimic the phosphate backbone by a certain spatial separation of the negative charges have a stronger stabilizing effect on the folded conformation of BrkDBD and thus show higher temperatures of the transition from the folded to the unfolded state (melting temperature). We investigated the temperature-unfolding of BrkDBD at low concentrations ($\leq$ 35 μM) at various buffer conditions by CD. The investigated buffers included non-carboxylic (phosphate), monocarboxylic (acetate) and dicarboxylic (oxalate, malonate, succinate, glutarate, adipate) buffers with increasing chain-lengths (Table 1.1, Figures 1.4 and 1.5). For all samples, the pH was adjusted to 5.5 by addition of small amounts of NaOH to ensure identical pH contributions to the melting temperature. Due to very strong absorbance of the carbonyl groups of buffer anions between 190 and 215 nm, the residue molar ellipticity has been monitored at 222 nm to indicate the presence of α-helices.

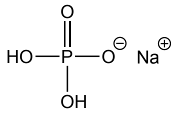
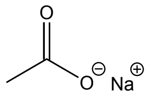
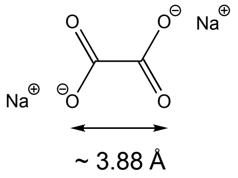
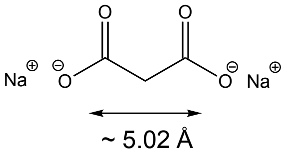
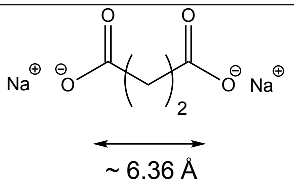
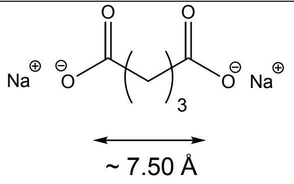
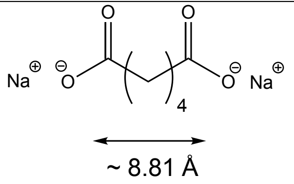
buffer 20 mM	c(BrkDBD) (μM)	structure	pKa[92]	I_c (mM)	T_{melt} (°C)
phosphate	14		2.16 7.21 12.32	21.3	20
acetate	21		4.76	17.7	-
oxalate	20		1.25 3.81	59.9	39
malonate	11		2.85 5.70	35.8	29
succinate	35		4.21 5.64	37.0	22
glutarate	13		4.32 5.42	41.1	18
adipate	13		4.41 5.41	41.1	16

Table 1.1 Melting temperatures (T_{melt}) of BrkDBD and calculated ionic strengths (I_c) for various anionic buffers. The pH was adjusted to 5.5 for all samples by addition of small amounts of NaOH. Ionic strengths have been calculated taking into account monovalent counterions for all buffer anions and the charged BrkDBD side chains. Melting temperatures have been estimated from the CD melting curves (Figure 1.4).

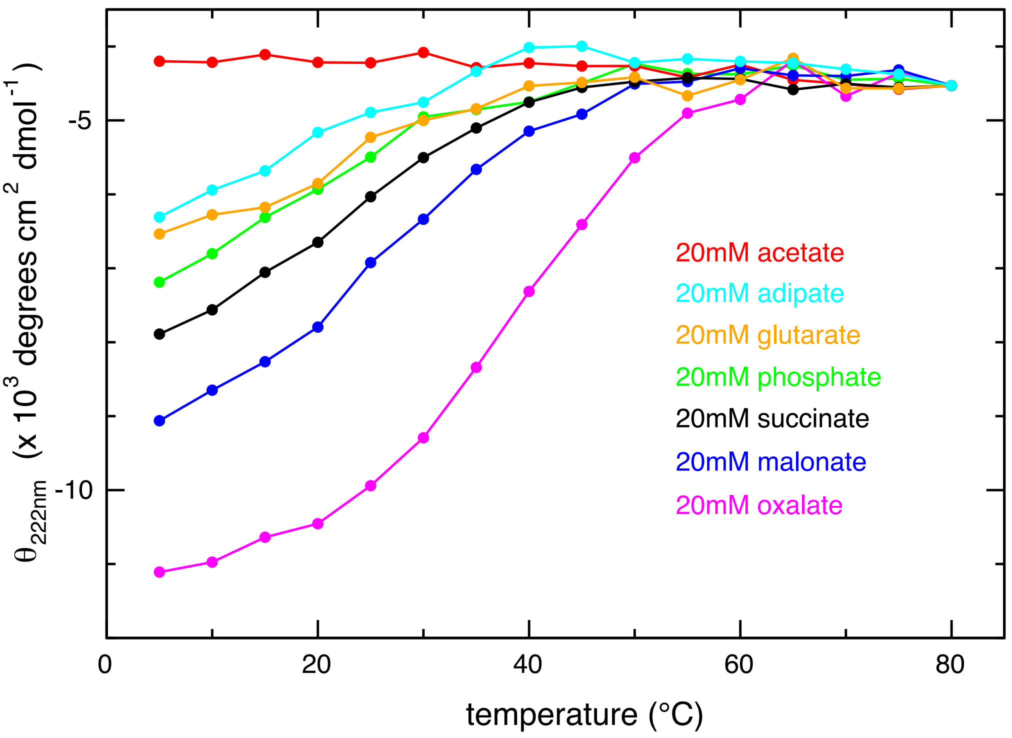


Figure 1.4 CD melting curves measured at 222 nm for various anionic buffers. The mean residue ellipticity as a function of temperature indicates unfolding of BrkDBD at higher temperatures. In order to facilitate comparison, the curves have been slightly shifted vertically, such that they match the value measured in 20 mM oxalate at 80 °C. Obviously, shorter chain-lengths of dicarboxylic buffer anions correlate with higher melting temperatures (Table 1.1). All buffers contained sodium as counterion.

Apo BrkDBD dissolved in 20 mM sodium phosphate buffer showed a melting temperature of about 20 °C (Table 1.1, Figure 1.4 (green)) which is very similar to the reported NMR-based result[41]. No plateau for the mean residue ellipticity has been observed at low temperatures, indicating that convergence to the folded state is reached only below 5 °C. Similar to the CD melting curve in 20 mM phosphate, BrkDBD showed a transition from folded to unfolded in all tested dicarboxylic buffers. In contrast, a flat temperature-independent mean residue molar ellipticity at 222 nm is found in 20 mM acetate (Figure 1.4 (red)), which indicates the absence of folded BrkDBD over the entire temperature range.

The melting temperature correlated with the distance between the negative charges. While the shortest dicarboxylic buffer (oxalate) showed the highest melting temperature, it decreased with increasing chain-length (Table 1.1). This can also be concluded from the mean residue ellipticity at 5 °C as a function of charge separation (Figure 1.5), which is minimal for oxalate-buffered BrkDBD. The distance between the negative charges has been calculated for each compound, assuming standard C-C single bond lengths and a mean oxygen-carbon bond lengths (between double

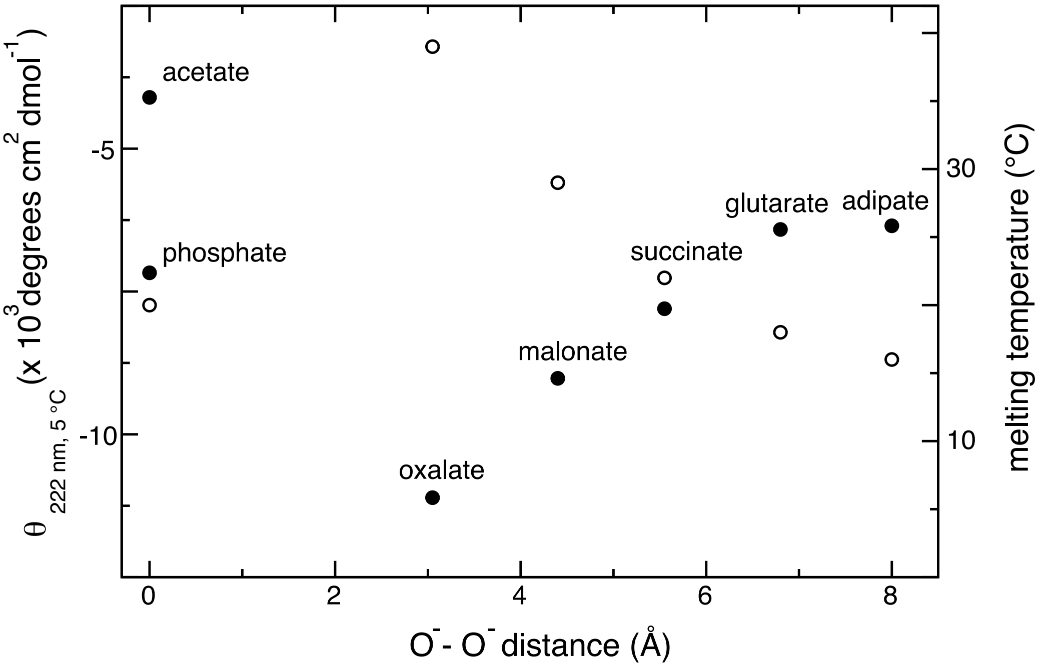


Figure 1.5 Mean residue ellipticity at 5 °C (filled circles) and melting temperature (open circles) as a function of the distance between intramolecular anionic charges. The shortest distance correlates with the highest melting temperature and the most pronounced propensity for α-helices at 5 °C. This indicates that the folding energy landscape of BrkDBD depends on anionic charge density. Note that for acetate, no folded conformation and subsequently no melting temperature was observed.

and single bond) due to the delocalization of the negative charge on each carboxylic group (Table 1.1). In contrast to the mean charge distance of 6.5-7.5 Å between two adjacent phosphate oxygens in BrkDBD-specific DNA (omb12T5[41]), as quantified from the NMR structure, the melting temperature is not maximal for similar distances between intramolecular anionic charges. The observed trend of higher melting temperatures at shorter distances between adjacent negative charges indicates that charge density is more important for folding of BrkDBD than a certain separation of charges. Due to the different pKa values of the investigated buffers, further pH-dependent experiments are required to completely rule out the initial hypothesis, that buffers, which mimic the backbone of cognate DNA have the most stabilizing effect on folded BrkDBD. It is noteworthy that in a comparison of the melting curves, the melting point in phosphate buffer is between those in succinate and glutarate, which have an estimated charge-separation of about 6.4 Å and 7.5 Å, respectively (Figure 1.5, Table 1.1). This may indicate, that these anions bind to similar sites, but are unable to compensate for specific contacts from the nucleobases to the protein. Furthermore, no simple correlation between ionic strength and melting temperature has been found in these experiments (Table 1.1), again indicating that rather the charge density is the decisive factor for folding of BrkDBD in the absence of specific interactions with DNA. Remarkably, the heat denaturation of BrkDBD seems fully reversible due to the superposition of full wavelength CD spectra at 5°C, recorded before and after exposure to high temperatures ($T_{max} = 80$ °C), for all tested conditions.

Heat denaturation of BrkDBD monitored by NMR

Phosphate buffer

The ^{1}H-^{15}N HSQC NMR spectrum of 0.4 mM BrkDBD dissolved in phosphate buffer (20 mM NaH_2PO_4/Na_2HPO_4, 10 mM NaCl, 3 mM DTT, pH 5.6) revealed the presence of folded and unfolded BrkDBD at 5 °C (Figure 1.6). This observation is consistent with the CD melting curve of BrkDBD dissolved in 20 mM phosphate buffer, since no plateau had been reached at low temperature, indicating incomplete folding under these conditions. At 35 °C, the overall HSQC resembled that of an unfolded protein due to the narrow dispersion of ^{1}H resonances, even though the corresponding CD spectrum indicated incomplete unfolding at this temperature. Many of the resonances showed extreme broadening under these conditions. Such line broadening is typically attributed to conformational exchange in the microsecond to millisecond time scale, i.e. the chemical shift time scale. However, in this case, the extreme extent of line broadening and the absence of several resonances indicates, that most residues exchange on an intermediate time scale. Only at very high temperatures (50 °C), the resonances became narrower again, but still suffered from broadening compared to spectra completely devoid of conformational exchange.

Protein precipitation and aggregation had been observed after exposure to high temperatures for more than a total of four days. This was apparent since the NMR resonance intensities decreased and solid aggregates became visible. After recooling to

5°C, only very low folded state resonance intensities were observed. The unfolded state was much higher populated than initially at low temperatures and the resonances showed even more pronounced broadening. This result is in contrast to the reversibility concluded from CD experiments. Presumably, the much higher concentration of the protein and the longer exposure to high temperatures compared to the CD samples caused these effects.

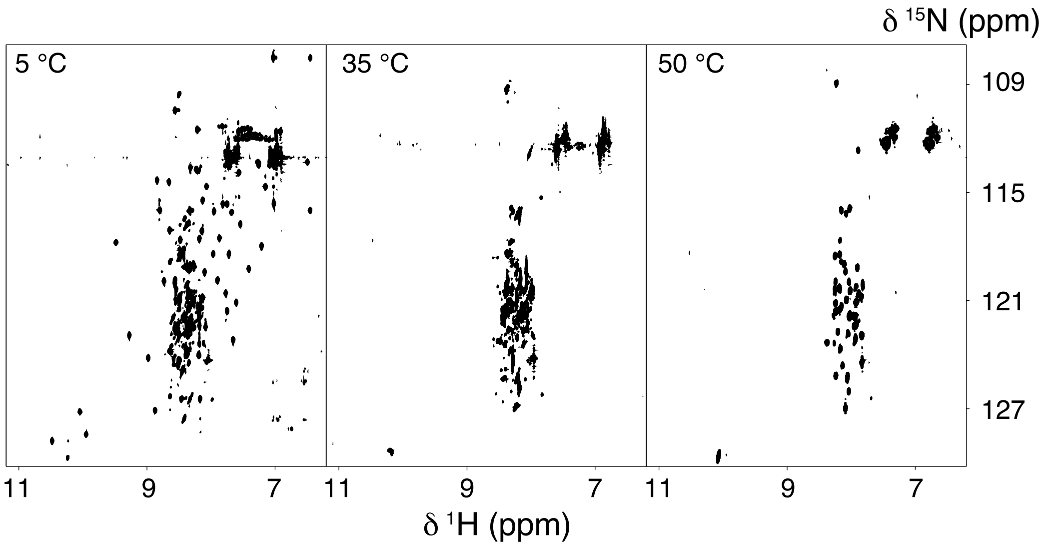


Figure 1.6 ^{1}H-^{15}N HSQC spectra of BrkDBD dissolved in buffer containing 20 mM phosphate at various temperatures. At 5 °C (left) resonances of folded and unfolded BrkDBD are observed, while at 35 °C (middle) and 50 °C (right) the dispersion of ^{1}H chemical shifts indicates the presence of only unfolded protein. Several resonances exhibit severe line broadening at 35 °C, indicating conformational exchange, whereas the resonances become narrower again at 50 °C due to faster dynamical averaging.

Succinate buffer

The ^{1}H-^{15}N HSQC NMR spectrum of BrkDBD dissolved in 20 mM succinate buffer (20 mM succinate, 10 mM TCEP, pH 5.6) is very similar to that in 20 mM phosphate buffer at 5 °C. The dominant resonances corresponded again to folded and unfolded BrkDBD, consistent with the CD experiments. Additionally, two sets of significantly weaker resonances have been observed. These two resonances sets had the ^{15}N chemical shift of folded, but the ^{1}H chemical shift of unfolded BrkDBD and *vice versa* (Figure 1.7). This indicates that the conformational exchange between folded and unfolded BrkDBD is on the slow time scale at these conditions, but fast enough to cause the build-up of exchange peaks during the ^{1}H-^{15}N HSQC experiment (Figure 1.7). Therefore, the exchange between folded and unfolded must be faster than a few milliseconds.

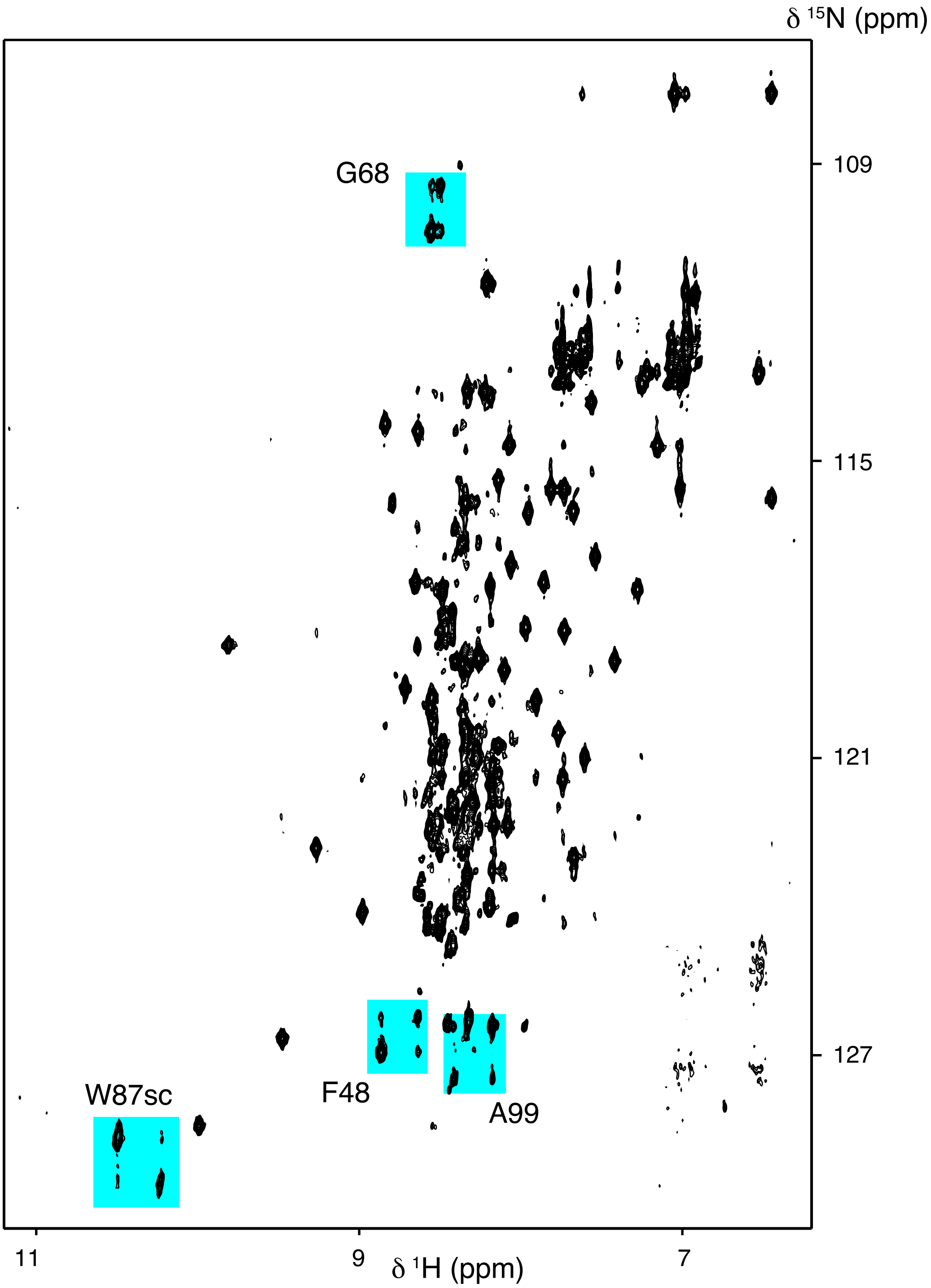


Figure 1.7 ^{1}H-^{15}N HSQC spectrum of BrkDBD in 20 mM succinate buffer, acquired at 5 °C. Selected resonances with their corresponding exchange resonances are highlighted. Such exchange resonances are found for nearly all residues.

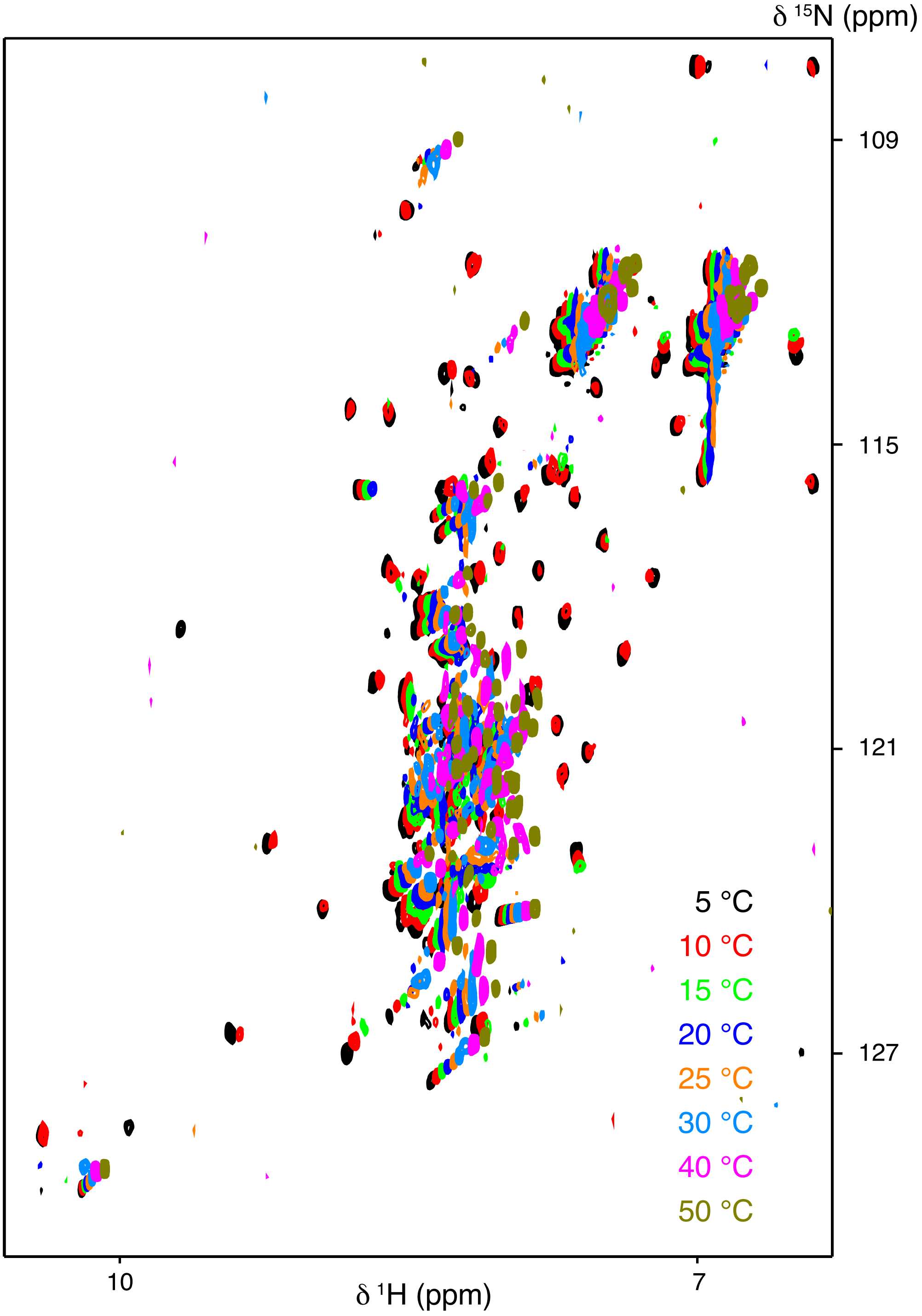


Figure 1.8 Temperature series of ^{1}H-^{15}N HSQC spectra of BrkDBD dissolved in buffer containing 20 mM succinate. Color coding for temperatures is indicated in the Figure.

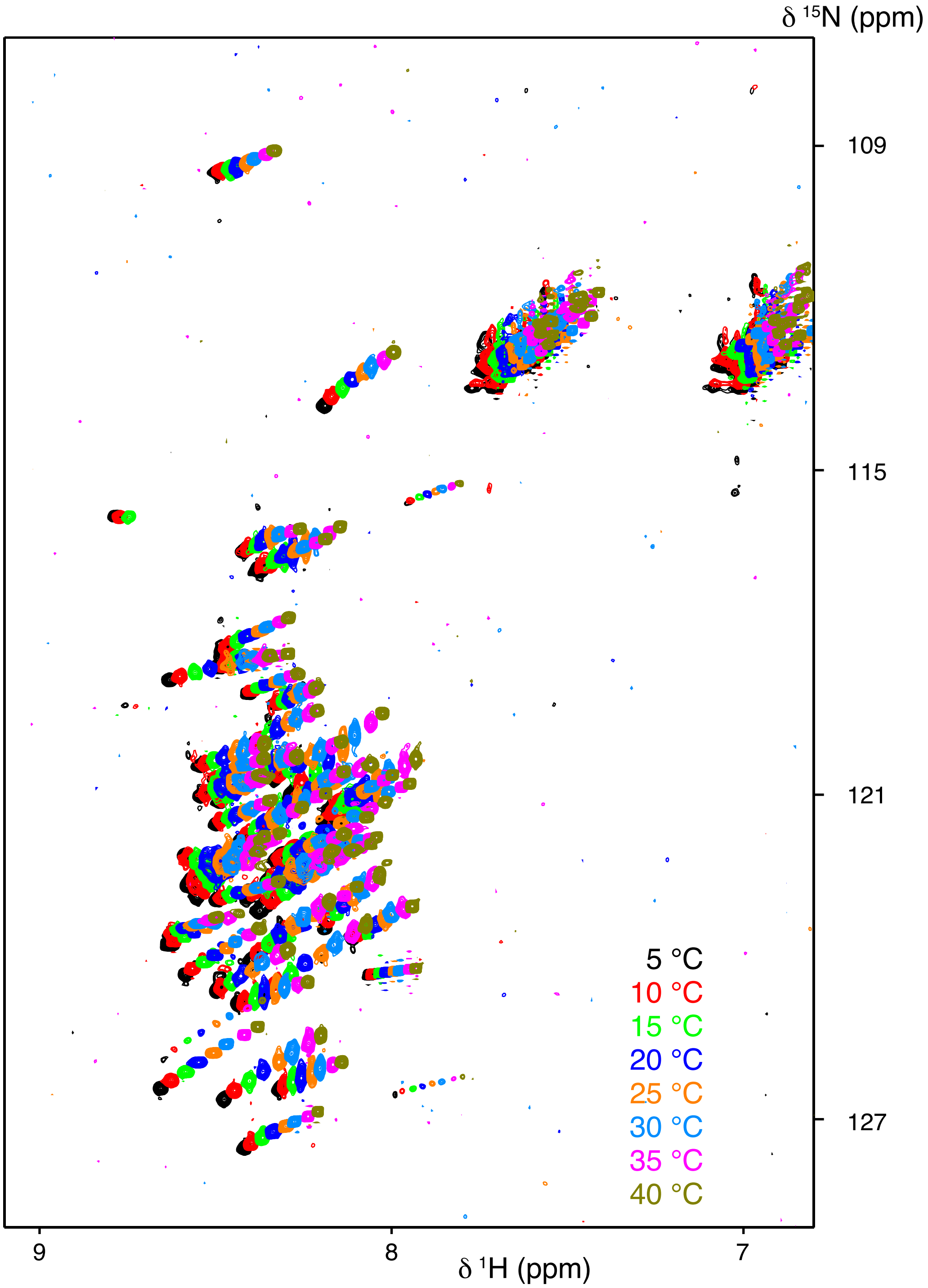


Figure 1.9 Temperature series of ^{1}H-^{15}N HSQC spectra of BrkDBD dissolved in buffer containing 20 mM acetate. Color coding for temperatures is indicated in the Figure.

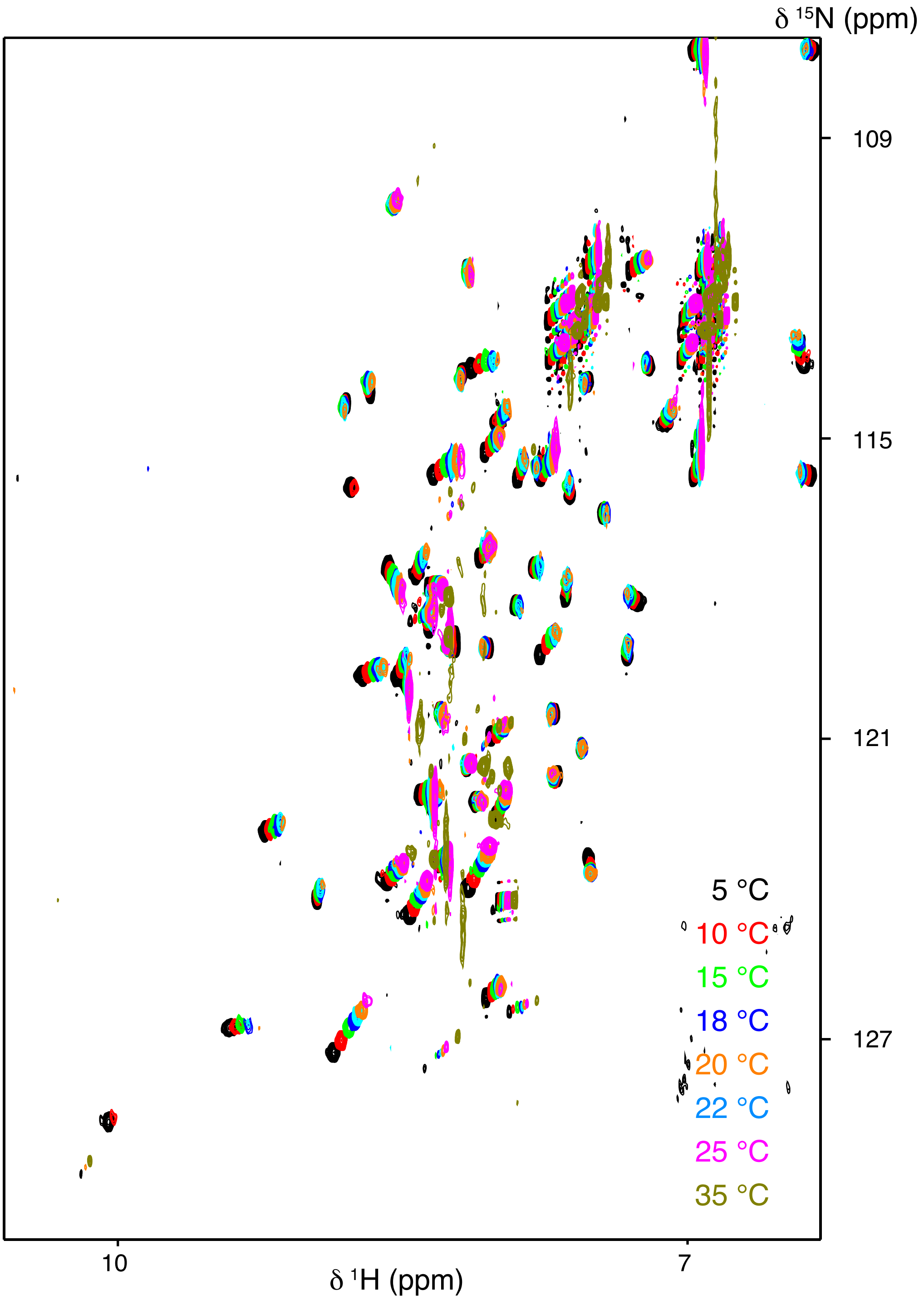


Figure 1.10 Temperature series of ^{1}H-^{15}N HSQC spectra of BrkDBD dissolved in buffer containing 20 mM oxalate. Color coding for temperatures is indicated in the Figure.

In order to gain further insight into the conformational exchange, a temperature series of ^{1}H-^{15}N HSQC spectra has been recorded (Figure 1.8). With increasing temperature, several resonances of the folded state disappeared or experienced broadening, indicating both decreasing stability of folded BrkDBD and a change of the exchange regime. The dispersion of resonances indicated unfolded BrkDBD at 15 °C, even though few resonances of the folded state remained visible. At 20 °C, no resonances attributed to the folded state were detected. The non-uniform broadening of the remaining unfolded state resonances was not surprising, since the exchange time scale depends on the frequency difference between the individual resonances of the interconverting states. Some N- and C-terminal residues of BrkDBD give rise to sharp resonances, even at this intermediate temperature. With increasing temperature, the extent of line broadening was reduced again and many sharp resonances were observed at 40 °C. However, at that temperature, some residues were still severely broadened. Only few residues showed broadened or diminished resonances at 50 °C, similar to the observations in 20 mM phosphate buffer. The sample stability also decreased after exposure to high temperatures.

Acetate buffer

In contrast to the findings at low temperatures in 20 mM phosphate and 20 mM succinate and consistent with the CD data, only unfolded BrkDBD was observed in ^{1}H-^{15}N HSQC spectra in 20 mM acetate buffer (20 mM acetate, 5 mM DTT, pH 5.5) at all investigated temperatures (Figure 1.9). Remarkably, the resonances of unfolded BrkDBD had a much narrower linewidth over the full temperature range compared to those in the previously described buffers. However, for individual resonances (e.g. S59, A72, A74, A99) line broadening was observed at several intermediate temperatures indicating the presence of conformational exchange, even though to a much lesser extent. Interestingly, no protein precipitation was observed in 20 mM acetate buffer even after exposure to high temperatures.

Oxalate buffer

In order to further investigate the heat denaturation of BrkDBD at conditions, which lead to the highest melting temperature (Table 1.1) according to CD data, a temperature series of ^{1}H-^{15}N HSQC spectra has been recorded in 20 mM oxalate buffer (20 mM oxalate, 5 mM DTT, pH 5.5). Since CD experiments showed that oxalate-buffered BrkDBD had the most pronounced helical propensity of the tested buffers, a heat denaturation series was also measured by NMR (Figure 1.10). Below 22 °C, the resonance dispersion in the ^{1}H-^{15}N HSQC spectrum indicated the presence of mainly folded BrkDBD. This is in contrast to the other NMR results, where no such resonances had been observed at higher temperatures than 15 °C. Upon increasing temperature, several resonances vanished or experienced broadening. A similar extent of resonance broadening as at 25 °C for BrkDBD dissolved in 20 mM phosphate or succinate buffer were observed only at 35 °C, which reflects the increased folding propensity in 20 mM oxalate buffer. Since the resonance broadening was not uniform during the heating procedure, a detailed analysis of the sequential broad-

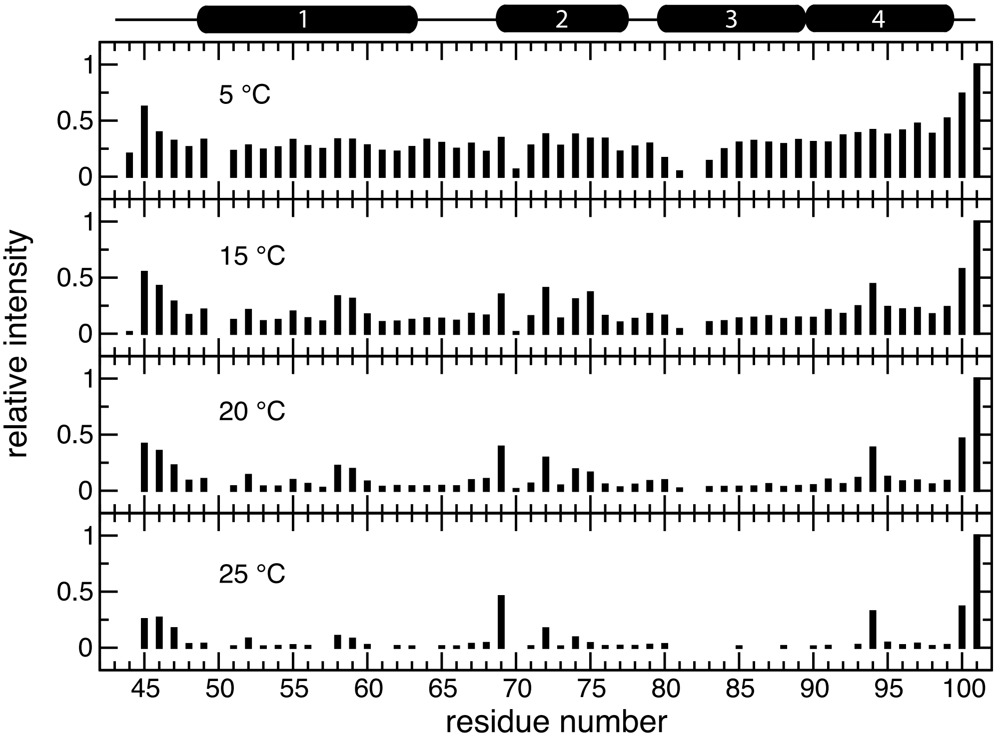


Figure 1.11 Temperature series of the relative resonance intensities (normalized on residue N101) of BrkDBD in 20 mM oxalate buffer as a function of the residue number.

ening events could shed light on the unfolding pathway. If folding and unfolding of BrkDBD proceed along the same reaction path, this may elucidate the underlying mechanism.

To this aim, the intensities of individual folded state resonances have been quantified at various temperatures and were compared to the corresponding ^{1}H-^{15}N HSQC spectra (Figure 1.11). Already at 5 °C, R82 was invisible and residues Q70 and R81 were broadened, corroborated by their lower intensities. Broadening of these two residues became even more severe at 10 °C and Q70 was almost undetectable at 15 °C. Residues T49 (N-terminal end of helix 1), E58, S59, A72, A74 and H80 were broadened at 18 °C. Even at temperatures at which the resonances of most residues had disappeared, few residues remained relatively intense (35 °C, Figure 1.10). These comprised not only residues close to the termini (R45, R46, I47, N100 and N101) but also residues F52, N69 and L94. For the latter two, however, overlap with unfolded state resonances could be the reason for the strong resonance intensities.

The resonance intensities did not correlate strongly with the apparent extreme broadening of individual resonances (Figure 1.11). For example, the resonances of E58, A72 and A74 at 20 °C were more intense than expected from the obvious resonance broadening in the ^{15}N dimension (Figure 1.10). Remarkably, a closer inspection of the chemical shifts of these particular residues revealed that the ^{1}H chemical shifts of the folded and the unfolded state are virtually identical, whereas the ^{15}N chemical shifts differ by about 2.5 ppm each. These findings illustrate the difficulty of such an analysis due to the increased resonance overlap caused by the line broadening (conformational exchange) in both ^{1}H and ^{15}N dimensions. The exchange regime itself depends on the effective kinetics of the exchange process and on the respective frequency differences, which may not be identical for both nuclei. Even though the apparent broadening of resonances is non-uniform, no simple correlation to unfolding could be established at the atomic level.

Nevertheless, from the fact that most residues of the folded state vanished simultaneously at about 25 °C, while unfolded state resonances gained intensity, this can be interpreted as a two-state folding mechanism. The outliers which are not located at the termini of the protein (N69 and L94) seem to be less affected by temperature. However, C66 and F52 of the unfolded state have identical chemical shifts in 20 mM acetate as folded N69 and L94 at this temperature. Therefore, an overall two-state folding mechanism seems plausible.

In contrast to the reduced stabilitiy of BrkDBD in 20 mM phoshate or succinate buffer after exposure to higher temperatures, only little precipitation has been observed in 20 mM oxalate presumably due to the low concentration (100 μM) and the shorter time of exposure to high temperatures (1.5 hours at 35 °C). At about 300 μM concentration, which was used for resonance assignment experiments, precipitation occured even at low temperatures (5 °C), again indicating that BrkDBD is not stable in salt-folded states.

Salt-dependence of the population of folded BrkDBD

The CD and NMR data of folded and unfolded BrkDBD at 20 mM buffer concentration revealed that the unfolding temperature as well as the population of folded and unfolded states depend on the type of anion in solution. Therefore, we expected that an increase in total anion concentration may be sufficient to force the protein to adopt its folded state. Indeed, resonances attributed to both the folded and the unfolded state were detected at 5 °C in 200 mM acetate buffer, proving that predominantly electrostatics alters the energy landscape of BrkDBD.

This observation explains the difference between NMR spectra of BrkDBD solutions where TCEP instead of DTT was used as a reducing agent. Resonances attributed to the folded state of BrkDBD were observed in 20 mM acetate buffer, in the presence of 6 mM TCEP, but not in samples using solely DTT as reducing agent (Table 1.2). This can be rationalized by the structure of TCEP, which contains three carboxylic groups (Figure 1.12), which are mostly deprotonated at pH 5.5. The original intention of TCEP addition was to increase the protection of the cysteine residues from oxydation due to the higher efficiency of TCEP compared to DTT. It is therefore important to take all species in solution into account (including the reducing agent), which could influence the investigated process.

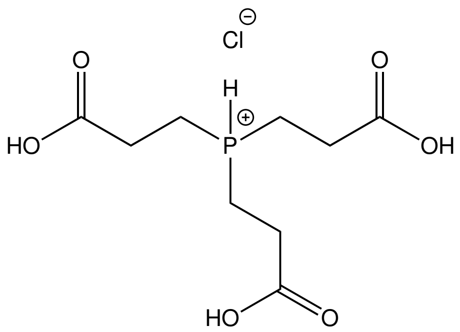


Figure 1.12 Chemical structure of commercially available TCEP/HCl. TCEP contains three carboxyl moieties (pKa values: 2.87, 3.59, 4.31[133]). The phosphorus pKa of TCEP is 7.68. Thus, at experimental pH values around 5.5, most TCEP molecules were deprotonated at all three carboxyl groups, while the phosphorus was protonated, leading to a net charge of -2.

Therefore, addition of a sufficient concentration of any anion was expected to increase the population of the folded state of BrkDBD at low temperatures as well. Even the addition of 100 mM NaCl to unfolded BrkDBD dissolved in acetate buffer of low concentration (5 mM) caused folding of BrkDBD (Table 1.2). Generally, the observation that the folded state is populated at higher salt concentrations is striking, even though samples at high salt concentrations were more prone to precipitation, which makes quantitative comparisons less reliable (Table 1.2). It is noteworthy that precipitation could be attenuated by additional precautions (working in a glovebox), minimizing potential oxydation contributions to the precipitation process.

In contrast to the already published data of apo BrkDBD[41], where no purely folded state spectrum was observed even at 5 °C, we expected that it would be possible to observe only the folded state at high salt concentrations and at higher charge density of buffer anions than in 200 mM acetate buffer. This was indeed the case at a succinate concentration of 95 mM (Figure 1.13) which is nearly identical to 200 mM acetate buffer with respect to ionic strength. However, the ratio of folded to unfolded BrkDBD is clearly much higher at 95 mM succinate concentration. For oxalate buffer, even a concentration of 20 mM was sufficient to observe solely folded BrkDBD at 5 °C (Table 1.2 and Figure 1.13). The ionic strength is even less for 20 mM oxalate than for 95 mM succinate buffer. Despite the severe reduction of protein stability at high anion concentrations and the reversibility problem after temperature jumps, the hypothesis that BrkDBD can be forced to fold by a change of electrostatics, and that rather charge density than purely ionic strength is the main determinant of the conformational energy landscape of BrkDBD, has been confirmed.

A closer inspection of folded BrkDBD in various buffers at 5 °C revealed that most of the resonances experience minor chemical shift perturbations while residues H80-Q83 were much more affected (Figure 1.13). Quantitative average ^{1}H-^{15}N shift changes ($\Delta\delta_{\text{avg}}(^{1}\text{H}\text{-}^{15}\text{N})$) in spectra from 20 mM to 95 mM succinate revealed major changes for residues H80-Q83, which compose the binding interface for specific DNA (Figure 1.14). In line with the CD results, the folded state chemical shifts in 20

buffer	reducing agent	% folded	% unfolded	stability
20 mM acetate	5 mM DTT	0 (0)	100 (100)	stable
20 mM phosphate + 10 mM NaCl	10 mM TCEP + 3 mM DTT	50 (50)	50 (50)	precipitation
20 mM succinate	10 mM TCEP	60 (64)	40 (36)	precipitation
20 mM succinate	5 mM DTT	71 (74)	29 (26)	precipitation
95 mM succinate	5 mM DTT	100 (100)	0 (0)	precipitation
200 mM acetate	5 mM DTT	74 (70)	26 (30)	precipitation
20 mM oxalate	5 mM DTT	100 (100)	0 (0)	stable at 5 °C
20 mM acetate	3 mM DTT	0 (0)	100 (100)	stable
20 mM acetate + 5 mM succinate	3 mM DTT	22 (23)	78 (77)	precipitation
20 mM acetate + 10 mM succinate	3 mM DTT	41 (50)	59 (50)	precipitation
20 mM acetate +	3 mM DTT + 6 mM TCEP	34 (38)	66 (62)	slight precip.
20 mM acetate +	3 mM DTT + 20 mM TCEP	39 (39)	61 (61)	precipitation
20 mM acetate + 20 mM phosphate	3 mM DTT + 20 mM TCEP	54 (55)	46 (45)	precipitation
20 mM acetate + 50 mM phosphate	3 mM DTT + 20 mM TCEP	63 (71)	37 (29)	precipitation
20 mM acetate + 100 mM NaCl	3 mM DTT	22 (23)	78 (77)	severe precip.
5 mM acetate + 100 mM NaCl	5 mM DTT	83 (84)	17 (16)	precipitation
10 mM acetate + 20 mM phosphate	5 mM TCEP	80 (73)	20 (27)	precipitation

Table 1.2 Populations of folded and unfolded BrkDBD and the sample stability in various buffers. At high salt concentrations, the samples were prone to precipitation, particularly after exposure to high temperatures. The precipitation increased with higher protein concentrations. Populations were obtained from the intensities of folded and unfolded state resonances of F48 (G68), respectively. The exchange peaks were usually below the detection limit and not taken into account. However, comparison of absolute values is difficult due to different sample 'histories' (heating and cooling cycles, which showed that unfolding of BrkDBD is not fully reversible) and precipitation of the salt-induced folded state.

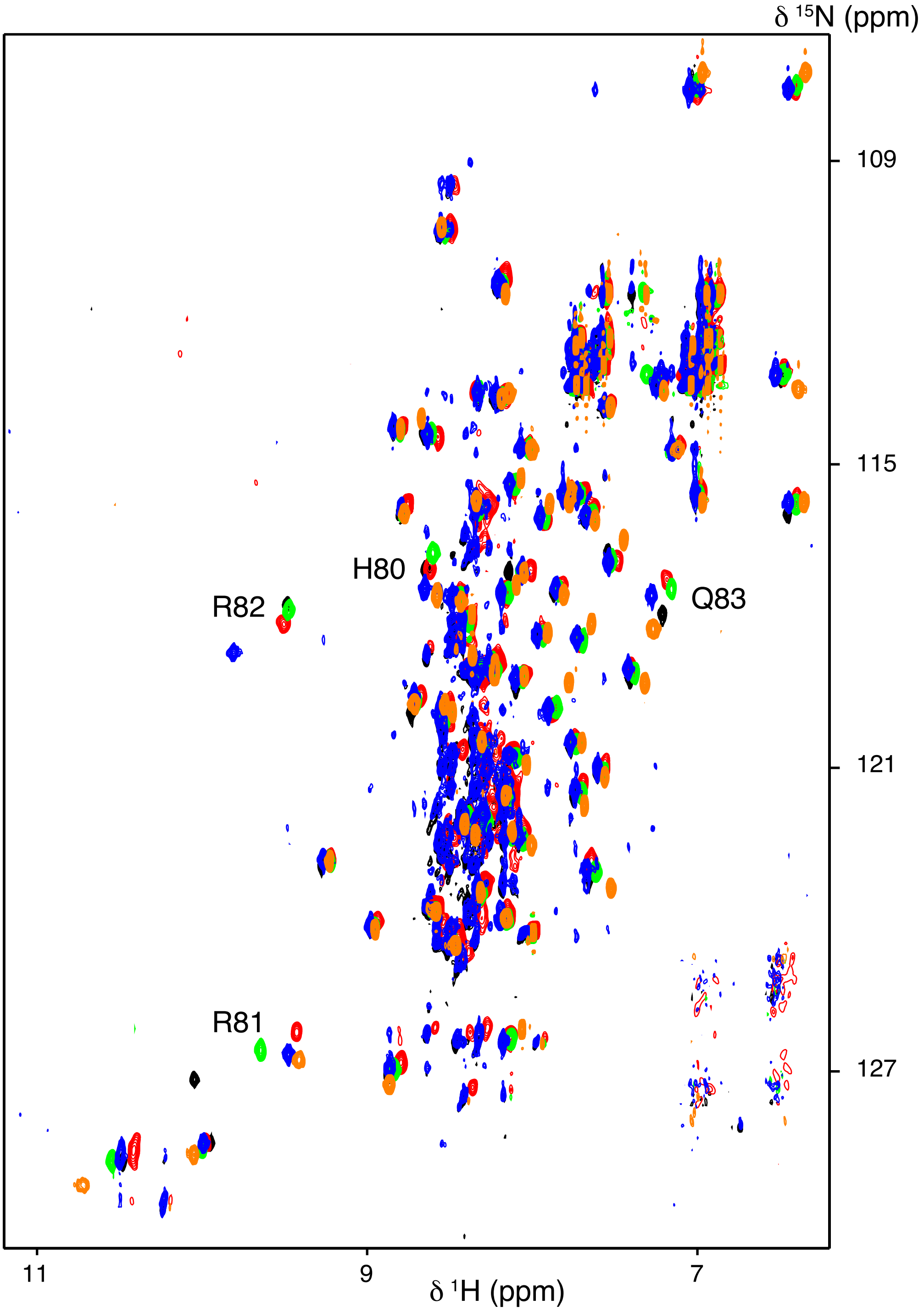


Figure 1.13 Superposition of ^{1}H-^{15}N HSQC spectra acquired at 5 °C and various buffer conditions. Colors indicate salt conditions as: 20 mM phosphate (black), 200 mM acetate (red), 20 mM succinate (blue), 95 mM succinate (green), 20 mM oxalate buffer (orange). The annotated residues experience the most pronounced chemical shift changes upon variation of salt type and concentration.

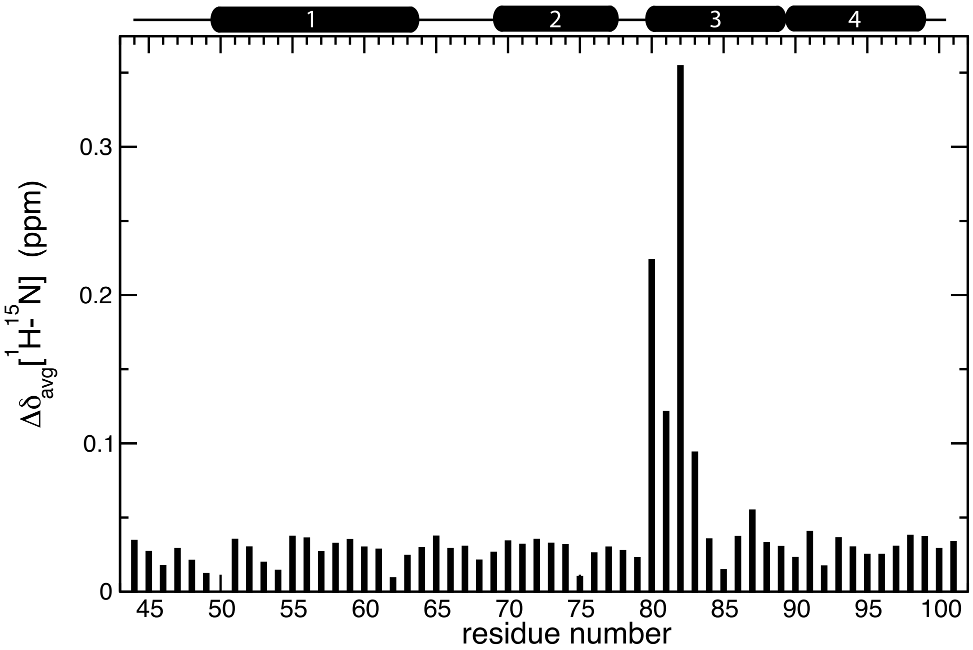


Figure 1.14 Averaged ^{1}H-^{15}N chemical shift difference between BrkDBD dissolved in 95 mM and 20 mM succinate from HSQC spectra acquired at 5 °C. Residues, which interact with the DNA major groove in the BrkDBD-omb12T5 complex show the most pronounced differences.

mM oxalate differ from those in other buffers, indicating more pronounced changes of their local environment of the backbone NH groups (Figure 1.13 orange). The reason for this may be a higher α-helical propensity of BrkDBD dissolved in 20 mM oxalate, indicated by the convergence to a plateau in the corresponding CD experiment (Figure 1.4).

BrkDBD in the presence of DNA

Due to the anion-dependence of the population of BrkDBD, we expected that BrkDBD would also fold in the presence of unspecific DNA, which also contains a negatively charged phosphate backbone. BrkDBD adopts indeed a completely folded state based on our ^{1}H-^{15}N HSQC data (Figure 1.15) in complex with a representative BrkDBD-unspecific DNA (AHD-DNA, which is specific for antennapedia homeodomain[68,186]). However, superposition of the spectra of BrkDBD in complex with specific, GC-rich DNA (omb12T5) and unspecific, AT-rich AHD-DNA indicates different local environments for many residues (Figure 1.15). A stretch of BrkDBD residues at the N-terminal part of helix 3 (H80-Q83) as well as Q70 and R71 (N-terminal part of helix 2), which make specific contacts with omb12T5 DNA, could not be detected in complex with AHD-DNA at 25 °C. This indicates that these residues undergo conformational exchange resulting in broadened resonances, which are below the detection limit.

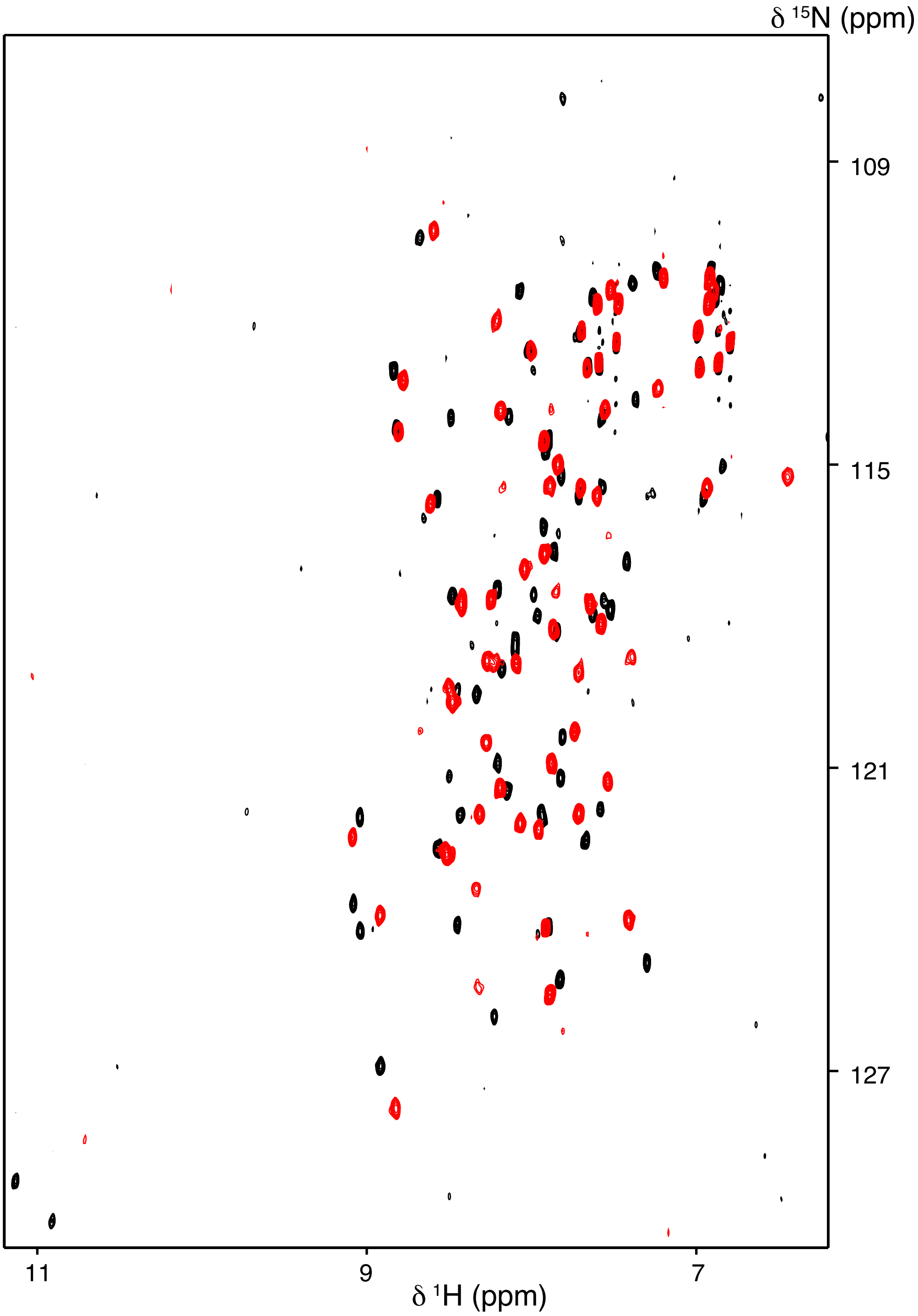


Figure 1.15 Superposition of ^{1}H-^{15}N HSQC spectra acquired at 25 °C in presence of specific omb12T5 DNA (black) and unspecific AHD-DNA (red). Most resonances experience relatively strong chemical shift changes, indicating different local environments.

Two of these residues (R81 and R82) were also not detected in complex with specific omb12T5 DNA, again indicating conformational exchange. This is corroborated by the faster T_2 relaxation times for residues close to H80[41].

In this context, it seems particularly interesting that resonances H80-Q83 of the major groove binding residues experience the most pronounced chemical shift changes in apo BrkDBD depending on buffer conditions and R82 was undetectable in 20 mM oxalate buffer at low temperatures. Clearly, the enhanced R_2 relaxation rates, as well as the invisibility of these residues under certain conditions indicates additional conformational exchange in this region of the protein.

Presumably, the recognition motif formed by these residues has intrinsically high dynamics which accelerates its search process and facilitates its probing for contacts with the target site. After binding to the recognition site, the conformational exchange of these residues is more restricted due to specific interactions with the binding partner. This would explain the detection of residues Q70, R71, H80 and Q83 in the presence of omb12T5 DNA in contrast to AHD-DNA.

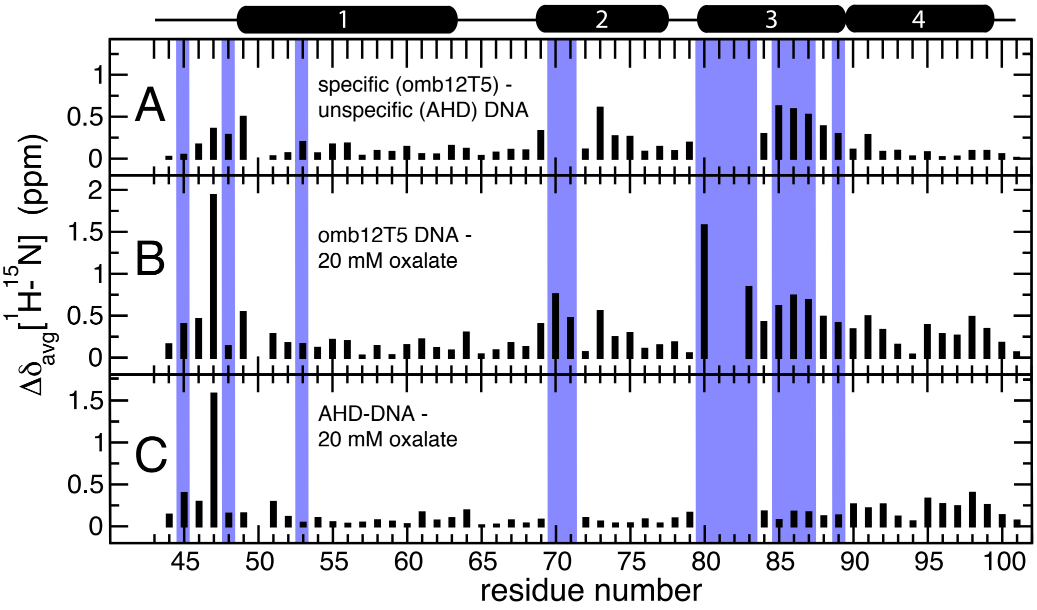


Figure 1.16 Average ^{1}H-^{15}N chemical shift differences ($\Delta\delta_{\text{avg}}[^1\text{H} - ^{15}\text{N}]$) of folded BrkDBD in the presence of DNA or salt. BrkDBD residues involved in binding of specific omb12T5 DNA are highlighted in blue. Major differences between spectra of specific and unspecific DNA (**A**) are observed in close proximity to the residues which make specific contacts to omb12T5 DNA, including the full recognition helix 3. In contrast, resonances of helices 1 and 4 are nearly identical. While BrkDBD resonances in 20 mM oxalate differ significantly from those in the specific DNA complex (**B**), they are very similar to the unspecific DNA complex (**C**). Positively charged residues on the N-terminal end of BrkDBD (R45, R46) interact with the DNA minor groove in the presence of DNA. This also affects I47, which may interact with other hydrophobic residues in the absence of DNA.

A comparison of the chemical shifts of BrkDBD-AHD-DNA-complex with those of folded BrkDBD in 20 mM oxalate buffer reveals a lot of similarities (Figure 1.16 C). Nearly negligible differences are found for residues in helices 1 and 2. Only slightly larger differences are observed for helix 3. The large differences close to the N-terminus of BrkDBD may be caused by electrostatic interactions of residues R45 and R46 upon binding to the minor groove, which also affects I47. In the absence of DNA, electrostatic interactions of the N-terminal tail of BrkDBD with buffer anions are not restricted to a certain spatial position corresponding to the DNA minor groove. This increased conformational freedom could lead to a reorientation of the N-terminus in the absence of DNA, which may favour hydrophobic interactions of I47 with other hydrophobic protein residues.

In contrast, very pronounced chemical shift differences are found between the specific BrkDBD-DNA complex and the folded state in the presence of 20 mM oxalate (Figure 1.16 B). These differences are maximal for residues located at the N-terminus (due to the interaction with the minor groove) and helices 3 and 4. Additionally several residues in helix 2 show differences.

The averaged ^{1}H-^{15}N shifts of BrkDBD in 20 mM oxalate buffer have nearly identical differences to both DNA complexes for residues at the N-terminus as well as helix 4, indicating that the local environment in oxalate is significantly different at these sites. Averaged ^{1}H-^{15}N chemical shift differences between the two DNA complexes of BrkDBD were most pronounced for the detectable part of helix 3 and also for residues in relatively close vicinity to the other DNA binding sites of BrkDBD (Figure 1.16 A). Residues located in two interaction sites (Q70, R71 and H80-Q83) are not visible in the presence of AHD-DNA. Only relatively small differences are observed for resonances in helix 1 and the loop between helices 1 and 2.

Taken together, these findings illustrate that BrkDBD in complex with specific DNA shows pronounced differences to the oxalate-induced folded state. The folded state in the presence of 20 mM oxalate is very similar to the complex with unspecific DNA. Therefore, the specific interactions with DNA lead to a significantly different local environment.

It has to be mentioned that ^{1}H-^{15}N HSQC spectra of BrkDBD in complex with DNA at 5 °C and 25 °C are very similar and no pronounced resonance broadening has been observed in contrast to all investigated DNA-free buffer conditions at higher temperatures. It is noteworthy that NMR samples with a BrkDBD concentration of more than 100 μM in complex with AHD-DNA showed significantly reduced stability, even at low temperatures, while maximal stability was observed for BrkDBD in complex with omb12T5 DNA at equimolar concentrations.

In band-shift (PAGE) experiments, only weak bands have been observed for the BrkDBD-AHD-DNA complex. This may indicate a much weaker binding than the reported nanomolar dissociation constant for the BrkDBD-omb12T5 complex[41]. However, although no precipitates have been observed optically, the stability problems of BrkDBD in AHD-DNA could have influenced this result. Therefore, no quantitative binding affinity between BrkDBD and AHD-DNA could be determined, but it is likely that the presence of purely folded state as indicated by NMR is a result of electrostatic effects as observed for apo BrkDBD at high salt concentrations rather than DNA binding. Even though equistoichiometric NMR samples of BrkDBD and omb12T5 DNA proved to be stable for a long time, samples of unequal concentrations of either omb12T5 or AHD-DNA and BrkDBD always precipitated, even at low micromolar concentrations.

Secondary chemical shift analysis

The presence of folded and unfolded states of BrkDBD have been discussed with respect to two-dimensional ^{1}H-^{15}N HSQC spectra which yield information about individual NH groups, and CD data which provide average information on all molecules in solution, but do not describe the behavior of individual residues. Differences of individual chemical shifts of a residue (comprising all types of atoms) to so-called random coil chemical shifts, usually determined on short denatured model polypeptides[267], are commonly used to obtain information about secondary structure propensities. These differences are referred to as secondary chemical shifts.
^{13}Cα secondary chemical shifts have proven to be an exceptionally useful measure for secondary structure determination by NMR spectroscopy[267]. Positive ^{13}Cα secondary chemical shifts indicate α-helical propensity and are thus well-suited for the characterization of folded and unfolded BrkDBD, due to its helix-turn-helix fold. Experiments such as 3D-HNCA or 3D-CBCA(CO)NH are measured routinely for resonance assignment of proteins and serve not only as source of information to establish sequential connectivities, but yield also ^{13}Cα chemical shifts. To assess structural propensities, secondary chemical shifts under various buffer conditions have been determined for both folded and unfolded BrkDBD.

The ^{13}Cα secondary shifts of BrkDBD in complex with both specific omb12T5 DNA and unspecific AHD-DNA correlate well with the secondary structure of the folded protein (Figure 1.17). Only negligible differences between the ^{13}Cα chemical shifts indicate that the protein adopts the same fold in both DNA complexes. This observation proves that BrkDBD does not only fold in the presence of unspecific DNA, but also adopts the native secondary structure. Thus, the differences detected in the ^{1}H-^{15}N HSQC spectra are likely caused by small local differences due to interactions with DNA, which do not significantly alter the secondary structure of BrkDBD.
Nearly identical ^{13}Cα secondary shifts as for both DNA complexes are also observed for the folded state of apo BrkDBD in 20 mM succinate buffer at 5 °C (Figure 1.17 A). Therefore, we conclude that BrkDBD always adopts the same helix-turn-helix secondary structure in the folded state.
Based on the positive ^{13}Cα secondary chemical shifts, unfolded BrkDBD shows some helical propensity at all investigated buffer and temperature conditions (Figure 1.17 B). This propensity seems very pronounced around helix 2 and to a slightly reduced extent at the positions of native helices 3 and 4. Remarkably, the N-terminal part of helix 2 seems to be extended, spanning from amino acids D63 to N69. This is interesting, since these residues form a loop in folded BrkDBD which connects

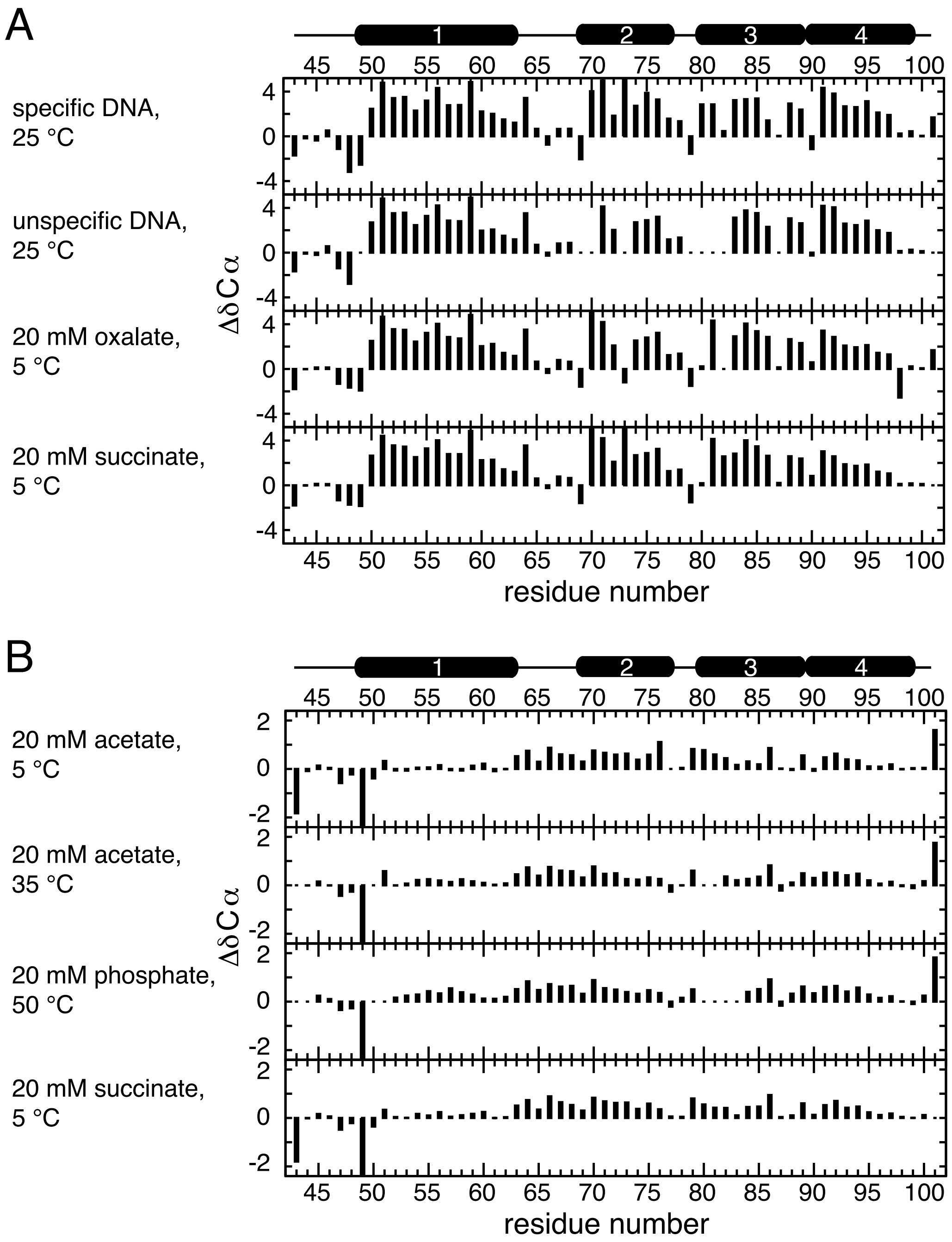


Figure 1.17 ^{13}Cα secondary chemical shifts of folded (A) and unfolded (B) BrkDBD at various conditions. The secondary structure of the native state is indicated on top. The large negative ^{13}Cα secondary chemical shift of T49 (preceding P50) can be explained by the used reference set of ^{13}Cα random coil shifts from Wishart *et al.*[267], which are not corrected for sequence-dependence. The nearest neighbor effects have been shown to be strongest (about -2.0 ppm for ^{13}Cα chemical shifts[218]) for residues that precede proline due to the restricted conformational freedom of the proline ring and the absence of an amide proton.

helices 1 and 2 *via* a hydrogen bond network, involving residues D63, D65, C66 and K76. Virtually no helical propensity is observed at the position of helix 1 of the native state. The ^{13}Cα secondary chemical shifts of unfolded BrkDBD in 20 mM acetate and 20 mM succinate are virtually identical at 5 °C, corroborating the observation of very similar ^{1}H-^{15}N HSQC spectra.

Surprisingly, nearly identical ^{13}Cα secondary chemical shifts are observed for unfolded BrkDBD at various buffer conditions and temperatures (Figure 1.17 B). This indicates that about 20% of the unfolded state ensemble adopts α-helical structure at helices 2-4 and helix 2 is extended at the N-terminus, including residues D63-N69, independent of the tested experimental conditions.

Secondary chemical shifts have also been calculated for ^{1}HN, ^{15}N and ^{13}CO[21,267]. While the ^{13}CO and ^{13}Cα secondary chemical shifts correlated well with the secondary structure of folded BrkDBD, the other two nuclei showed several differences. This can be attributed to the reduced sensitivity of ^{1}HN and ^{15}N on secondary structure[267].

^{15}N relaxation dispersion

Due to the presence of microsecond to millisecond conformational exchange evident from the severe line broadening in unfolded BrkDBD, the requirements for the application of ^{15}N relaxation dispersion NMR experiments were fulfilled. Therefore, identical ^{15}N relaxation experiments were performed at two magnetic fields (800 and 600 MHz). Non-uniform ^{15}N relaxation dispersion curves were observed for BrkDBD in 20 mM acetate at 5 °C at 80 ms of relaxation dispersion time (Figure 1.18). In addition to residues at the termini of BrkDBD (S44 to R46 and N101), only few other resonances (N69, W87 and N93) showed flat dispersion profiles. All other residues showed dispersion curves, which were fitted to a two-state exchange model (by an in-house written Matlab script, according to eq. 1.3) yielding excellent fits. Some representative fit curves are shown in Figure 1.19. With this fitting procedure, the exchange rate (k_{ex}), the population of the invisible state (p_b) and the chemical shift difference between the invisible and the dominant state ($|\Delta\omega|$) were obtained for each residue. Due to the quadratic dependence of the effective R_2 relaxation rate on $\Delta\omega^2$, the sign of the chemical shift difference cannot be determined by this type of experiments.

The obtained $|\Delta\omega|$ parameters correlated reasonably well with the ^{15}N chemical shift difference between folded and unfolded states of BrkDBD at 5 °C in 20 mM succinate buffer (Figure 1.20). Only few outliers apart from those with spectral overlap have been observed. The few exceptions include residues located at the N-terminal part of helix three which make specific contacts to cognate DNA in the BrkDBD-omb12T5 complex[41]. Combined with T_2 relaxation data of BrkDBD in complex with specific DNA[41] and apo BrkDBD in 20 mM oxalate buffer, this strongly indicates that these residues experience additional conformational exchange. For these residues, the two-state exchange model is not appropriate.

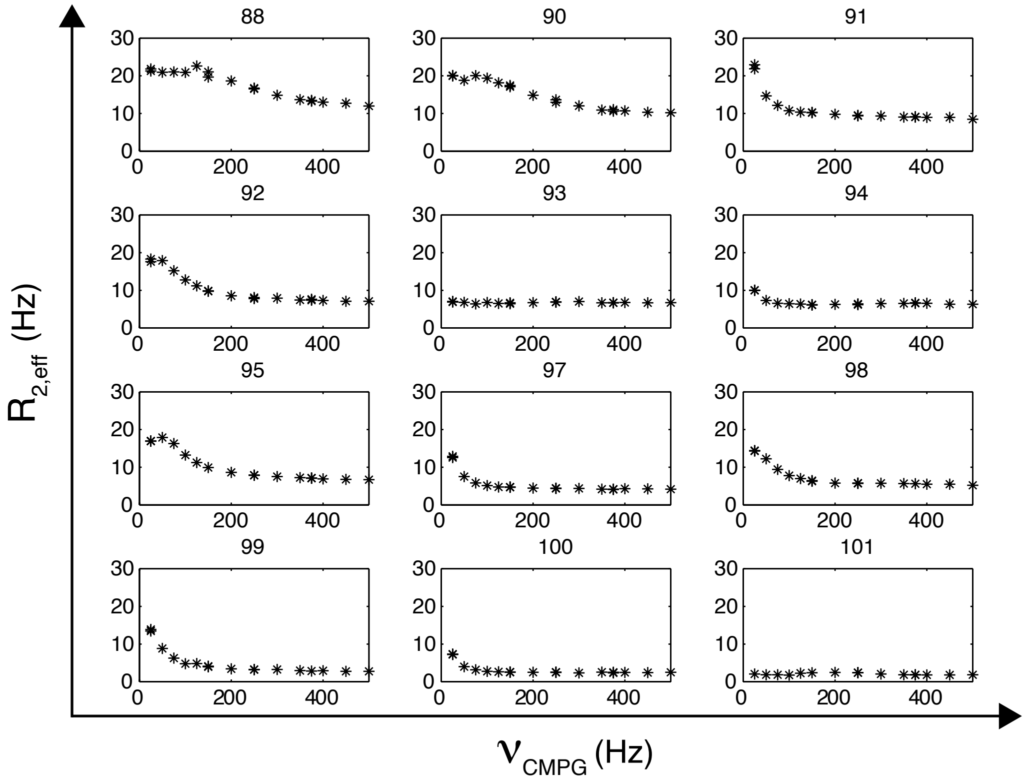


Figure 1.18 Relaxation dispersion curves of unfolded BrkDBD in 20 mM acetate at 5 °C. The effective relaxation rate ($R_{2,\text{eff}}$) as a function of the CPMG pulse frequency (ν_{CPMG}) shows non-uniform conformational exchange along the amino acid sequence. Residue numbers are indicated at the top of each plot.

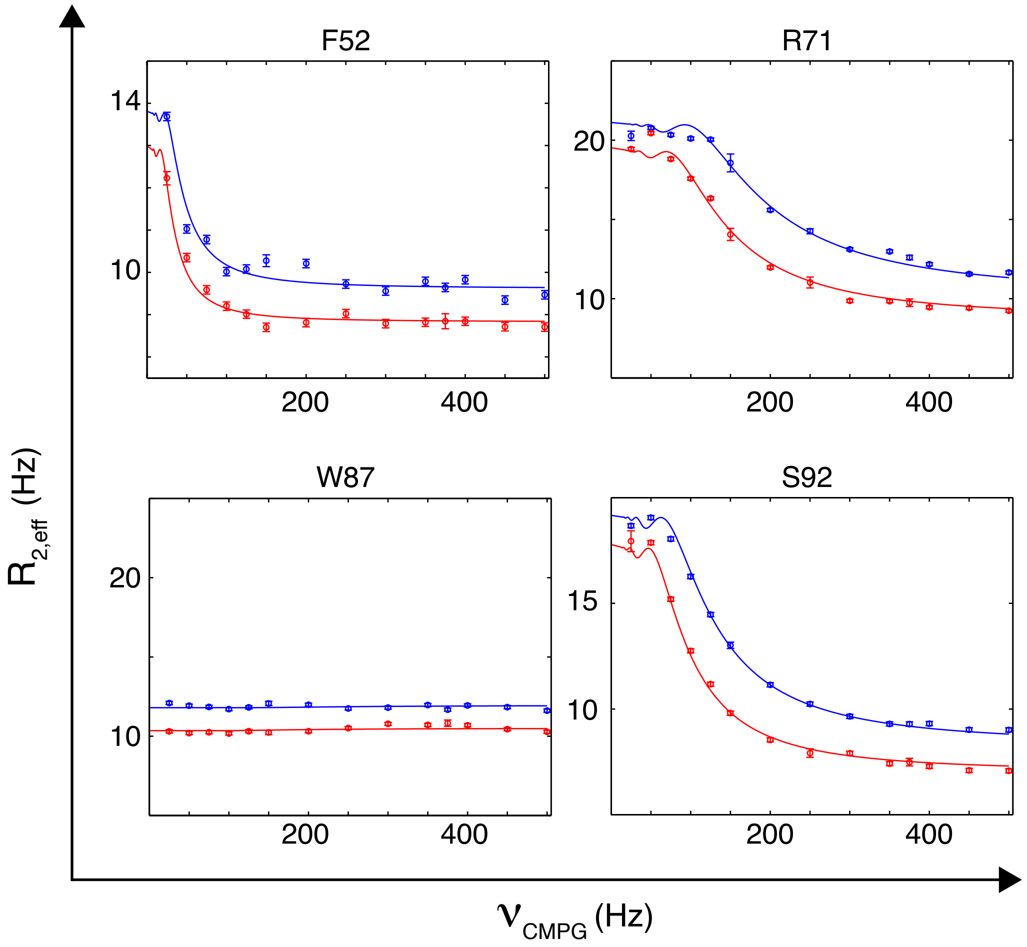


Figure 1.19 Representative fitted relaxation dispersion curves of unfolded BrkDBD in 20 mM acetate at 5 °C. The effective relaxation rate ($R_{2,eff}$) as a function of the CPMG pulse frequency (ν_{CPMG}) is shown for measurements at two magnetic fields (blue: 800 MHz; red: 600 MHz). Residue type and number are displayed at the top of each plot. The solid lines correspond to the back-calculated $R_{2,eff}$ rates based on the fit parameters.

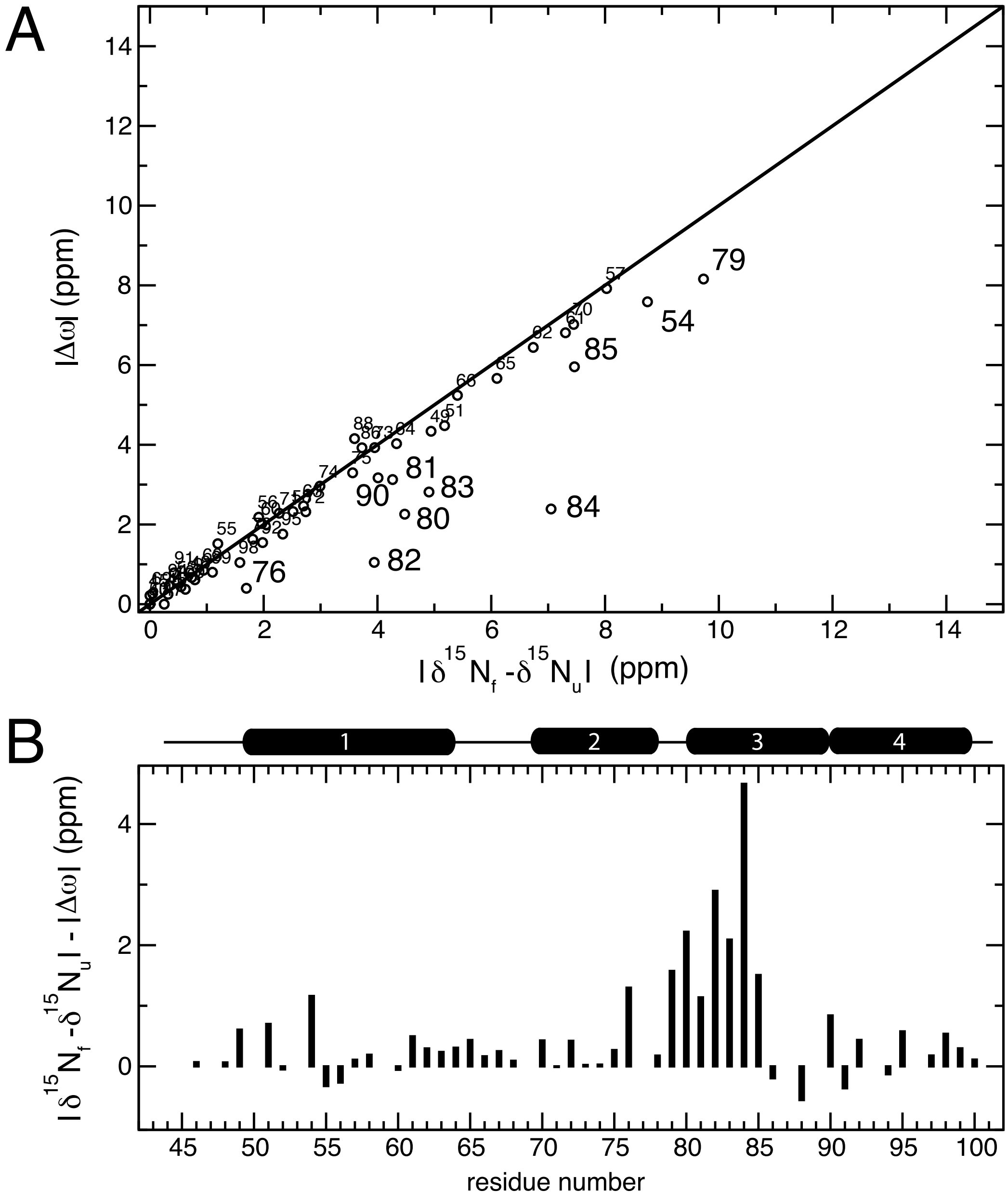


Figure 1.20 Relation between $|\Delta\omega|$ from unfolded BrkDBD and the ^{15}N chemical shift difference between folded and unfolded BrkDBD ($|\delta^{15}\mathrm{N_f} - \delta^{15}\mathrm{N_u}|$). $|\Delta\omega|$ parameters were obtained by fitting the ^{15}N relaxation dispersion profiles of unfolded BrkDBD in 20 mM acetate buffer measured at two different magnetic fields. **A:** $|\Delta\omega|$ as a function of ($|\delta^{15}\mathrm{N_f} - \delta^{15}\mathrm{N_u}|$) obtained in 20 mM succinate at 5 °C. The nearly linear correlation for most residues is consistent with a two-state folding model. **B:** The difference between $|\delta^{15}\mathrm{N_f} - \delta^{15}\mathrm{N_u}|$ and $|\Delta\omega|$ as a function of the residue number indicates that the largest deviations are observed for residues in the DNA recognition helix 3.

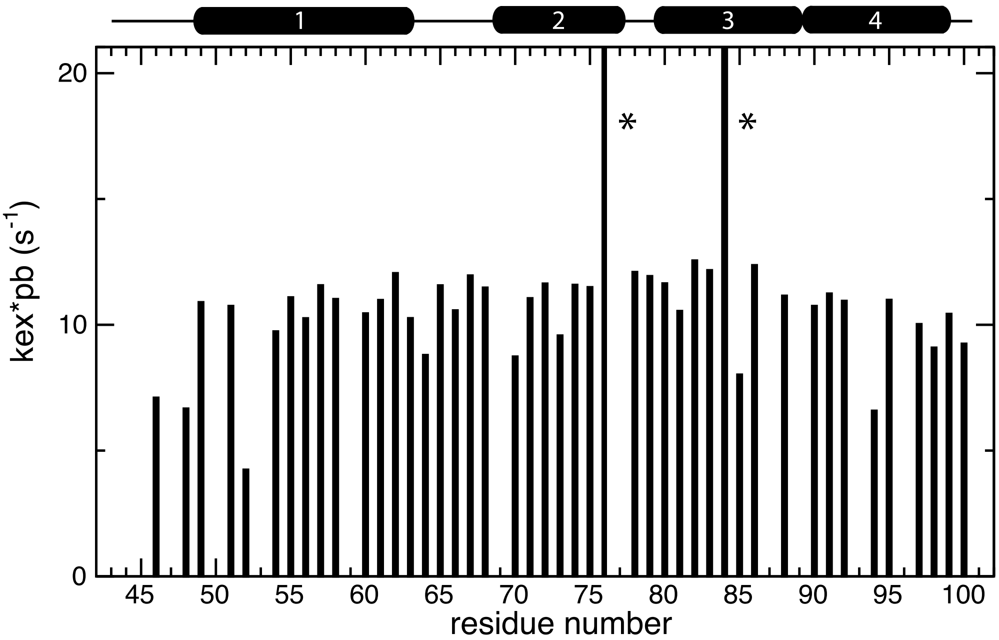


Figure 1.21 The product of fitted exchange rates (k_{ex}) and the population of the invisible state (p_b) as a function of residue number. Only few outliers are observed, indicating that the majority of residues shows a uniform exchange behavior, which is consistent with a two-state folding process. Residues marked by an asterisk show high χ^2 values, possibly caused by additional exchange.

Even though individual exchange rates and populations differed significantly, the product of these two parameters was very similar for nearly all residues (Figure 1.21). Thus, measurements at two magnetic fields were not sufficient to quantitatively determine all the parameters. Nevertheless, the linear correlation between $|\Delta\omega|$ and ^{15}N chemical shift differences for most of the residues corroborates the two-site exchange, which had been suggested in the literature for small single domain proteins[9,111,145].

However, based on the available data at this time, it cannot be completely excluded that the low-populated state, which exchanges with the dominant unfolded BrkDBD, could be a folded-like conformation with slight local differences to the native state. This issue may be addressed by acquisition of additional relaxation dispersion data comprising ^{13}Cα, ^{13}CO, ^{1}HN and ^{1}Hα.

Alternatively, relaxation dispersion of folded BrkDBD could reveal whether the $|\Delta\omega|$ parameters again match the chemical shift differences between resonances from the folded and the unfolded state. Therefore, ^{15}N relaxation dispersion curves at 600 MHz were also measured for folded BrkDBD in the presence of 95 mM succinate at 5 °C (Figure 1.22). However, these data could not be analyzed quantitatively, since the protein precipitated during the actual relaxation dispersion and the reference experiments. Nevertheless, qualitative analysis indicated the presence of exchange for most residues and the absence of exchange for residues S44 to I47 at the N-terminus, N101 at the C-terminus and residues E91, N93 and L94, located in helix 4. The reason for the flat dispersion profiles of the latter three residues remains unclear, but it is remarkable that N93 is neither subject to ^{15}N relaxation dispersion in the folded nor in the unfolded state. All other residues showed dispersion profiles which were not a consequence of precipitation, since the experiment was recorded in an interleaved fashion, providing nearly identical conditions for individual resonances.

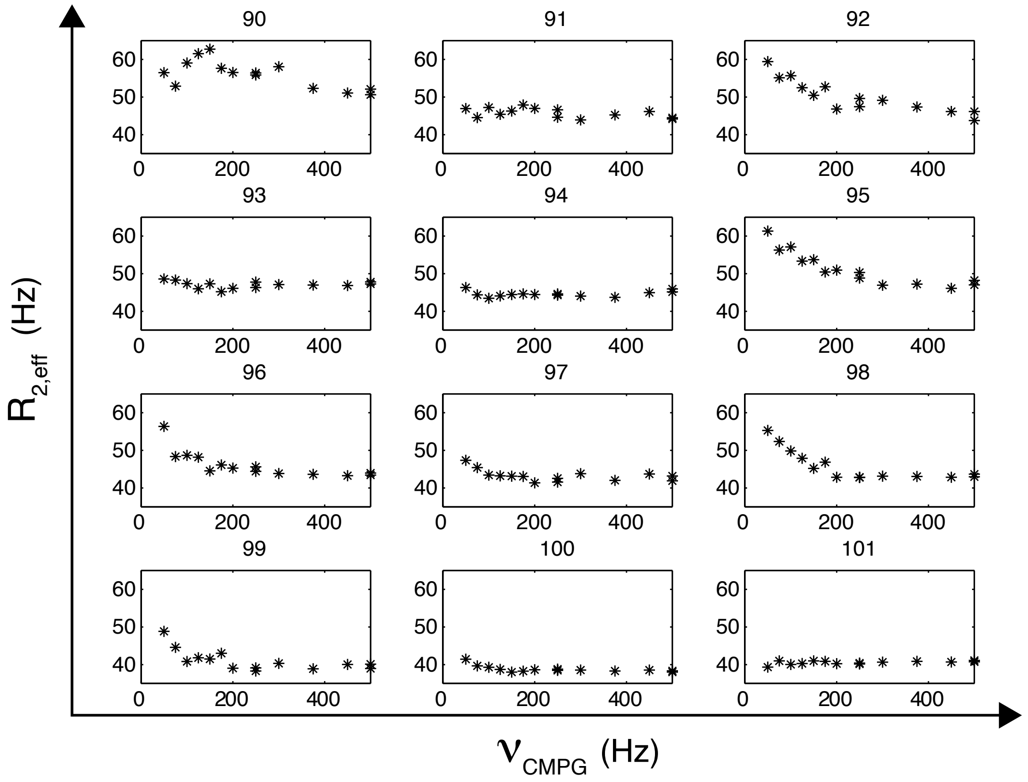


Figure 1.22 Relaxation dispersion curves of folded BrkDBD in 95 mM succinate at 5 °C (600 MHz magnet). The effective relaxation rate ($R_{2,eff}$) as a function of the CPMG pulse frequency (ν_{CPMG}) shows non-uniform conformational exchange along the amino acid sequence. However, the absolute values of $R_{2,eff}$ are less reliable due to protein precipitation. Residue numbers are indicated at the top of each plot.

For apo BrkDBD in the folded form (succinate) or unfolded form (acetate), clear dispersion profiles can be observed. In contrast, no dispersion curves could be detected for BrkDBD in complex with both specific and unspecific DNA at 25 °C. This indicates either a severe restriction of conformational exchange or exchange outside of the microsecond to millisecond regime. Possibly, higher temperatures could induce exchange or modulate the exchange regime also in the presence of DNA, which then could be studied by relaxation dispersion.

Taken together, we conclude that BrkDBD shows strong microsecond to millisecond conformational exchange in the absence, but not in the presence of DNA, for both folded and unfolded states. Unfolded BrkDBD is clearly in a two-state exchange equilibrium with the folded state due to the linear correlation of $|\Delta\omega|$ with the ^{15}N chemical shift difference of these two states for most residues.

MTSL labeling and mutants of BrkDBD

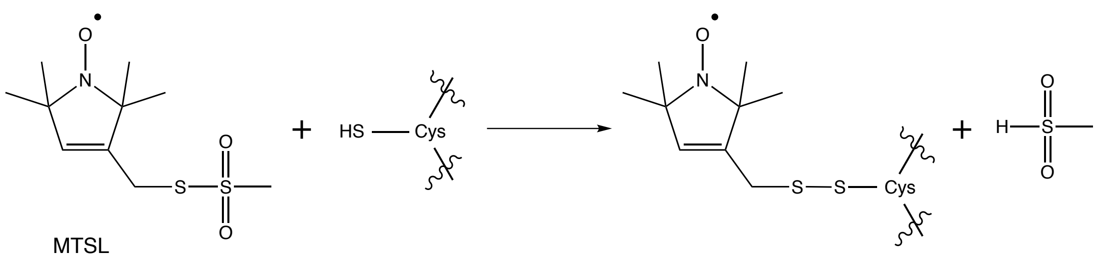


Figure 1.23 Scheme of the reaction of the MTSL spin label with a cysteine side chain.

In order to probe for transient interactions between BrkDBD and DNA and their influence on the exchange behavior, we tried a similar approach as Iwahara *et al.*[108]. In that approach, PREs were measured for a homeodomain (HOXD9) in complex with DNA which carried a strategically positioned spin label (radical). For their method, fast exchange between bound and unbound protein was required, which was the case in the presence of 160 mM NaCl. The PRE profiles, arising from transient encounter complexes, were detected for spin labels at 4 different DNA sites and indicated that no free protein was populated, while intra- and intermolecular translocation occured with the bound state in the fast exchange regime. PREs of mixtures of labeled and unlabeled (specific and unspecific) DNA with the protein gave evidence that indeed sliding (intramolecular) and hopping (intermolecular) translocation occured with both specific and unspecific DNA[108,110].

In order to study such encounter complexes, labeling of either protein or DNA with a stable radical was required to measure PREs and detect the formation of transient encounter complexes. To this aim, we tested labeling of BrkDBD with the MTSL spin label (Figure 1.23). Both cysteine residues (C66 and C90) of wild type (wt) BrkDBD were expected to react with MTSL. The success of the labeling reaction was confirmed by mass spectrometry and indicated that two MTSL moieties had been attached to the protein. As expected, only few resonances where observed in the corresponding ^{1}H-^{15}N HSQC spectrum due to bleaching of resonances, caused by the PRE effects of both spin labels.

The spin labels were quenched by addition of ascorbic acid in order to recover the diamagnetic resonance intensities. This facilitated the comparison of the ^{1}H-^{15}N HSQC spectra of quenched, MTSL-labeled and not MTSL-labeled BrkDBD. Purely

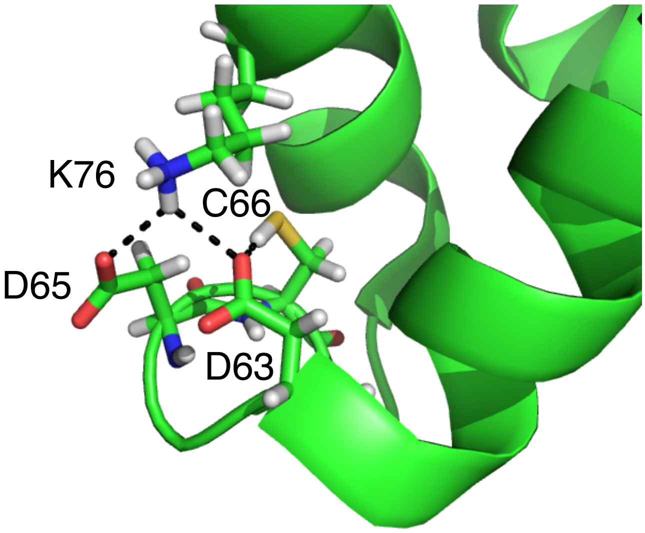


Figure 1.24 Hydrogen bond network comprising D63, D65, C66 and K76, extracted from the BrkDBD pdb structure (2GLO.pdb[41]). Hydrogen bonds are indicated by dashed lines. The cognate omb12T5 DNA was removed for clarity.

unfolded BrkDBD has been observed for the quenched MTSL-labeled BrkDBD in 20 mM acetate buffer based on the ^{1}H-^{15}N HSQC spectrum, as expected.

In order to investigate folded BrkDBD, the MTSL labeling was repeated at high salt concentration (95 mM succinate) at low temperature (4 °C). The ^{1}H-^{15}N HSQC again showed only few resonances due to the PREs of the spin labels. However, only unfolded BrkDBD was observed after quenching the MTSL radicals. This was unexpected since unlabeled BrkDBD was always folded at such high succinate concentrations.

To further investigate this phenomenon and since single spin labeling was required at selected positions, we mutated cysteine 90 to serine (C90S BrkDBD). C90 is located at the kink between helices 3 and 4 and its side chain is solvent acessible. The C90S BrkDBD mutant showed an identical expression and yield as wt BrkDBD and the ^{1}H-^{15}N HSQC spectra were very similar. At high salt concentrations, the (not MTSL-labeled) C90S mutant was folded based on the resonance dispersion in the ^{1}H-^{15}N HSQC NMR spectrum. However, after quenching of the (C66) MTSL-labeled C90S BrkDBD, the ^{1}H-^{15}N HSQC experiment showed again only unfolded protein, even though it was folded before the reaction with MTSL.

These observations indicate that the reaction of MTSL with C66 prevents wt BrkDBD and the C90S mutant from folding. This was initially unexpected due to the localization of C66 in the loop between helices 1 and 2 in the folded structure of BrkDBD.

A closer inspection of the 20 NMR structures of BrkDBD (pdb code 2GLO) revealed that all structures showed a hydrogen bond network involving residues D63, D65, C66 and K76, which connects the C-terminal ends of helices 1 and 2. In this network, NHζ of K76 and SHγ of C66 are hydrogen bond donors and Oδ of D63 and D65 are acceptors, respectively (Figure 1.24). While K76 interacts with both D63 and D65, C66 only forms one hydrogen bond with D63.

Thus it seemed as if the contribution of C66 to the network was rather small. Furthermore, it is apparent from the structure that MTSL labeling of C66 would be incompatible with the folded structure, due to steric clashes. Consequently, the loop would have to reorient, leading to an increase in K76 to D63 and D65 distances which would eventually disrupt this entire network of hydrogen bond interactions. Therefore, a combination of both steric clashes and prevention of hydrogen bond formation may be the reason for the folding inability of BrkDBD variants with the MTSL label attached to C66.

In order to attach the spin label to C90, we mutated C66 to serine (C66S BrkDBD), which should be capable of forming the described hydrogen bond network. The yield of the expression of the C66S mutant was much lower than for wt and C90S BrkDBD and more impurities have been observed. Nevertheless, the identity of C66S BrkDBD was confirmed by mass spectrometry and a ^{1}H-^{15}N HSQC spectrum was recorded at 95 mM succinate concentration at 5 °C. Surprisingly, only unfolded C66S BrkDBD (not MTSL-labeled) was observed, even at such high succinate concentration.

In another attempt to force the C66S to adopt the folded structure, we added an equimolar concentration of omb12T5 DNA. About 95% of C66S BrkDBD precipitated subsequently. However, the ^{1}H-^{15}N HSQC spectrum remaining 5% vaguely resembled that of folded BrkDBD in complex with omb12T5. This indicates that the integrity of C66 is crucial for folding of BrkDBD.

MD simulations of C66 variants

Sequence alignment of BrkDBD from several organisms shows that C66 is highly conserved in contrast to C90[41]. We currently lack full understanding of the folding inability of the C66S mutant, since S66 could theoretically participate in the described hydrogen bond network in contrast to MTSL-labeled C66.

In order to get more insight into the difference between C66 mutants (C66S, C66A and C66N) and wild type BrkDBD, a series of all-atom MD simulations in explicit water has been performed. The starting structure for all mutants was that of folded BrkDBD (pdb code 2GLO), as previously determined by NMR[41] during the MD simulations. The length of each trajectory was 40 ns and the distance between Nζ of K76 and Cγ of D63 was measured for various time points during the simulations to indicate whether the hydrogen bond network remained intact.

Even wt BrkDBD showed slight conformational flexibility, since the Nζ-Cγ distance slightly increased, but subsequently reduced again to the original value of about 3 Å. Therefore, the hydrogen bond network remained intact for the majority of the simulation time of 40 ns, indicating that this interaction is stable (Figure 1.25). This observation was supported by 10 independent 40 ns trajectories.

In contrast, the hydrogen bond network of the C66S mutant was disrupted for about 50% of each trajectory in all 10 independent runs of 40 ns length. This may cor-

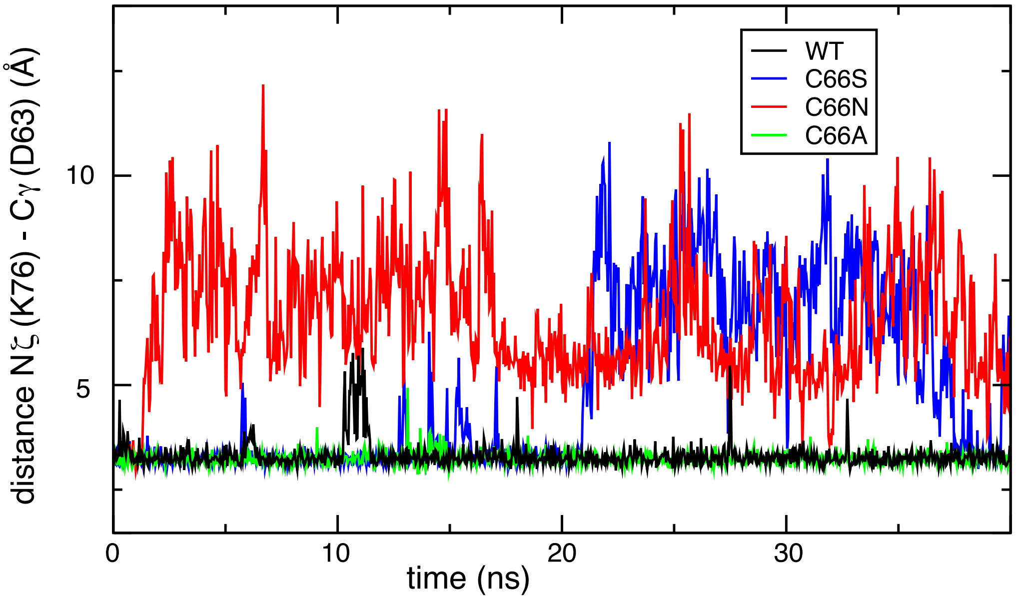


Figure 1.25 Distance between Nζ of K76 and Cγ of D63 as a function of time during MD simulations. Wild type (wt) BrkDBD and the C66A mutant show nearly constant distances of 3.2 Å, indicating that the hydrogen bond is stable. In contrast, C66S looses the hydrogen bond for 50% of the simulation time and C66N for nearly 100%.

respond to the experimental observation of solely unfolded C66S BrkDBD, but the reason remains elusive. Possibly, the different geometrical requirements of hydrogen bonds of sulphur compared to oxygen contribute to these observations. Generally, SH is a weaker hydrogen bond donor than OH due to the difference in electronegativity and the larger size of sulphur. Another reason may be the different preferences for the side-chain χ_1 angles. For unfolded regions, the prefered χ_1 angle of cysteine is -60 degrees, while that of serine is +60 (derived from the protein coil library according to Vajpai *et al.*, 2010[250]). However, the energetic barrier is not expected to be high enough to keep this angle fixed, particularly if a hydrogen bond can be formed upon a -120 degree rotation.

A complete loss of the hydrogen bond network has been expected for C66N, since the side chain of asparagine is sterically much more demanding than the side chain of cysteine or serine. Indeed, the contacts between K76 and D63 were lost immediately and never recovered during the 40 ns simulation. Combined with the experimental observation of unfolded protein in all BrkDBD mutants that contained an MTSL-labeled C66, it can be concluded that steric repulsion always disrupts these hydrogen bond interactions.

Such steric repulsions are not possible for C66A due to the small side chain. The MD simulation of the C66A BrkDBD mutant indeed had a similar trajectory as the wild type protein with respect to the K76-D63 hydrogen bond. An experimental verification did not succeed, since C66A did not overexpress in *E. coli* either due to cell toxicity or infection by bacteriophages. It still remains unclear from the MD simula-

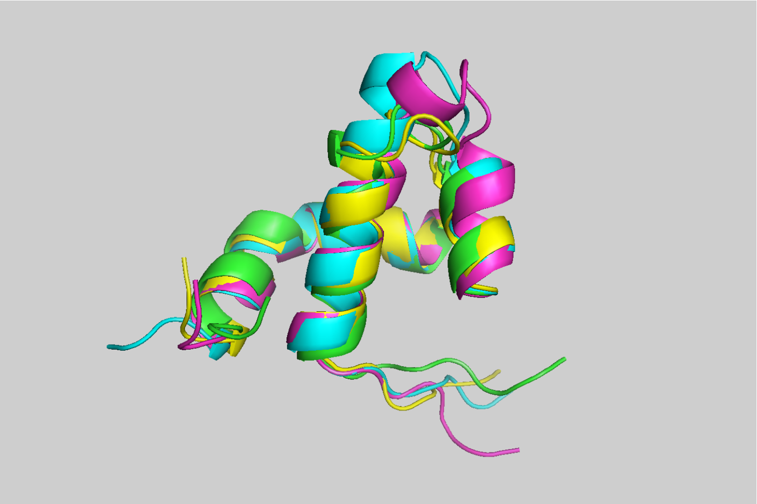


Figure 1.26 Superposition of 4 significantly different structures with similar energy, generated by CS-Rosetta with chemical shifts of apo BrkDBD. The experimental data are compatible with various conformations only in the region between native helix 1 and 2 (top of the Figure), while the other secondary structures are well-defined.

tions, whether the whole protein would really be unable to fold due to the loss of this network of hydrogen bonds.
Nevertheless, the fact that MTSL labeling of C66 as well as the C66S mutant are unable to fold indicates that the conserved loop between helices 1 and 2 plays a crucial role. Possibly, additional interactions in the unfolded state of both C66 MTSL-labeled BrkDBD variants and the C66S mutant could increase the α-helical propensity in this region and prevent the conformational transition to the native loop structure.

Chemical shift-based structure predicition allows assembly of a 3D structure determination by computational selection of short protein fragments from a database which are compatible with the experimental chemical shifts[30,221,266]. CS-Rosetta[221] prediction of 12500 structures based on ^{1}HN, 1Hα, ^{15}N, ^{13}Cα, and ^{13}CO chemical shifts of folded apo BrkDBD revealed that solely the loop around C66 is not precisely defined by the experimental data. These predicted structures indicate that the experimental data would also be compatible with an extended helix 1 or a single helix turn as shown in Figure 1.26. This may reflect the conformational flexibility required to proceed from a non-native structure with α-helical propensity in this region to the native loop with the specific intramolecular interactions in the folded state.

1.2.3 Conclusion and Perspective

Salt-dependence of BrkDBD folding
In conclusion, we found that folding of the Brinker DNA binding domain is strongly influenced by both anion type and concentration. At higher anion concentrations, the

folding energy landscape is modified such that the folded state is higher populated than at lower anion concentrations.

Longer chain lengths between two intramolecular carboxylic groups correlate with a lower denaturation temperature as indicated by CD measurements. Therefore, anionic buffer molecules do not mimic the phosphate backbone by a spatial separation of negative charges.

^{1}H-^{15}N HSQC NMR data reveal the presence of both folded and unfolded BrkDBD and additional resonances attributed to conformational exchange at 20 mM succinate concentration at 5 °C. Thus, conformational exchange is slow on the chemical shift time scale at these conditions, but fast enough to lead to the build-up of exchange resonances within several milliseconds.

Temperature series of ^{1}H-^{15}N HSQC spectra at various buffer conditions corroborate the observations by CD. BrkDBD dissolved in 20 mM acetate buffer shows only unfolded state resonances at all temperatures, whereas in all other investigated buffers, a broad transition from folded to unfolded BrkDBD is observed. This transition includes severe broadening of NMR resonances at intermediate temperatures, due to conformational exchange close to the intermediate regime. At high temperatures, sharper resonances attributed to the unfolded state ensemble are observed.

^{1}H-^{15}N HSQC spectra of folded apo BrkDBD at various buffer conditions (at 5 °C) are very similar, except for resonances of residues at the N-terminal part of the recognition helix 3. This indicates that these residues show an increased sensitivity towards anion concentrations and type.

NMR ^{1}H-^{15}N HSQC spectra indicate that folded apo BrkDBD and the unspecific DNA complex are very similar, despite some shift differences at the N-terminus and helix 4. In contrast, resonances of residues located in the DNA binding site change significantly from specific to unspecific BrkDBD-DNA complexes.

α-helical propensity in unfolded BrkDBD

^{13}Cα secondary chemical shifts reveal that the α-helical secondary structure propensity is strikingly similar for BrkDBD in complex with either specific or unspecific DNA and the folded apo form. Irrespective of how the folding has been induced, we find that in folded BrkDBD, all four helices are formed by the same residues as in the solution NMR structure of BrkDBD[41].

Even for unfolded BrkDBD, ^{13}Cα secondary chemical shifts indicate α-helical propensity for helices 2, 3 and 4, which are populated to about 20% in the unfolded state ensemble. Remarkably, α-helical propensity is also observed for residues D63-N69 in unfolded BrkDBD. These residues form a loop in the folded state, which connects the C-terminal ends of helices 1 and 2 *via* a network of hydrogen bonds.

Two-state folding of BrkDBD indicated by ^{15}N relaxation dispersion

NMR ^{15}N relaxation dispersion experiments of unfolded BrkDBD are consistent with microsecond to millisecond conformational exchange with the folded state, indicated by the linear correlation of the $|\Delta\omega|$ parameter of most residues and the ^{15}N chemical shift difference between the two states. This indicates that BrkDBD is a two-state folder, which is expected for single domain proteins of this size[9,111,145]. The simultaneous disappearance of many folded state resonances and the appearance of unfolded state resonances of BrkDBD in 20 mM oxalate buffer corroborates the two-state folding mechanism, which has also been suggested for BrkDBD in a recent molecular dynamics study[201].

The few residues (H80-Q85, W87, N93) which could not be fitted to an acceptable level by the two-state approach include those residues located at the N-terminus of helix 3, which showed significant chemical shift variations depending on anions. This may indicate, that these residues experience additional conformational exchange which so far has not been characterized in detail. The same BrkDBD residues could not be detected in the complex with unspecific DNA and locally larger transverse relaxation rates have been observed both in the complex with specific DNA and in 20 mM oxalate buffer. Possibly, these residues show higher dynamics since they probe for specific binding interactions. The detectable BrkDBD resonances in the complex with unspecific DNA have lower intensities than those of the specific DNA complex at comparable experimental conditions. This indicates an increased overall conformational flexibility of BrkDBD in the presence of unspecific DNA, which likely causes the absence of several protein amide resonances in the DNA binding sites of helices 2 and 3. The observation of higher dynamics in the DNA recognition helix 3 compared to other parts of BrkDBD in complex with specific DNA has been attributed to an induced fit mechanism according to all-atom MD simulations[201].

C66 is a crucial residue for folding of BrkDBD

MTSL labeling of C66 of wild type BrkDBD or the C90S mutant inhibited folding at high salt concentrations. Furthermore, the C66S variant of BrkDBD was also unable to fold. This inability to fold and the problem of precipitation at unequal protein and DNA concentrations prevented the detection of transient interactions of BrkDBD with specific and unspecific DNA by PRE[108].

The NMR structure of BrkDBD[41] revealed that C66 is involved in a hydrogen bond network with D63, D65 and K76 which connects the C-terminal parts of helices 1 and 2. Even though the side chain of C66 mutants (e.g. serine or asparagine) could act as hydrogen bond donor and thus participate in the H-bond network, all-atom MD simulations indicate that both the C66S and the C66N BrkDBD mutants loose this network for the majority of the 40 ns simulation time.

Clearly, the hydrogen bond network could not remain intact in MTSL-labeled BrkDBD due to the size of the MTSL moiety. The steric clashes would directly lead to a rearrangement in this region, preventing the formation of the described hydrogen bonds.

CS-Rosetta[221] structure calculations indicate, that the experimental chemical shifts would also be compatible with other conformations of the loop around C66, while the rest of the structure is well-defined. It remains to be proven experimentally, that the region around C66 has to change its conformation from slightly α-helical to a loop during the folding event.

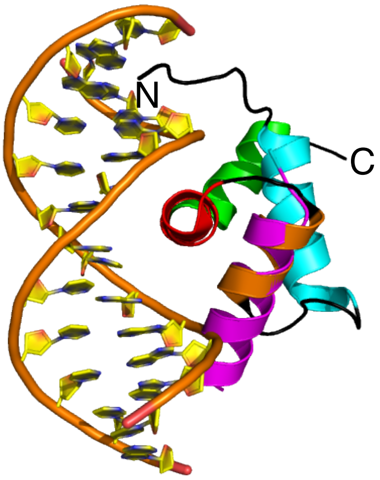


Figure 1.27 Superposition of the pdb structure of BrkDBD[41] and a model of an extended helix 2 (magenta). The native helices are colored cyan (helix 1), orange (helix 2), red (helix 3) and green (helix 4). N and C denote the N- and C-terminal ends of BrkDBD, respectively.

Model of binding and folding of BrkDBD

Altogether, the following model for DNA recognition by BrkDBD may be envisioned (Figure 1.28). First, the disordered apo BrkDBD with helical propensity around helices 2,3 and 4 approaches its target, consistent with a fly-casting model to accelerate the search process for binding partners[223]. Here, the preformed helices could serve as initial contact points with the interaction partner[70]. Once a potential (negatively charged) binding partner is approached, the folded state is stabilized by electrostatic interactions. The N-terminus of the extended helix 2 must unwind, since its N-terminal end would create steric clashes with the DNA (Figure 1.27). Only then helix 1 can be formed which orients the other 3 helices via several hydrophobic interactions as well as the hydrogen bond network around C66. Strong binding (specific interactions) to its partner restrict the conformational degrees of freedom of the protein, such that it does not explore unfolded or partially unfolded conformations. In the absence of a specific binding partner, BrkDBD either partially or completely dissociates and searches for another target molecule or it moves along the DNA and the N-terminal part of helix 3 probes for specific contacts until a high affinity binding site has been recognized (Figure 1.29).

This may be a relevant model for the coupled binding and folding of BrkDBD, where folding can be induced by (unspecific) electrostatic interactions but possibly enables unfolding if no negative charges are in close vicinity. Similar to the observation of sliding and hopping along DNA as well as intersegment transfer between adjacent DNA strands for the HOXD9 homeodomain, without going through the free protein as intermediate[108,110], the search for its target could also be accelerated by this mechanism for BrkDBD. Even if the unfolded free protein would be populated, it

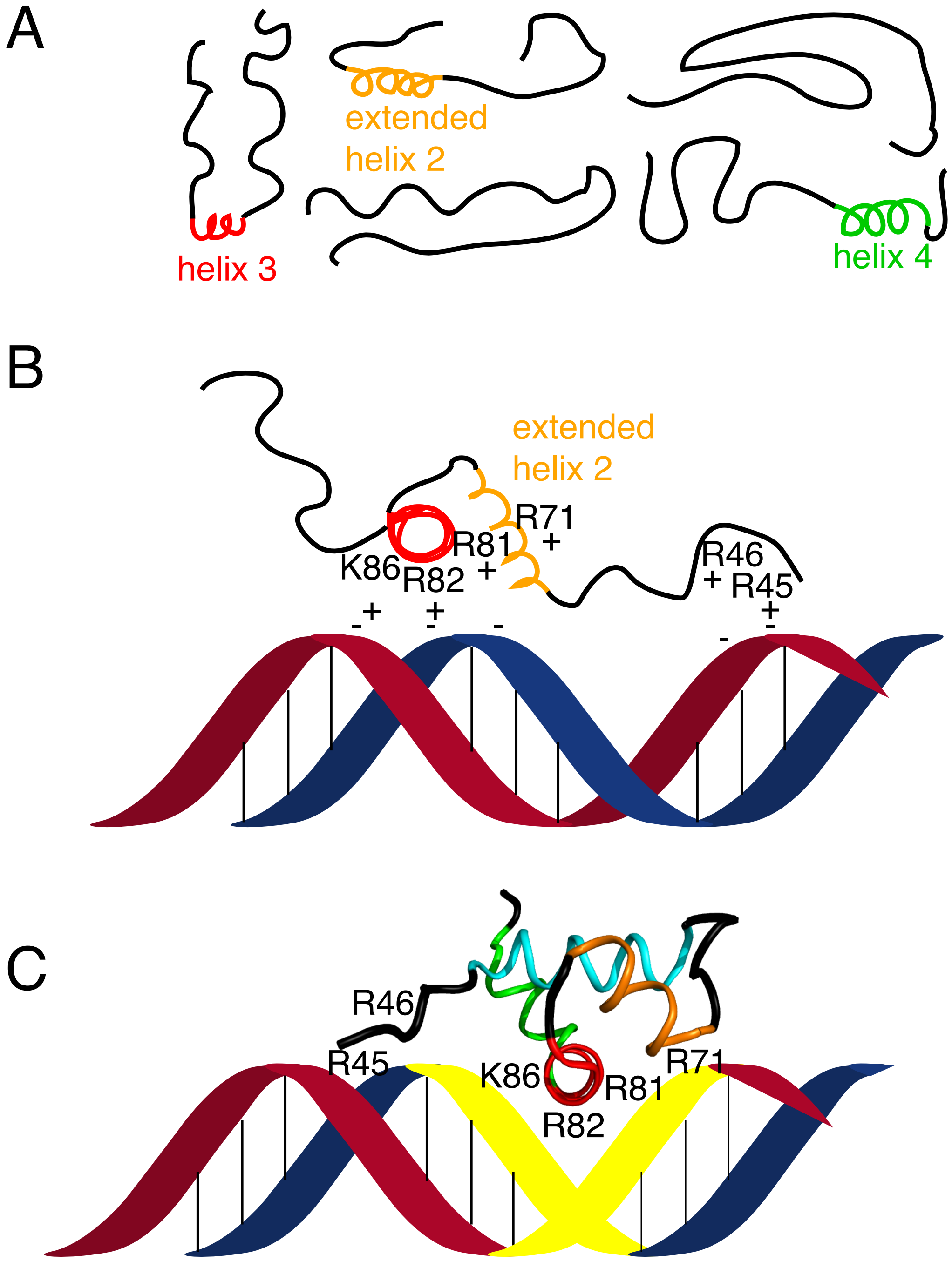


Figure 1.28 Proposed mechanism of coupled binding and folding of BrkDBD upon DNA interactions. **A:** BrkDBD exists as unfolded state ensemble with about 20% propensity for helices 2-4 in the absence of DNA. The extended conformation increases the capture radius. Helices are colored as follows: extended helix 2 (orange), helix 3 (red) and helix 4 (green). **B:** The protein approaches the DNA and folding is induced due to electrostatic interactions. The preformed helices, particularly helix 3 (red) may serve as initial contact points. **C:** BrkDBD is folded and interacts with DNA. The specific DNA motif is shown in yellow.

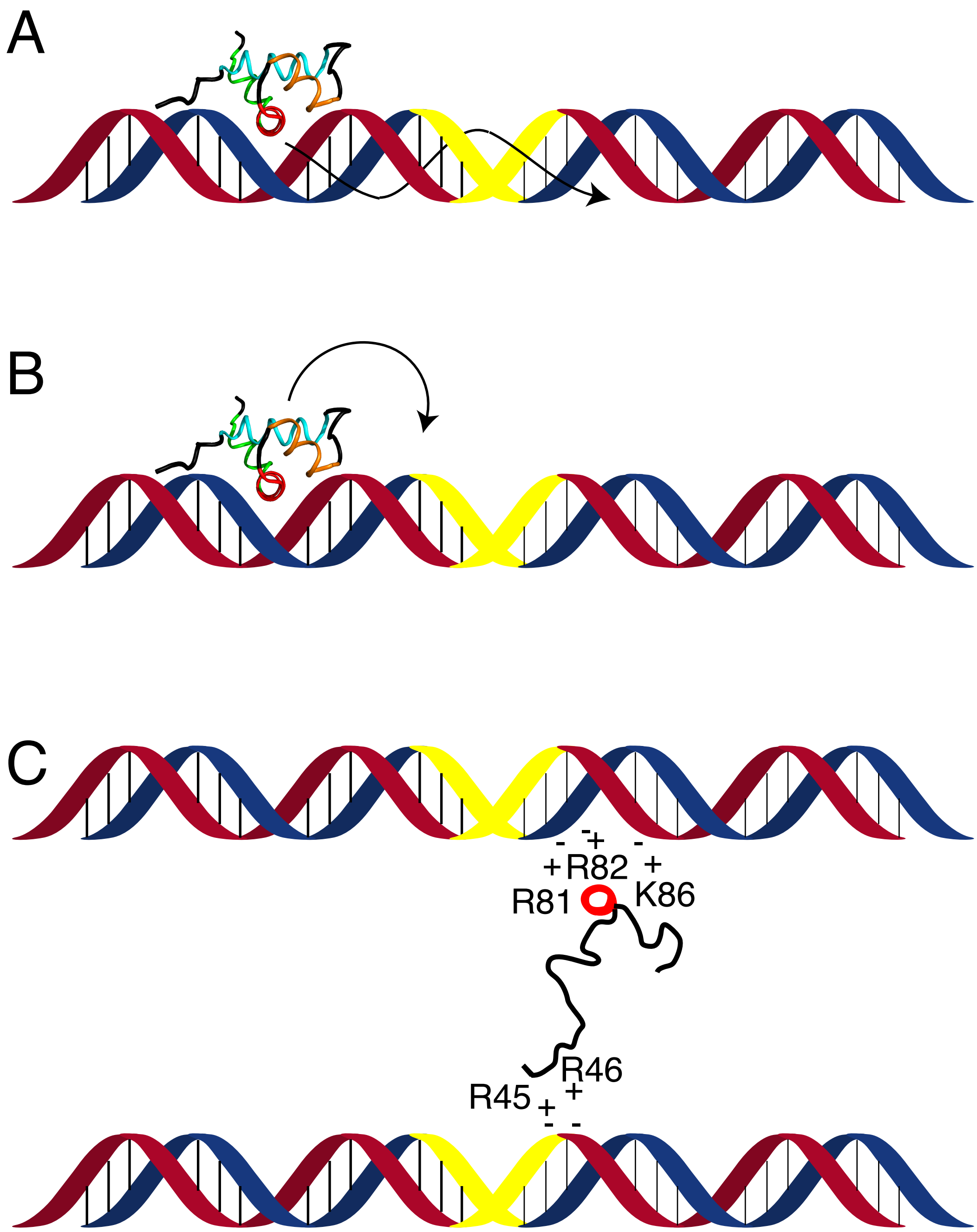


Figure 1.29 Translocation mechanisms of DNA-bound BrkDBD. **A:** Sliding of the protein along the major groove rail. **B:** Hopping along a DNA. **C:** Intermolecular translocation *via* the 'monkey-bar' mechanism[256]. The 3D diffusion mechanism is not shown. Depending on salt conditions, BrkDBD translocates along the DNA strand (sliding and hopping) or changes its binding partner by intersegment transfer. These mechanisms accelerate the search for the specific binding site (yellow). Once the cognate binding site is detected, BrkDBD binds and is immobilized due to the strong specific interactions.

would have an increased capture radius, again accelerating its search.

Clearly, despite the finding that salt 'induces' folding of BrkDBD, it is not possible from the present data to distinguish induced fit from conformational selection. Since preformed helices may represent initial contact points, both mechanisms could contribute to the folding of BrkDBD.

Remarkably, recent computational studies revealed that for homeodomains, all depicted translocation mechanisms (Figure 1.29) take place in a salt-dependent manner[256]. This salt dependence is due to positively charged residues at the N-terminal tail, which is a crucial component of these mechanisms. Therefore, changes in salt concentrations as well as mutations in the N-terminus (particularly mutations of charged residues) alter the preference of individual mechanisms. Our results indicate, that the situation is even more complicated for BrkDBD, since salt does not only modify the strength and the mechanism of interactions with DNA, but also the folding propensity of the protein.

Further investigations

Further investigations should include whether small, highly charged anions are also able to force BrkDBD to fold. The pH dependence of the folding behavior is another topic to be addressed in the future.

It may be of interest to study whether folding of the full-length Brk DNA binding domain (Brk1, residue 1-101) can be induced at high salt concentrations, since only its unfolded apo form has been observed[41]. Similar experiments as for BrkDBD could also be performed to assess the effects of charge density or charge-separation on Brk1. In addition, the long unfolded N-terminus may prevent aggregation, since unfolded BrkDBD was not prone to form aggregates or precipitate, even after exposure to high temperatures. Transient complexes of the Brk1 construct with specific and unspecific DNA could be investigated, if non-equimolar concentration ratios are stable in solution. Since MTSL labeling would probably also prevent folding of Brk1, a radical label could be introduced at the DNA[108]. This may contribute to the understanding of the full domain and its folding and DNA interaction mechanism.

1.2.4 Materials and Methods

Protein preparation
BrkDBD (Brinker residues 43-101) was expressed and purified as described elsewhere[41]. An HPLC purification step was introduced on a 95 ml protein-C4 column (GRACE/Vydac) in a buffer system of buffer A (99.9% water, 0.1% TFA) and buffer B (90.0% acetonitrile, 9.9% water and 0.1% TFA) using a gradient from 20% to 50% buffer B content in 40 minutes at a flow rate of 10 ml/min. BrkDBD elution was typically in the range from 32 to 35 % buffer B. Lyophilized powder was stored at -70 °C and dissolved in water before buffer exchange. Buffer exchange and concentration of the protein as desired for the experiments was achieved with AmiconUltra 3 kDa cutoff concentrators. The concentrations were determined by UV spectroscopy using a molar extinction coefficient of 8480 M^{-1} cm^{-1} at 280 nm[74].
Mutations were performed with the QuickChange (Stratagene) site-directed mutagenesis kit. Desalted primers were obtained from Microsynth.

Preparation of DNA
DNA single strands for BrkDBD-specific omb12T5[41] (5'-TGAGGCGTCAAC-3' and 5'-GTTGACGCCTCA-3') and BrkDBD-unspecific (antennapedia-specific DNA[68,186]: 5'-AGAAAGCCATTAGAG-3' and 5'-TCTCTAATGGCTTTC-3'), were purchased from Microsynth (desalted purity). For annealing, equimolar single strands were dissolved in water, mixed and heated to 95 °C for 15 minutes in a 5 l water bath. The sample was left in the water bath overnight to ensure slow cooling to room temperatures. NMR spectra recorded on the two single strands and on the double strand served as control of the annealing procedure by formation of base pairs. This procedure has been applied to both BrkDBD-specific omb12T5 DNA and BrkDBD-unspecific AHD-DNA.

Circular dichroism
Protein concentrations ranged from 11 to 35 μM (Table 1.1). For each measurement, a 350 μl Hellma quartz cuvette (optical path length: 1 mm) was washed with water (3 times), then with the target buffer (3 times). A full wavelength scan was performed at 5 °C and at 80 °C. An additional full wavelength measurement at 5 °C was recorded after completion of the heating procedure in the presence of BrkDBD to assess the reversibility of the denaturation. The range of the full wavelength scan was 190 to 260 nm. The molar ellipticity at 222 nm was measured at each 5 °C step of the temperature series from 5 to 80 °C. 20 measurements at 222 nm were recorded and averaged for each step. Heating rates of 1 °C/min and 3 min equilibration time were used for each. Averages of 20 measurements were recorded at each temperature point.

In CD, the difference (ellipticity) between absorption of left- and right-handed circularly polarized light is measured. Protein secondary structure elements (α-helices,

β-sheets, polyproline II and random coil loops) show significantly different CD spectra, which can be exploited to assess the content of each[79,231]. For α-helical proteins, the transition temperature from the folded to the unfolded state can be determined by monitoring characteristic wavelengths (a maximum at 192 nm and minima at 208 and 222 nm[231]). The buffers used in this study (Table 1.1) created strong background signals (due to the strong absorption of buffer carbonyl groups) from 190 to 215 nm, obscuring two of the three characteristic α-helix bands. Therefore, the mean residue molar ellipticity at 222 nm was monitored over the temperature range from 5 to 80 °C to determine the transition temperatures from folded to unfolded states. The mean residue molar ellipticity ($\theta_{\text{res}}(\lambda)$) was calculated according to equation 1.4.

$$\theta_{\text{res}}(\lambda) = \Delta\theta(\lambda)/(c \cdot l \cdot n) \tag{1.4}$$

$\Delta\theta(\lambda)$ is the experimental molar ellipticity, c is the concentration in M, l is the pathlength in cm and n is the number of residues.

NMR spectroscopy

Protein concentrations varied from 30 to 400 μM. Resonance assignments were achieved by standard HNCA, HNCO and CBCA(CO)NH[82] experiments using uniformly ^{1}H/^{15}N/^{13}C labeled protein. NMR data were processed with the NMRPipe program suite[47] and resonances were assigned using NMRView[117]. NH-assignments of folded BrkDBD have been achieved for BrkDBD in complex with AHD-DNA (25 °C, 90% assigned), in 20 mM succinate buffer (5 °C, 98%) and in 20 mM oxalate buffer (5 °C, 97%). For unfolded BrkDBD, NH resonances have been assigned in 20 mM acetate buffer (5 °C, 98%; 35 °C, 97%), in 20 mM succinate buffer (5 °C, 98%) and in 20 mM phosphate buffer (50 °C, 84%).

Averaged ^{1}H-^{15}N chemical shift differences

Averaged ^{1}H-^{15}N chemical shift differences were calculated according to the following equation:

$$\Delta\delta_{\text{avg}}[^{1}\text{H} -^{15}\text{N}] = \sqrt{((\Delta\delta^{1}\text{H})^2 + (\Delta\delta^{15}\text{N}/5)^2)/2} \tag{1.5}$$

Secondary chemical shifts

For the calculation of secondary chemical shifts, all random coil shifts were taken from Wishart *et al.*[267], except for ^{15}N, for which sequence-corrected values were used[21].

^{15}N relaxation dispersion experiments

NMR spectra were recorded at 800 and 600 MHz Bruker DRX spectrometers (81.09 and 60.03 MHz for ^{15}N, respectively), equipped with a triple-resonance, Z-gradient TCI cryo-probe and a triple-resonance, triple-gradient TXI probe, respectively. Typically, BrkDBD concentrations of 200 μM in 270 μl buffer (e.g. 20 mM acetate, 5 mM DTT, pH 5.5) in shigemi tubes were used. Spectra were recorded at 5 °C and

an in-house written pulse sequence was used. The total time for relaxation dispersion was 80 ms, filled with various repetitions (4 - 80) of 180 degree pulses, each creating spin-echos. To ensure similar heating effects for different CPMG repeats in 2D sections, a complementary number of pulses (to give in total 84 pulses in this case) within the same time frame was applied directly after the recovery delay and before the initial 90 degree pulse. The following number of arbitrarily ordered pulse repetitions was used for individual 2D sections recorded in an interleaved manner: 4, 4, 8, 12, 16, 20, 24, 24, 32, 40, 40, 48, 56, 60, 60, 64, 72, 80. Experimental errors were calculated from the duplicate data points. Reference intensities were obtained by replacing the part of CPMG repeats by two 180 degree pulses within 200 μs. For the reference experiment, comparable heating effects as for the interleaved measurements was achieved by 82 additional (uniformly spaced) CPMG repeats directly after the recovery delay.
A total of 40 ms relaxation dispersion time was applied for ^{15}N relaxation dispersion experiments on folded BrkDBD in the presence of equimolar amounts of DNA or dissolved in 95 mM succinate buffer. The number of repeated 180 degree pulses for individual 2D sections within 40 ms CPMG time was: 4, 6, 8, 10, 12, 14, 16, 20, 20, 24, 30, 36, 40, 40. NMR data were processed with the NMRPipe program suite[47]. Peak picking of the processed spectra was achieved with PIPP[73]. Relaxation rates were obtained from peak intensities, determined with NLINLS[47].

Fitting of ^{15}N relaxation dispersion curves

Relaxation rates as a function of CPMG frequency were fitted by an in-house written Matlab (MATHWORKS) script by minimization of the following target function (equation 1.6):

$$\chi^2 = \frac{1}{\mathrm{n}} \sum_{i=1}^{t} \left(\frac{(\mathrm{R}_{2,\mathrm{exp},800}(i) - \mathrm{R}_{2,\mathrm{calc},800}(i))^2}{\sigma_{800}(i)^2} + \frac{(\mathrm{R}_{2,\mathrm{exp},600}(i) - \mathrm{R}_{2,\mathrm{calc},600}(i))^2}{\sigma_{600}(i)^2} \right) \quad (1.6)$$

where n is the number of data points minus the number of fit parameters, t is the number of data points of an individual experiment, σ denotes experimental errors (obtained from duplicate data points), with indexes of 600 and 800 for the respective proton frequency of the magnets and $\mathrm{R}_{2,\mathrm{eff}}$ and $\mathrm{R}_{2,\mathrm{calc}}$ denote experimental and calculated effective transverse relaxation rates, defined according to equation 1.3. The exchange rate (k_{ex}), the population of the invisible state (p_{b}), the intrinsic relaxation rate in the absence of exchange ($\mathrm{R}_{2,0}$) and the absolute chemical shift difference between the observed and the invisible state ($|\Delta\omega|$) are obtained for each residue from minimization of the target function (eq. 1.6).

MTSL labeling of BrkDBD variants

A 10- to 20-fold excess of MTSL dissolved in acetonitrile was added to BrkDBD dissolved in the desired NMR buffer, devoid of DTT. For example, 0.2 mg MTSL dissolved in acetonitrile (50 μl) were added to 200 μl of 250 μM BrkDBD dissolved in 100 mM succinate buffer. This amount of MTSL corresponds to a 15-fold excess of the spin label relative to wild type BrkDBD, which contains two cysteines. The reaction mixture was stored in the dark at 4 °C overnight. The success of the reaction was controled by mass spectrometry. Before NMR experiments were carried out, the reaction mixture was washed 5 times with the NMR buffer (filled to 5 ml and concentrated to 250 μl in each step) to remove excess MTSL.

MD simulations

All-atom MD simulations were performed in NAMD (2.7b2[194]) with the CHARMM27 force field[152] in a cubic box with explicit solvent (TIP3P water[118]) and periodic boundary conditions at 298 K. Each simulation had a length of 40 ns with 1 fs timesteps. Snapshots were taken every 100 ps of each trajectory. The following protocol was used: 5000 steps solvent minimization (steepest descent), 300 ps solvent equilibration, 5000 steps full system minimization, 1 ns full system equilibration and 40 ns of free dynamics. This protocol was repeated 10 times for wild type and C66S BrkDBD using different initial velocities.

For each snapshot, the distance between Nζ of K76 and Cγ of D63 was calculated to indicate the presence or the absence of a hydrogen bond. Increasing distances indicated loss of the hydrogen bond network and conformational change of the loop around the crucial residue C66.

CS-Rosetta calculations

CS-Rosetta[221] structure prediction was performed with chemical shifts (^{1}HN, ^{15}N, ^{13}CO, ^{13}Cα, and ^{1}Hα) of apo BrkDBD in 20 mM phosphate buffer. 12500 structures were generated and rmsd values calculated relative to the 20 lowest energy structures.

1.3 Side-Chain χ_1 Conformations in Urea-Denatured Ubiquitin and Protein G from ^3J Coupling Constants and Residual Dipolar Couplings

Original Publication

Vajpai, N., Gentner, M., Huang, J.-r., Blackledge, M., and Grzesiek, S.

Side-Chain χ_1 Conformations in Urea-Denatured Ubiquitin and Protein G from ^3J Coupling Constants and Residual Dipolar Couplings

Journal of the American Chemical Society 132 (2010) 3196-3203.

J|A|C|S
ARTICLES

Published on Web 02/15/2010

Side-Chain χ_1 Conformations in Urea-Denatured Ubiquitin and Protein G from ${}^3J$ Coupling Constants and Residual Dipolar Couplings

Navratna Vajpai,[†] Martin Gentner,[†] Jie-rong Huang,[†] Martin Blackledge,[‡] and Stephan Grzesiek*[,†]

Biozentrum, University of Basel, Klingelbergstrasse 50/70, 4056 Basel, Switzerland, and
Institut de Biologie Structurale Jean-Pierre Ebel, CEA, CNRS, UJF UMR 5075,
41 Rue Jules Horowitz, Grenoble 38027, France

Received December 7, 2009; E-mail: stephan.grzesiek@unibas.ch

Abstract: Current NMR information on side-chain conformations of unfolded protein states is sparse due to the poor dispersion particularly of side-chain proton resonances. We present here optimized schemes for the detection of ${}^3J_{H\alpha H\beta}$, ${}^3J_{NH\beta}$, and ${}^3J_{C'H\beta}$ scalar and ${}^1D_{C\beta H\beta}$ residual dipolar couplings (RDCs) in unfolded proteins. For urea-denatured ubiquitin and protein G, up to six ${}^3J$-couplings to ${}^1H^{\beta}$ are detected, which define the χ_1 angle at very high precision. Interpretation of the ${}^3J$ couplings by a model of mixed staggered χ_1 rotamers yields excellent agreement and also provides stereoassignments for ${}^1H^{\beta}$ methylene protons. For all observed amino acids with the exception of leucine, the chemical shift of ${}^1H^{\beta 3}$ protons was found downfield from ${}^1H^{\beta 2}$. For most residues, the precision of individual χ_1 rotamer populations is better than 2%. The experimental χ_1 rotamer populations are in the vicinity of averages obtained from coil regions in folded protein structures. However, individual variations from these averages of up to 40% are highly significant and indicate sequence- and residue-specific interactions. Particularly strong deviations from the coil average are found for serine and threonine residues, an effect that may be explained by a weakening of side-chain to backbone hydrogen bonds in the urea-denatured state. The measured ${}^1D_{C\beta H\beta}$ RDCs correlate well with predicted RDCs that were calculated from a sterically aligned coil model ensemble and the ${}^3J$-derived χ_1 rotamer populations. This agreement supports the coil model as a good first approximation of the unfolded state. Deviations between measured and predicted values at certain sequence locations indicate that the description of the local backbone conformations can be improved by incorporation of the RDC information. The ease of detection of a large number of highly precise side-chain RDCs opens the possibility for a more rigorous characterization of both side-chain and backbone conformations in unfolded proteins.

Introduction

A detailed, quantitative description of the unfolded state of proteins is crucial for understanding protein folding,[1] protein misfolding and aggregation in amyloidogenic diseases such as Alzheimer's and Parkinson's,[2] and function of intrinsically disordered proteins.[3,4] Such a description is both experimentally and theoretically highly challenging, because only a limited number of measurable parameters are available to describe the vast space of possible unstructured conformations.

During recent years, the conformations of the backbone of unfolded proteins have been described in some detail using paramagnetic relaxation enhancements PREs[5,6] and RDCs.[7,8] These two parameters report on well-defined ensemble averages of the long- and short-range backbone geometry and are thus more amenable to a rigorous quantitative interpretation than chemical shifts, NOE, or relaxation data. Both PREs and RDCs have revealed long-range contacts and residual structure in many denatured proteins showing that such states contain structural bias that may drive them toward a folded structure. The correct prediction of such structural propensities of denatured states from the amino acid sequence may be an important step toward solving the protein folding problem.

RDCs offer particular advantages for the characterization of unfolded states, because they do not require additional chemical labeling and can be detected with ease for a large number of internuclear vectors; for example, a recent study showed that up to seven RDCs per peptide unit could be determined for urea-unfolded ubiquitin.[9] For a number of unfolded proteins, trends of backbone RDCs along the polypeptide sequence could be reproduced in structural ensembles created according to the

---

[†] University of Basel.

[‡] Institut de Biologie Structurale Jean-Pierre Ebel.

(1) Shortle, D. *FASEB J.* **1996**, *10*, 27–34.
(2) Dobson, C. M. *Nature* **2003**, *426*, 884–890.
(3) Dunker, A. K.; Silman, I.; Uversky, V. N.; Sussman, J. L. *Curr. Opin. Struct. Biol.* **2008**, *18*, 756–764.
(4) Wright, P. E.; Dyson, H. J. *Curr. Opin. Struct. Biol.* **2009**, *19*, 31–38.
(5) Gillespie, J. R.; Shortle, D. *J. Mol. Biol.* **1997**, *268*, 170–184.
(6) Mittag, T.; Forman-Kay, J. *Curr. Opin. Struct. Biol.* **2007**, *17*, 3–14.
(7) Shortle, D.; Ackerman, M. *Science* **2001**, *293*, 487–489.
(8) Meier, S.; Blackledge, M.; Grzesiek, S. *J. Chem. Phys.* **2008**, *128*, 052204.
(9) Meier, S.; Grzesiek, S.; Blackledge, M. *J. Am. Chem. Soc.* **2007**, *129*, 9799–9807.

10.1021/ja910331t

amino-acid-specific phi/psi angle propensities in non-alpha, non-beta conformations of PDB structures (PDB coil libraries).[8,10,11] This indicates that the so-called coil model[12,13] is a good, first approximation of the unfolded state ensemble. In turn, deviations from the coil model point to residual order within the unfolded state. Such deviations have revealed highly populated turn conformations in the natively unfolded Tau protein[14] and have shown that urea binding drives the backbone to more extended conformations for ubiquitin.[9] Additional long-range RDCs between amide protons have given evidence for a remaining, significant (10−20%) population of the first β-hairpin (residues 1−18) in (8 M) urea-denatured ubiquitin.[15]

In contrast to the backbone of unstructured polypeptides, experimental information on side chains is rather sparse. Such investigations are severely hampered by the poor dispersion of side-chain signals resulting from the conformational averaging. A small number of $^3J_{\mathrm{H\alpha H\beta}}$ couplings have been determined in shorter unfolded peptides without ^{13}C labeling,[16,17] which correlated with predictions from χ_1 coil distributions. However, no stereoassignments of methylene $\mathrm{H}^{\beta2}$ and $\mathrm{H}^{\beta3}$ resonances were obtained. A more advanced study[18] determined heteronuclear $^3J_{\mathrm{NC\gamma}}$ and $^3J_{\mathrm{C'C\gamma}}$ couplings in urea-denatured lysozyme. Assuming staggered χ_1 rotamers, estimates for their populations were derived for approximately 50 amino acids, which also showed correlations to coil model predictions for most amino acids with the exception of aromatics. Precision in this analysis was limited by the lack of precise Karplus coefficients for $^3J_{\mathrm{NC\gamma}}$ and $^3J_{\mathrm{C'C\gamma}}$ and the fact that the three populations $p_{-60°,60°,180°}$ (two independent parameters because $p_{-60°} + p_{60°} + p_{180°} = 1$) were determined from only two experimental values.

In the present study, we have improved the description of χ_1 conformations in unfolded proteins by optimized heteronuclear experiments involving β-protons, which are able to resolve most methylene $\mathrm{H}^{\beta2}$ and $\mathrm{H}^{\beta3}$ pairs. Stereoassignments and χ_1 angle information could be obtained for the predominant part of residues in urea-denatured ubiquitin and protein G from an extensive set of up to six three-bond scalar couplings ($^3J_{\mathrm{NH\beta2,3}}$, $^3J_{\mathrm{C'H\beta2,3}}$, and $^3J_{\mathrm{H\alpha H\beta2,3}}$). A combined analysis of all ^3J couplings according to the staggered conformer model yields individual populations with a maximal error of 2%. This analysis is corroborated by independent $^1D_{\mathrm{C\beta H\beta2,3}}$ RDC data detected in strained polyacrylamide gels.[19,20] These side-chain RDCs agree well with theoretical RDCs calculated from the ^3J-derived χ_1 conformer distribution and a coil model ensemble of backbone conformations generated by the program *Flexible*-Meccano.[11] The obtained χ_1 conformer populations cluster around coil model averages, but individual variations in particular for serines and threonines of up to 40% are significant and indicate sequence- and residue-specific preferences.

Materials and Methods

Sample Preparation and NMR Spectroscopy. ^{15}N/^{13}C-labeled human ubiquitin and protein G (GB1 sequence ^{1}MQYKLILNGK ^{11}TLKGETTTEA ^{21}VDAATAEKVF ^{31}KQYANDNGVD ^{41}GEWTYDDATK ^{51}TFTVTE) were prepared according to standard protocols.[21] Ubiquitin NMR samples contained 1.0 (0.6) mM ^{15}N/^{13}C-labeled protein in 10 mM glycine, 8 M urea, pH 2.5, 95/5% H_2O/D_2O for measurement under isotropic (anisotropic) conditions. Protein G samples contained 0.6 mM ^{15}N/^{13}C-labeled protein in 10 mM glycine, 7.4 M urea, pH 2.0, 95/5% H_2O/D_2O. Residual alignment of urea-denatured proteins was achieved by introducing the protein solutions into 7% (w/v) polyacrylamide gels and horizontal compression (aspect ratio 2.9:1) in NEW-ERA sample tubes[22] yielding maximal $|^1D_{\mathrm{NH}}|$ RDCs of about 13 Hz for both proteins.

All NMR experiments were carried out at 298 K on a Bruker Avance DRX 800 spectrometer equipped with a TCI cryoprobe. Spectra were processed with NMRPipe[23] and evaluated with NMRView[24] and PIPP.[25]

Assignments. Assignments of urea-denatured ubiquitin (BMRB entry 4375)[26] and protein G[27] were transferred to our sample preparations and extended by a combination of CBCA(CO)NH,[28] HNCO,[29] HBHA(CO)NH,[28] and HNHB[30] experiments. To obtain higher resolution, the constant time ^{15}N acquisition period was increased in these experiments to about 40 ms. Note that for CBCA(CO)NH, HNCO, and HBHA(CO)NH, this still achieves a transfer of about 95% via the $^1J_{\mathrm{NC'}}$ (∼15 Hz) coupling. Almost complete assignments of all $^1\mathrm{H}^{\mathrm{N}}$, ^{15}N, ^{13}C′, $^{13}\mathrm{C}^{\alpha}$, $^{13}\mathrm{C}^{\beta}$, $^1\mathrm{H}^{\alpha}$, and $^1\mathrm{H}^{\beta}$ resonances were obtained from this procedure. Missing assignments mainly comprise amino acids preceding proline or were due to signal degeneracy of some geminal protons. The obtained chemical shifts are close to the published data with the exception of residues in the vicinity of the mutated T2Q site in protein G. They also extend the previous data by the stereoassignments of β-methylene protons and the ^{13}C′ chemical shifts (protein G). The assignments are deposited in the BMRB data bank under accession numbers 16626 (ubiquitin) and 16627 (protein G).

Determination of Scalar and Residual Dipolar Coupling Constants. ^3J scalar couplings carrying information on the χ_1 angle of the denatured proteins were obtained from modified versions of quantitative $^3J_{\mathrm{NH\beta}}$-HNHB,[30] $^3J_{\mathrm{C'H\beta}}$-HN(CO)HB,[31] and $^3J_{\mathrm{H\alpha H\beta}}$-HAHB-(CACO)NH[32] experiments. $^1D_{\mathrm{C\beta H\beta}}$ RDCs for χ_1 angle information were determined from an HBHA(CO)NH[28] experiment, where an IPAP detection scheme[33] was introduced into the mixed constant

---

[10] Jha, A.; Colubri, A.; Freed, K.; Sosnick, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13099–13104.

[11] Bernado, P.; Blanchard, L.; Timmins, P.; Marion, D.; Ruigrok, R.; Blackledge, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17002–17007.

[12] Serrano, L. *J. Mol. Biol.* **1995**, *254*, 322–333.

[13] Smith, L.; Bolin, K.; Schwalbe, H.; MacArthur, M.; Thornton, J.; Dobson, C. *J. Mol. Biol.* **1996**, *255*, 494–506.

[14] Mukrasch, M.; Markwick, P.; Biernat, J.; Bergen, M.; Bernado, P.; Griesinger, C.; Mandelkow, E.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* **2007**, *129*, 5235–5243.

[15] Meier, S.; Strohmeier, M.; Blackledge, M.; Grzesiek, S. *J. Am. Chem. Soc.* **2007**, *129*, 754–755.

[16] West, N. J.; Smith, L. J. *J. Mol. Biol.* **1998**, *280*, 867–877.

[17] Mathieson, S. I.; Penkett, C. J.; Smith, L. J. *Pacific Symp. Biocomputing* **1999**, 542–553.

[18] Hennig, M.; Bermel, W.; Spencer, A.; Dobson, C. M.; Smith, L. J.; Schwalbe, H. *J. Mol. Biol.* **1999**, *288*, 705–723.

[19] Sass, H.; Musco, G.; Stahl, S.; Wingfield, P.; Grzesiek, S. *J. Biomol. NMR* **2000**, *18*, 303–309.

[20] Tycko, R.; Blanco, F.; Ishii, Y. *J. Am. Chem. Soc.* **2000**, *122*, 9340–9341.

[21] Sass, J.; Cordier, F.; Hoffmann, A.; Cousin, A.; Omichinski, J.; Lowen, H.; Grzesiek, S. *J. Am. Chem. Soc.* **1999**, *121*, 2047–2055.

[22] Chou, J.; Gaemers, S.; Howder, B.; Louis, J.; Bax, A. *J. Biomol. NMR* **2001**, *21*, 377–382.

[23] Delaglio, F.; Grzesiek, S.; Vuister, G.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277–293.

[24] Johnson, B.; Blevins, R. *J. Biomol. NMR* **1994**, *4*, 603–614.

[25] Garrett, D.; Powers, R.; Gronenborn, A.; Clore, G. *J. Magn. Reson.* **1991**, *95*, 214–220.

[26] Peti, W.; Smith, L.; Redfield, C.; Schwalbe, H. *J. Biomol. NMR* **2001**, *19*, 153–165.

[27] Frank, M. K.; Clore, G. M.; Gronenborn, A. M. *Protein Sci.* **1995**, *4*, 2605–2615.

[28] Grzesiek, S.; Bax, A. *J. Biomol. NMR* **1993**, *3*, 185–204.

[29] Grzesiek, S.; Bax, A. *J. Magn. Reson.* **1992**, *96*, 432–440.

[30] Archer, S.; Ikura, M.; Torchia, D.; Bax, A. *J. Magn. Reson.* **1991**, *95*, 636–641.

[31] Grzesiek, S.; Ikura, M.; Clore, G.; Gronenborn, A.; Bax, A. *J. Magn. Reson.* **1992**, *96*, 215–221.

[32] Lohr, F.; Schmidt, J.; Ruterjans, H. *J. Am. Chem. Soc.* **1999**, *121*, 11821–11826.

[33] Ottiger, M.; Delaglio, F.; Bax, A. *J. Magn. Reson.* **1998**, *131*, 373–378.

time $^1H^{\alpha/\beta}$ evolution/$^1H^{\alpha/\beta} \rightarrow {}^{13}C^{\alpha/\beta}$ transfer period. $^1D_{C\beta H\beta}$ RDCs were calculated as the difference in couplings observed under anisotropic conditions in strained polyacrylamide gels and a second experiment under isotropic conditions. Details of these experiments are given in Supporting Information Figures S1−S4. Similar to the assignment experiments, acquisition times in the indirect ^{15}N and $^1H^{\alpha/\beta}$ dimensions were set to 40 and 30 ms, respectively, to obtain sufficient resolution. Each experiment for scalar and dipolar couplings was carried out twice, and the reported coupling constants and error estimates refer to mean and standard deviations from such repeated experiments. The *J*-coupling constants were not corrected for effects of scalar relaxation of the second kind,[34] because these are expected to be small due to the fast effective correlation times in unfolded proteins.

Analysis of ^3J Coupling Constants and RDCs. Analysis of the ^3J data according to a model of staggered χ_1 rotamers (see main text) was carried out using in-house written Matlab (MathWorks, Inc.) routines and its QUADPROG function for constrained linear minimization. For the analysis of RDC data, theoretical RDCs from steric alignment were calculated for the three staggered rotamers by an in-house written C program[35] from an ensemble of 50 000 unfolded protein structures generated by the *Flexible*-Meccano program.[11] Experimental RDCs were then compared and fitted to these predicted values by the same staggered rotamer model (see main text).

Coil Library Rotamers. The experimentally derived χ_1 rotamer populations were also compared to average populations from a protein coil library. This library was downloaded as version 20080310_pc20_res1.6_R0.25 generated on 3/10/2008 from the Rose lab server[36] and contained 16 856 protein fragments from nonhomologous proteins of X-ray structures with resolution better than 1.6 Å in non-alpha, non-beta conformations.

Results and Discussion

χ_1 Torsion Angle Information from Scalar Couplings and RDCs. The NMR analysis of unfolded proteins in solution is made difficult by the low spectral dispersion resulting from conformational averaging. In particular, this applies to side-chain resonances. Thus, so far, the potential of side-chain protons has not fully been used as a source of conformational information. Here, we have probed the χ_1 torsion angles in urea-denatured ubiquitin and protein G by ^3J scalar and residual dipolar couplings involving H^β protons. Scalar couplings were detected by modified versions of the quantitative-$^3J_{NH\beta}$ HNHB,[30] $^3J_{C'H\beta}$ HN(CO)HB,[31] and $^3J_{H\alpha H\beta}$ HAHB(CACO)NH[32] experiments (Supporting Information). These experiments overcome the problems of low dispersion in both backbone and side-chain resonances by making use of the long transverse relaxation times in the unfolded state to achieve maximal frequency resolution. Acquisition times of 30−35 ms for $^1H^\alpha$ and $^1H^\beta$ resonances and of 40 ms for ^{15}N in the indirect dimensions proved to be sufficient to resolve most of the overlap in the side-chain experiments. Additional information from similarly optimized 3D CBCA(CO)NH, HBHA(CO)NH experiments was used to establish sequential assignments. Figure 1 shows the good resolution of the side-chain $^1H^\beta$ resonances in the quantitative-*J* spectra for residues L7 and K10 of urea-denatured protein G. The intensity ratios of the resonances in these spectra yield information on the χ_1 torsion angles and the stereospecific assignments for *β*-methylene protons. The high population of −60° χ_1 conformations is obvious from the higher intensities of the $^1H^{\beta 2}$ resonances relative to the $^1H^{\beta 3}$ resonances in the quantitative-$^3J_{H\alpha H\beta}$ HAHB(CACO)NH experiment, which corresponds to $^3J_{H\alpha H\beta 2} > {}^3J_{H\alpha H\beta 3}$, from the inverse situation in the quantitative-$^3J_{NH\beta}$ HNHB experiment, and from more equal intensities in the quantitative-$^3J_{C'H\beta}$ HN(CO)HB experiment.

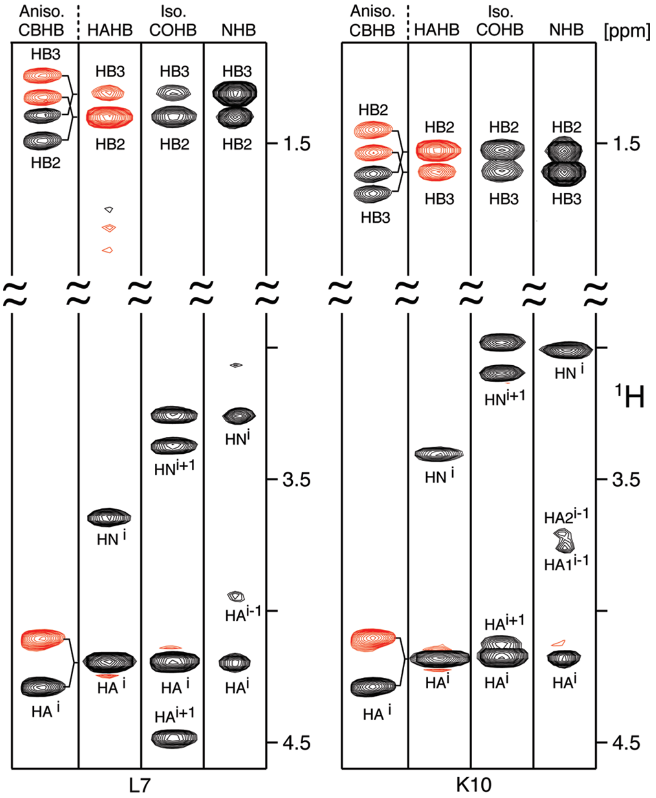


Figure 1. Strip plots of 3D spectra used for determination of $^3J_{H\alpha H\beta}$, $^3J_{NH\beta}$, $^3J_{C'H\beta}$ coupling constants and $^1D_{C\beta H\beta}$ RDCs in urea-denatured protein G. Data are shown for two typical residues L7 (left) and K10 (right) from the quantitative $^3J_{H\alpha H\beta}$-HAHB(CACO)NH (HAHB), $^3J_{C'H\beta}$-HN(CO)HB (COHB), and $^3J_{NH\beta}$-HNHB (NHB) experiments recorded under isotropic conditions as well as from the $^1D_{C\beta H\beta}$-IPAP-HBHA(CO)NH (CBHB) experiment recorded under anisotropic conditions. Resonances are labeled with assignment information. Amide proton signals are folded in the indirect proton dimension to reduce experimental time. These signals are split by the $^1J_{NH}$ coupling for the HN(CO)HB experiment. For the quantitative-*J* HAHB, COHB, and NHB experiments, resonances shown in red are negative signals. For the IPAP-HBHA(CO)NH experiment, red is used to distinguish the $^1H^{\alpha/\beta}$ upfield from the downfield (black) components.

In addition to ^3J couplings, also RDCs yield information on the χ_1 conformations. The detection of $^1D_{C\beta H\beta}$ RDCs induced in strained polyacrylamide gels proved particularly easy from an IPAP-HBHA(CO)NH experiment (Supporting Information). Figure 1 shows as an example the data of this experiment on the oriented urea-denatured protein G. The quality of the spectra is excellent and allowed the unambiguous determination of 91 (61) $^1D_{C\beta H\beta}$ RDCs in unfolded ubiquitin (protein G).

In total for ubiquitin (protein G), 353 (246) $^3J_{NH\beta}$, $^3J_{C'H\beta}$, $^3J_{H\alpha H\beta}$, and $^1D_{C\beta H\beta}$ couplings (Supporting Information Table S1) could be derived from the quantitative analysis of the spectra. These data cover 82% of all side chains with variable χ_1 angle, that is, 55 out of 68 (ubiquitin) and 40 out of 46 (protein G) non-(Gly, Ala) residues.

Analysis of ^3J Coupling Constants by Staggered χ_1 Rotamer Populations. Our analysis of the ^3J couplings in terms of side-chain conformations assumes as a first approximation that the conformations are a population mixture of three staggered χ_1 (−60°, +60°, 180°) rotamers. Previously deter-

---

(34) Bax, A.; Vuister, G. W.; Grzesiek, S.; Delaglio, F.; Wang, A. C.; Tschudin, R.; Zhu, G. *Methods Enzymol.* **1994**, *239*, 79–105.
(35) Huang, J.-r.; Grzesiek, S. *J. Am. Chem. Soc.* **2010**, *132*, 694−705.
(36) Fitzkee, N.; Fleming, P.; Rose, G. *Proteins* **2005**, *58*, 852–854.

mined amino-acid-specific Karplus coefficients[37] were used to predict theoretical $^3J_{\mathrm{H\alpha H\beta i}}$, $^3J_{\mathrm{NH\beta i}}$, and $^3J_{\mathrm{C'H\beta i}}$ coupling constants for these rotamers according to the Karplus relation:

$$^3J_{ij}^{\mathrm{calc}}(\chi_1) = C_{ij}^0 + C_{ij}^1 \cos(\vartheta_{ij}(\chi_1)) + C_{ij}^2 \cos(2\vartheta_{ij}(\chi_1)) \qquad (1)$$

where $\vartheta_{ij}(\chi_1)$ is the intervening dihedral angle between the nuclei i and j in a $^3J_{ij}$ coupling for a specific side-chain torsion angle χ_1. For all side chains, only single sets of resonances were observed. Thus, the side chains are in fast exchange on the time scale of the chemical shift, that is, faster than milliseconds. Accordingly, the observed coupling constants should be population averages over the individual rotamers:[38−40]

$$\langle {}^3J_{ij}^{\mathrm{calc}} \rangle = \sum_{\chi_1 = -60°, 60°, 180°} p_{\chi_1}\, {}^3J_{ij}^{\mathrm{calc}}(\chi_1) \qquad (2)$$

where $p_{-60°,60°,180°}$ are the individual populations.

To derive these populations from the experimental $^3J^{\mathrm{exp}}$ couplings, their deviation from the calculated average $\langle {}^3J^{\mathrm{calc}} \rangle$ was minimized with respect to $p_{-60°}$, $p_{60°}$, and $p_{180°}$ by a constrained linear least-squares fit of the target function

$$\chi^2 = \frac{1}{N} \sum_{ij} \left(\frac{{}^3J_{ij}^{\mathrm{exp}} - \langle {}^3J_{ij}^{\mathrm{calc}} \rangle}{\sigma_{ij}} \right)^2 \qquad (3)$$

under the conditions $p_{-60°}+p_{60°}+p_{180°} = 1$ and $0 \leq p_{-60°,60°,180°} \leq 1$. In eq 3, σ presents the statistical, experimental error of the coupling constant obtained from a repetition of the experiment, the summation runs over all individual nuclei i and j, for which a ^3J coupling could be determined for an individual side chain, and N indicates the total number of measured ^3J values.

The stereospecific assignments of geminal $\mathrm{H}^{\beta 2}$ and $\mathrm{H}^{\beta 3}$ protons in urea-denatured ubiquitin and protein G were not known prior to the current analysis. This information was also derived from the ^3J couplings by carrying out the fit procedure for both possible stereo assignments and using the assignment that corresponded to the lower χ^2 value. Typically these χ^2 values were about 10−20 times smaller than the values for the swapped assignment, such that discrimination was achieved easily. The stereospecific assignments obtained by this method comprise 82% of all β-methylene protons, corresponding to 41 and 25 residues in ubiquitin and protein G, respectively.

Figure 2 shows the experimental $^3J_{\mathrm{NH\beta}}$, $^3J_{\mathrm{C'H\beta}}$, and $^3J_{\mathrm{H\alpha H\beta}}$ coupling constants and their values according to the fit of eq 3 for a number of residues in ubqitin and protein G (the complete data for both proteins are shown in Supporting Information Figures S5 and S6). For most residues, the agreement between experimental and predicted data is excellent with average RMSDs between measured and predicted ^3J-values of less than 0.3 Hz. This indicates not only a high precision of both experimental ^3J data and Karplus coefficients, but also validates the staggered rotamer model as a reasonable approximation for the side-chain conformations. In total, 53 (39) χ_1 rotamer populations could be derived for those amino acids in ubiquitin (protein G) for which at least two ^3J coupling constants had been measured (Supporting Information Table S2). The constrained linear fit also provides error estimates for the $p_{-60°}$, $p_{60°}$, and $p_{180°}$ populations derived by propagation from the statistical experimental error. These errors range between 0.01 and 0.02, indicating a very high precision of the population estimates.

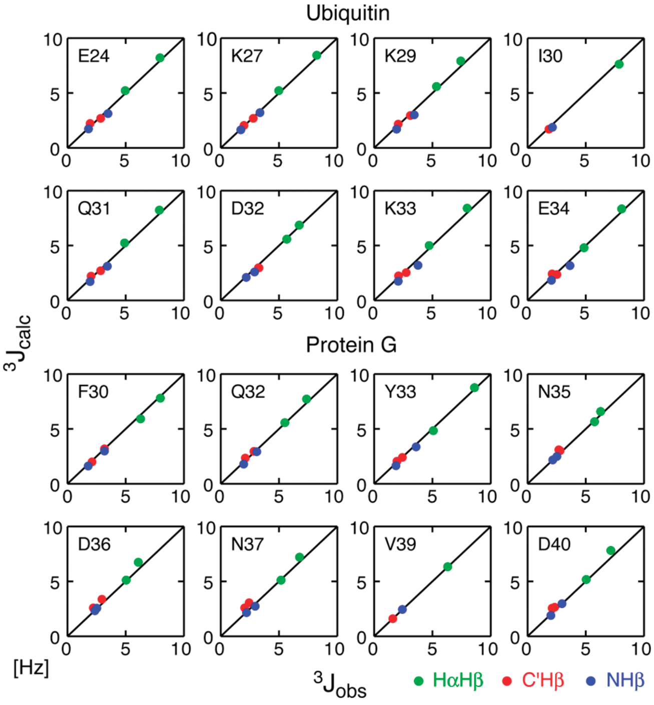


Figure 2. Comparison of experimental (obs) $^3J_{\mathrm{XH\beta}}$ (X = C′, N, HA) constants to values derived (calc) according to the χ_1 rotamer population fit of eq 3 for selected residues of urea-denatured ubiquitin (top) and protein G (bottom).

Analysis of RDC Data. To make use of the experimental $^1D_{\mathrm{C\beta H\beta}}$ RDCs for the analysis of χ_1 conformations, theoretical RDC estimates for all staggered rotamers were obtained from large simulated ensembles of unfolded ubiquitin or protein G structures. For both proteins, ensembles of 50 000 unfolded structures were generated by the *Flexible*-Meccano program[11] according to the amino-acid-specific phi/psi angle propensities in non-alpha, non-beta conformations of PDB structures (PDB coil library) and omitting structures with sterical clashes. Such ensembles have previously been shown to reproduce the trends of backbone RDCs along the polypeptide sequence.[8,10,11] Because the *Flexible*-Meccano algorithm represents side chains only by a pseudoatom at the C^β position, full coordinates for the C^β and H^β atoms for all staggered rotamers were generated from the N, C^α, and C′ positions using idealized tetrahedral geometry. The alignment tensor for each member k of the ensemble was then calculated on the basis of the assumption of steric exclusion[41,42] using an efficient in-house written algorithm.[35] In brief, this algorithm calculates the maximal extension of the molecule for each direction of the unit sphere. The probability for finding the molecule in a certain orientation is then derived as the volume that can be occupied by the molecule between two infinitely extended, parallel planes relative to the total volume between the planes. The alignment tensor then corresponds to the average over all orientations of

---

(37) Pérez, C.; Löhr, F.; Rüterjans, H.; Schmidt, J. M. *J. Am. Chem. Soc.* **2001**, *123*, 7081–7093.

(38) Pachler, A. *Spectrochim. Acta* **1963**, *19*, 2085–2092.

(39) Pachler, A. *Spectrochim. Acta* **1964**, *20*, 581–587.

(40) Dzakula, Z.; Westler, W.; Edison, A.; Markley, J. L. *J. Am. Chem. Soc.* **1992**, *114*, 6195–6199.

(41) Zweckstetter, M.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 3791–3792.

(42) van Lune, F.; Manning, L.; Dijkstra, K.; Berendsen, H. J.; Scheek, R. M. *J. Biomol. NMR* **2002**, *23*, 169–179.

second rank spherical harmonics weighted by this probability. Theoretical RDC values were derived for each χ_1 rotamer and in each individual structure as

$$D^{\text{calc}}_{k,ij}(\chi_1) = -\frac{\gamma_i\gamma_j\hbar\mu_0}{4\pi^2}\sqrt{\frac{4\pi}{5}}\sum_{m=-2}^{2} S^*_{k,m}\frac{Y_{2m}(\Theta_{k,ij}(\chi_1), \Phi_{k,ij}(\chi_1))}{r^3_{k,ij}(\chi_1)} \tag{4}$$

where $D^{\text{calc}}_{k,ij}$ represents the RDC between nuclei i and j for ensemble member k with individual alignment tensor $S_{k,m}$ (written in irreducible form[43]), Y_{2m} are spherical harmonics, $r_{k,ij}$, $\Theta_{k,ij}$, $\Phi_{k,ij}$ are the polar coordinates of the internuclear vector, and $\gamma_{i,j}$ are the nuclear gyromagnetic ratios. These RDC values for the individual structures were then averaged over all N members of the ensemble to obtain an estimate D^{calc}_{ij} for the RDC in the unfolded protein:

$$D^{\text{calc}}_{ij}(\chi_1) = \frac{1}{N}\sum_{k=1}^{N} D^{\text{calc}}_{k,ij}(\chi_1) \tag{5}$$

Because the absolute size of the alignment tensor $S_{k,m}$ is difficult to predict from the experimental conditions, an additional common overall scaling was used such that the mean square difference between measured and predicted average backbone ${}^1D_{\text{HN}}$ was minimized. Analogously to eq 2, the population average over the individual χ_1 rotamers was then calculated as

$$\langle D^{\text{calc}}_{ij}\rangle = \sum_{\chi_1=-60°,60°,180°} p_{\chi_1} D^{\text{calc}}_{ij}(\chi_1) \tag{6}$$

Figure 3 shows the experimental ${}^1D_{C\beta H\beta}$ RDCs and their values predicted by eq 6 from the ${}^3J$-derived χ_1 rotamer populations (red ●) for the same residues in ubiquitin and protein G as in Figure 2. For comparison, measured and predicted ${}^3J$-couplings (blue) are also shown. The complete data for both proteins are given in Supporting Information Figures S7 and S8. For many residues, such as E24, K29, I30, Q31, and E34 in ubiquitin and F30, Q32, N35, N37, V39, and D40 in protein G, measured and predicted RDCs agree within about 5 Hz. This agreement is very reasonable when compared to the full variation of about 20−30 Hz of observed RDCs. For other residues like D32 in ubiquitin or D36 in protein G, the deviations from the predictions are clearly larger, but still agree with the trends of the predictions. The correlations between all measured and predicted ${}^1D_{C\beta H\beta}$ RDCs have a Pearson's correlation coefficient of 0.86 and 0.70 for ubiquitin and protein G, respectively (Figure 4, ●). The correlation can be improved to some extent, when the ${}^1D_{C\beta H\beta}$ RDCs are also included into the fit of the χ_1 rotamer populations by extending the χ^2 function of eq 3 to the differences between measured and predicted D_{ij} couplings (Figures 3 and 4, ○). In this case, the χ_1 rotamer populations only change by a few percent (not shown), but the correlation coefficient increases to 0.92 (0.76) for ubiquitin (protein G).

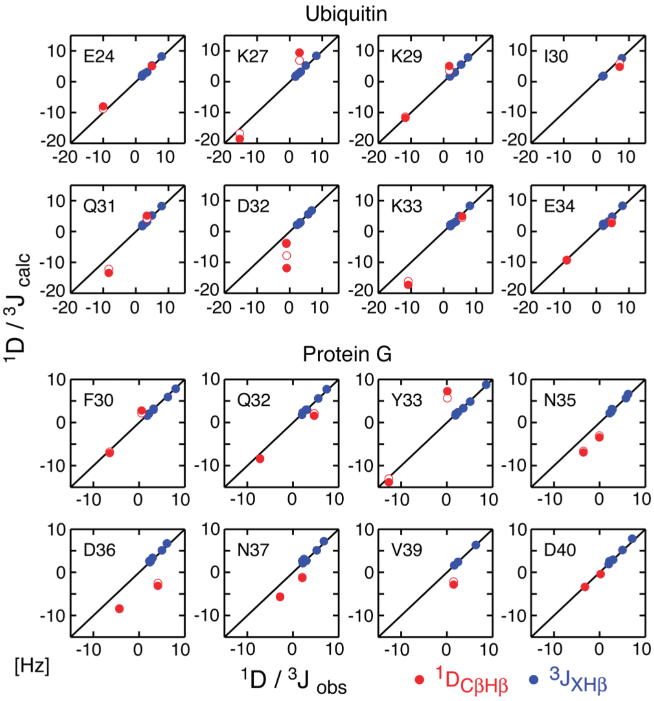


Figure 3. Comparison of experimental (obs) and predicted (calc) ${}^1D_{C\beta H\beta}$ RDCs (red ●) based on the χ_1 rotamer populations derived from the fit of eq 3 of ${}^3J_{XH\beta}$ constants and the coil model ensemble. For comparison, experimental and predicted ${}^3J_{XH\beta}$ couplings are also indicated (blue ●). The same residues are shown as in Figure 2 for urea-denatured ubiquitin (top) and protein G (bottom). Red ○ indicate predictions for ${}^1D_{C\beta H\beta}$ RDCs, when these RDC data were also included in the rotamer population fit of eq 3 (see text).

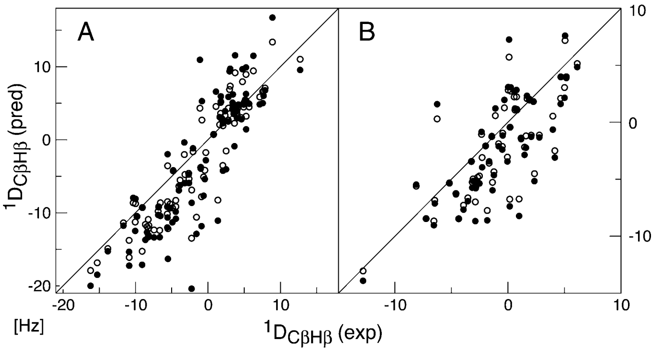


Figure 4. Correlation of all observed experimental (obs) and predicted (calc) ${}^1D_{C\beta H\beta}$ RDCs (●) based on the χ_1 rotamer populations derived from the fit of eq 3 of ${}^3J_{XH\beta}$ constants and the coil model ensemble for urea-denatured ubiquitin (A) and protein G (B). The Pearson's correlation coefficient is 0.86 and 0.70 for ubiquitin and protein G, respectively. The "○" indicate predictions for ${}^1D_{C\beta H\beta}$ RDCs, when these RDC data were also included in the rotamer population fit of eq 3. In this case, the correlation coefficient increases to 0.92 and 0.76 for ubiquitin and protein G, respectively.

Considering the crudeness of the assumptions for the coil model and the steric alignment of the unfolded model ensemble, the agreement between measured and predicted ${}^1D_{C\beta H\beta}$ RDCs is surprisingly good. This provides an independent confirmation for both the backbone coil model as implemented by the *Flexible*-Meccano algorithm[11] as well as the staggered χ_1 rotamer model. However, the deviations, which exceed the experimental errors, also clearly indicate shortcomings of this interpretation. Because the agreement of the staggered rotamer predictions is almost perfect for the ${}^3J$-couplings, it is very likely that the RDC deviations result from inaccuracies of the local backbone geometry predicted by the coil ensemble and from the unknown microscopic details of the alignment interaction, which may not be adequately covered by the simple steric alignment model. In principle, the inaccuracies of the backbone geometry may be reduced by refining the backbone conformations using information from additional backbone RDCs. Such

---

[43] Moltke, S.; Grzesiek, S. *J. Biomol. NMR* **1999**, *15*, 77–82.

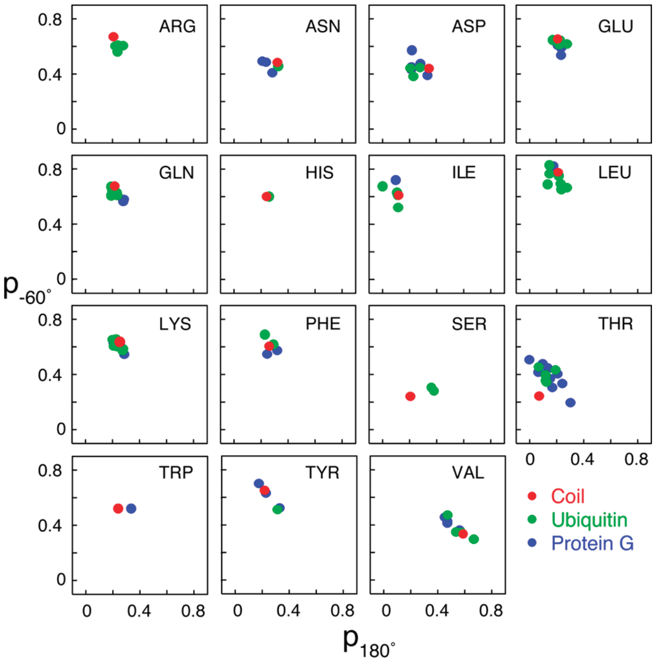


Figure 5. $p_{-60°}$ and $p_{180°}$ χ_1 rotamer populations obtained from the combined fit of ${}^3J_{NH\beta}$, ${}^3J_{C'H\beta}$, and ${}^3J_{H\alpha H\beta}$ couplings of eq 3. Populations are shown separately for ubiquitin (green), protein G (blue), and the average of the PDB coil structures (red).

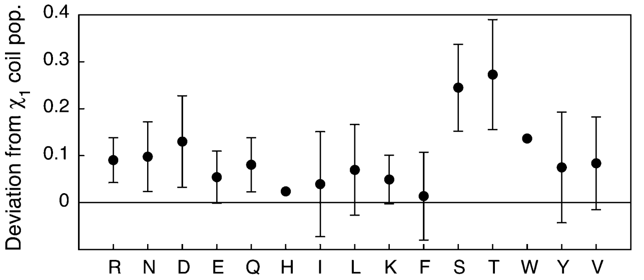


Figure 6. Deviations of the experimentally (${}^3J$-only) derived χ_1 rotamer populations in urea-denatured proteins from the average of the PDB coil structures. The deviations are calculated as $\|\vec{p}_{exp} - \vec{p}_{coil}\|$ with $\vec{p} = (p_{-60°,60°,180°})$. Average and standard deviations of $\|\vec{p}_{exp} - \vec{p}_{coil}\|$ are shown for all amino acids for which rotamer populations could be obtained in urea-denatured ubiquitin and protein G.

an approach is currently pursued by constrained molecular dynamics ensemble calculations, which include all of the RDCs as restraints.[35] The ease of detection of a large number of highly precise RDCs on side-chain nuclei may thus make it possible to increase the accuracy of the description of side-chain and backbone conformations in unfolded ensembles beyond the simple coil model.

Comparison to Coil χ_1 Populations. Previous analyses of the χ_1 rotamer conformations in unfolded proteins using more limited ${}^3J$ data[16−18] have concluded that their populations correlate to the average rotamer populations in the PDB coil conformations. Figure 5 shows the $p_{-60°}$ and $p_{180°}$ χ_1 rotamer populations according to the combined fit of ${}^3J_{NH\beta}$, ${}^3J_{C'H\beta}$, and ${}^3J_{H\alpha H\beta}$ couplings (Figure 2) together with the respective coil averages for all amino acids in ubiquitin and protein G, for which populations could be derived. The populations are indeed in the vicinity of the coil values. However, in some cases, they deviate by more than 30% from the coil values (Figure 6) and also vary around their mean by approximately 10%. These variations are significant considering that the errors of the individual populations are only about 1−2% as estimated from error propagation of the linear least-squares fit. Thus, they must reflect sequence-specific preferences of the side chains along the unfolded polypeptide chain (see below).

As remarked earlier,[18] many amino acids prefer the −60° χ_1 rotamer both in the coil model predictions and in ${}^3J$-derived populations due to the repulsion of substituents at the γ^1 position from the main chain. In our analysis (Figure 5), this is the case for all unbranched amino acids (R, N, D, E, Q, H, L, K, F, W, Y) as well as for the branched isoleucine. These residues have J-derived and coil average $p_{-60°}$ values of 40−80%, where the shortest side-chain asparagine and aspartic acid residues have the lowest and leucine the highest −60° preference. Valines have similar populations for the −60° (30−50%) and +180° (40−70%) χ_1 rotamers, which on average places both γ-methyl groups at the furthest distance from the backbone carbonyl.

Exceptions to this behavior are found for serines and threonines. For these residues, the coil model predicts $p_{-60°}$ populations of only 24% (S, T) and much larger $p_{60°}$ populations of 55% (S) and 68% (T). It has been speculated[18] that this is caused by favorable polar interactions between the side-chain hydroxyl and the main-chain amide group in the +60° χ_1 rotamer. Among all amino acid types, the experimentally derived rotamer populations for serines and threonines in urea-denatured ubiquitin and protein G show the strongest deviations from these coil predictions of folded protein structures. Thus, the J-derived $p_{60°}$ populations amount to only about 35% for serine and 40−50% for threonine, whereas $p_{-60°}$ and $p_{+180°}$ are correspondingly higher (Figure 5). The deviations are found for all serine and threonine residues, and no particular correlation to specific locations in the sequence is evident (Supporting Information Figure S9).

It is unlikely that this behavior is an experimental artifact connected to the particular ${}^{13}C^{\beta}$ chemical shift of these two amino acids, because the ${}^{13}C^{\beta}$ nuclei are not involved in the magnetization pathways of the quantitative ${}^3J_{NH\beta}$-HNHB, ${}^3J_{C'H\beta}$-HN(CO)HB, and ${}^3J_{H\alpha H\beta}$-HAHB(CACO)NH experiments. To further test for systematic errors from the ${}^3J_{H\alpha H\beta}$-HAHB(CA-CO)NH experiment, we have also fitted the serine and threonine χ_1 populations by using only the ${}^3J_{NH\beta}$- and ${}^3J_{C'H\beta}$-couplings (not shown). The deviations of the resulting populations from the all-${}^3J$-value results are in most cases smaller than 4%, that is, much smaller than the deviations from the coil values. It should be noted that the fit uses specific Karplus coefficients for serines and threonines.[37] Thus, particular effects of the side-chain oxygen on the size of the J-couplings are corrected. Deviations from the χ_1 coil populations for serines and threonines are also evident from the experimental RDCs, since their agreement with predictions from J-derived χ_1 populations is considerably better than with predictions from the coil populations (not shown).

We attribute this genuine difference of serine and threonine χ_1 populations from the coil average to a destabilization of the side-chain hydroxyl/amide interaction in the urea-denatured state, where urea or water could form hydrogen bonds to both groups. Interestingly, also deviations in backbone conformations have been found for these residues in a recent study using RDCs.[44] Such deviations had not been detected in the earlier studies of χ_1 conformations in unfolded proteins,[16−18] possibly due to the more limited precision. In contrast, stronger deviations were observed for aromatic residues in urea-denatured

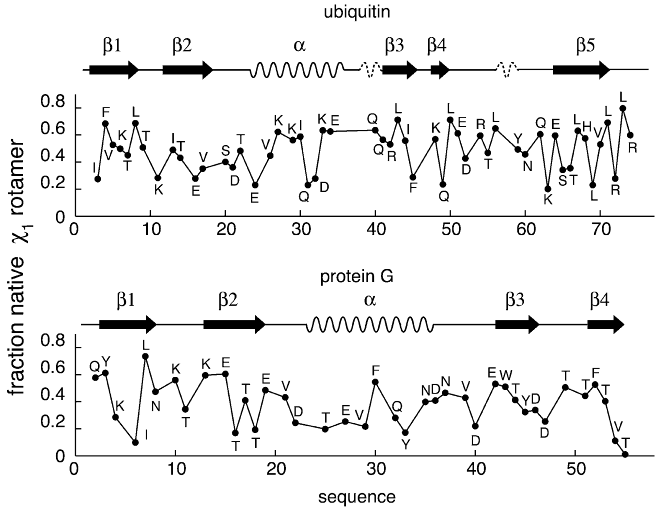


Figure 7. Fraction of native-state χ_1 angles contained in ^3J-derived rotamer populations of urea-denatured ubiquitin (top) and protein G (bottom). Native-state χ_1 angles are approximated as staggered rotamers and taken from the first entry of the folded native NMR structure (PDB code 1d3z)[46] and the 1.1 Å X-ray structure of protein G (PDB code 1igd).[47] The native-state secondary structure of both proteins is shown at the top of the respective panels.

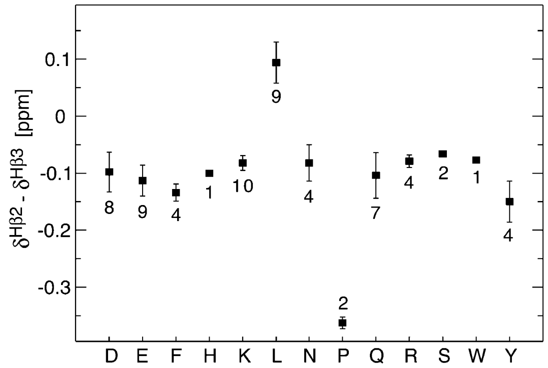


Figure 8. Differences of β-methylene ^{1}H chemical shifts in urea-denatured proteins. Average chemical shift differences of stereospecifically assigned $H^{\beta2}$ and $H^{\beta3}$ resonances are shown for all amino acids that could be observed in ubiquitin and protein G. Error bars indicate standard deviations. Data are labeled by the number of observations.

lysozyme.[18] This effect is not clearly visible for the 10 aromatic residues in urea-denatured ubiquitin and protein G (Figure 5), where the deviations from the coil predictions and their individual variations are not stronger than for many other amino acids such as, for example, valines and leucines.

It is expected that the observed variations in the χ_1 populations of the urea-denatured state correspond to sequence-specific structural preferences, such as the recently detected 10−20% population of the native-state, first β-hairpin (residues M1−V17) in ubiquitin.[15] However, the exact extent of the structural preferences is currently difficult to establish, because proper random coil "baseline" χ_1 populations are not known with sufficient precision in solution to be able to interpret population differences on the order of 10% with confidence. Nevertheless, it is clear that a high degree of native-state χ_1 conformations is contained in the observed χ_1 populations of the urea-denatured states. Figure 7 shows this fraction of native-state side-chain conformations in the χ_1 populations when native-state χ_1 angles are approximated by staggered rotamers. High native rotamer populations (>50%) are found for a number of residues in ubiquitin's β-strands $\beta1$, $\beta3$, $\beta4$, and $\beta5$, and in its α-helix, as well as in all β-strands of protein G. Average native rotamer populations are 49% for ubiquitin and 38% for protein G. This indicates that the transition to a fully formed structure does not require a particularly high entropic cost.

Stereospecific Methylene H^β Chemical Shifts. The availability of the stereospecific assignments for the $^1H^{\beta2/3}$ resonances allows one to analyze their chemical shift behavior. Average chemical shift differences, $\delta H^{\beta2} - \delta H^{\beta3}$, are shown in Figure 8 for all amino acids with distinct β-methylene resonances that could be observed in both proteins. For all observed amino acids (D, E, F, H, K, L, N, P, Q, R, S, W, Y) with the exception of leucine and proline, $\delta H^{\beta2}$ is smaller than $\delta H^{\beta3}$ by about 0.1 ppm. Thus, $H^{\beta2}$ is usually upfield from $H^{\beta3}$ (see also Figure 1, K10). Variations of individual amino acids around the mean shift difference are about 0.05 ppm. For prolines, the upfield shift of $H^{\beta2}$ is stronger (0.36 ppm). In contrast, for leucine, $H^{\beta2}$ is found downfield of $H^{\beta3}$ with a mean chemical shift difference of 0.09 ppm. This behavior is evident in Figure 1, where the intensity patterns of the upfield and downfield H^β protons are inverted for L10 relative to K10 in the two quantitative $^3J_{NH\beta}$ and $^3J_{H\alpha H\beta}$ experiments. We speculate that this unusual behavior is related to the extreme bulkiness of the two leucine δ-methyl groups, which restricts the entire side-chain conformations in folded structures to only two strongly populated classes ($\chi_1/\chi_2 = -60°/180°$ or $180°/+60°$) in folded structures[45] and also causes the extremely high (70−80%) p_{-60} and low (1−17%) p_{+60} values in the unfolded proteins (Figure 5).

Conclusion

In summary, we have presented optimized detection schemes for side-chain $^1H^\beta$ resonances in unfolded proteins that yield highly precise structural information about the χ_1 angle from up to six $^3J_{H\alpha H\beta}$, $^3J_{NH\beta}$, and $^3J_{C'H\beta}$ coupling constants and up to two $^1D_{H\beta C\beta}$ RDCs. Interpretation of the detected ^3J couplings in urea-denatured ubiquitin and protein G by a model of staggered χ_1 rotamers[38−40] and previously published Karplus coefficients[37] provides stereoassignments of $^1H^\beta$ methylene protons and yields excellent agreement. This corroborates both the staggered rotamer model and the high precision of the Karplus coefficients. For most residues, the precision of individual rotamer populations is better than 2% as estimated from error propagation. As found in earlier studies,[18] the rotamer populations are in the vicinity of averages obtained from coil regions of folded protein structures. However, individual variations from these averages of up to 40% are highly significant and must originate from sequence- and residue-specific interactions. Particularly strong deviations from the coil average are found for serine and threonine residues, an effect that may be explained by a weakening of side-chain to backbone hydrogen bonds in the urea-denatured state.

The measured $^1D_{H\beta C\beta}$ RDCs correlate well with predicted RDCs based on steric alignment of a coil model ensemble of the unfolded state generated by the program *Flexible*-Meccano,

---

(44) Nodet, G.; Salmon, L.; Ozenne, V.; Meier, S.; Jensen, M.; Blackledge, M. *J. Am. Chem. Soc.* **2009**, *131*, 17908–17918.

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

(47) Derrick, J.; Wigley, D. *J. Mol. Biol.* **1994**, *243*, 906–918.

(45) Ponder, J. W.; Richards, F. M. *J. Mol. Biol.* **1987**, *193*, 775–791.

where the side-chain conformations had been adjusted according to the J-derived χ_1 rotamer populations. This agreement validates the coil model as a good first approximation of the unfolded state. However, deviations between the measured and predicted values also indicate that the local backbone geometries may be improved by incorporation of the additional RDC information. The ease of detection of a large number of highly precise side-chain RDCs should make it possible to obtain such a more accurate description of backbone and side-chain conformations in unfolded states.

Acknowledgment. We thank C. Howald, K. Rathgeb-Szabo, and M. Rogowski for sample preparation, and we also thank H.-J. Sass and S. Kasprzak for helpful discussions. This work was supported by SNF Grant 31-109712.

Supporting Information Available: Details of quantitative $^3J_{NH\beta}$-HNHB, $^3J_{C'H\beta}$-HN(CO)HB, $^3J_{H\alpha H\beta}$-HAHB(CACO)NH, and IPAP-HBHA(CO)NH experiments, figures showing comparisons of experimental and calculated ^3J and RDC values for all observed residues in urea-denatured ubiquitin and protein G, and tables of measured ^3J and RDC values as well as of the fitted χ_1 rotamer populations. This material is available free of charge via the Internet at http://pubs.acs.org.

JA910331T

Chapter 2

C-di-GMP Equilibria at Physiological Concentrations

2.1 Introduction

Cyclic diguanylate (c-di-GMP) is an only recently discovered ubiquitous bacterial second messenger, which is not found as a signaling molecule in eukaryotes and archaea[165,206,236]. Among its many functions, c-di-GMP triggers the switch from motile to sessile bacterial lifestyles and is a key component in bacterial biofilm formation and maintenance[115]. The interest in c-di-GMP and its signaling mechanisms is considerable, since most bacteria contain proteins which are involved in either c-di-GMP synthesis, degradation or binding (as effector molecule) or a combination of these three[165,216].

Mostly functional studies have been performed in this field. In contrast, detailed structural and kinetic studies, both of involved proteins and c-di-GMP itself are less common.

C-di-GMP is a C_2-symmetric molecule, composed of two GMP moieties which are linked *via* the 5' phosphate of each GMP subunit to the 3' oxygen atom of the other monomer and *vice versa*, to form a 12-membered macrocycle (Figure 2.1 A). This allows several structures even for the monomer, which have been designated as open (a rather linear, streched form) and closed form (with the guanine bases pointing to the same direction, also called U-shaped), respectively (Figure 2.1 A and B). While many structures of effector proteins bind dimeric, intercalated c-di-GMP[165], only a smaller number of receptors binding monomeric c-di-GMP has been found. However, both monomers and dimers have been observed *in vitro* and also their importance *in vivo* is generally accepted, since both forms of c-di-GMP play important roles in bacterial signaling networks and are related to pathological effects as well as biofilm formation[94,165,216].

Nevertheless, it is still unknown whether c-di-GMP dimerization occurs in a state, where a c-di-GMP monomer is bound to the protein and an additional monomer binds or c-di-GMP dimers are bound directly as one entity. It is thus an important goal to study c-di-GMP equilibria and interactions at the atomic level for a better understanding of the regulation of cellular c-di-GMP concentrations, taking into account all involved dynamic equilibria. This could eventually lead to the development of strategies to disrupt bacterial biofilms.

2.1.1 Physiological context of c-di-GMP

In the original discovery by Benziman and co-workers, c-di-GMP has been found to stimulate cellulose synthase activity in *Gluconacetobacter xylinus* upon allosteric binding[209]. Today, it is known that c-di-GMP cannot only trigger the switch from the motile, planktonic to the sessile, biofilm-related lifestyle[94], but it is also related to virulence of pathogens[22,119,183], cell cycle progression and cell differentiation[1,55], production of antibiotics[245], modulation of secretion[169,258], transcriptional regulation[95], synthesis of metabolites[66], exopolysaccharide production[160–162,167,179,193] and several other functions. Recently, it has even been shown that c-di-GMP is involved in bioleaching[210], which is the process of metal extraction from their corresponding ores by microorganisms. C-di-GMP controls biofilm formation in *Acidothiobacillus ferrooxidans*, which are required for its bioleaching processes and which could make it interesting for metal production[210].

Most known bacteria possess the molecular machinery to produce and cleave c-di-GMP[165,216]. The characteristic GGDEF motif (named after the involved amino acids) is found in proteins which have diguanylate cyclase (DGC) function and perform c-di-GMP synthesis from two GTP molecules (Figure 2.1). Two proteins, each containing the characteristic GTP-binding GGDEF domain form an antiparallel homodimer which is mostly mediated via N-terminal domains, preceding the catalytic domain. Dimerization is often required for the enzymatic DGC activity[216].

For homodimerization, several motifs such as the tandem REC domains in PleD from *Caulobacter crescentus*[32,264], coiled-coil motifs (e.g. in DgcA and DgcB from *C. crescentus*[20,36], or the Leucine zipper motif[45] are found. Interestingly, most proteins containing the GGDEF domain also contain an inhibitory site (I-site) with an RXXD motif (where X represents any amino acid), which binds c-di-GMP and inhibits the cyclase activity[216]. This is referred to as allosteric feedback inhibition of DGCs[32,264]. To date, only one example of an active DGC protein without the canonical GGDEF sequence has been found[191]. Instead of GGDEF, this protein contains an SGDEF motif, which has been proposed to represent an alternative active-site motif.

In contrast, two motifs have been identified, which degrade c-di-GMP into the linear pGpG molecule[94,165,206,216,236] (Figure 2.1). These are EAL domains or the less commonly found HD-GYP domains, which have phosphodiesterase (PDE) activity. Monomeric c-di-GMP in the open form is bound to such domains before cleavage. The linear pGpG is then further degraded into two GMP molecules by other protein

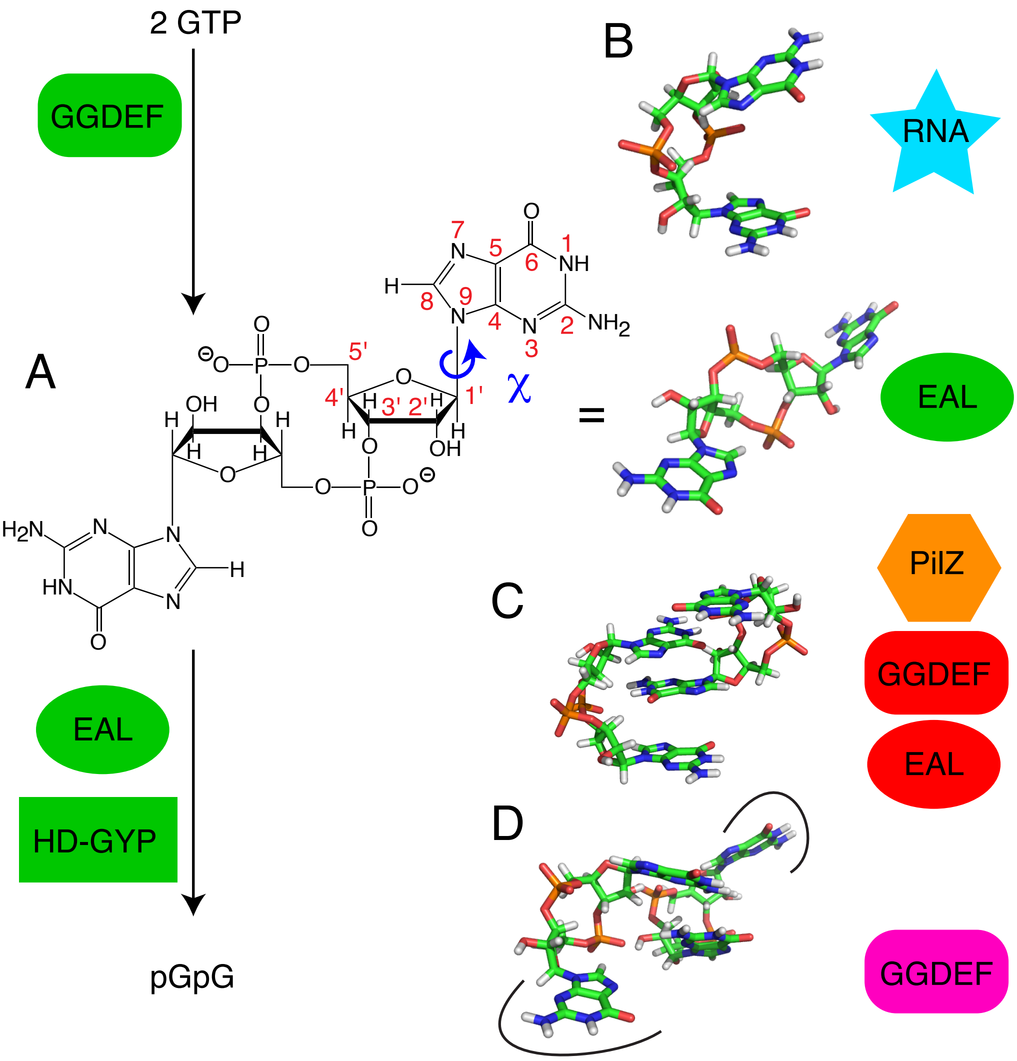


Figure 2.1 Physiological context of c-di-GMP. The characteristic sequence motifs for diguanylate cyclases (GGDEF) and phosphodiesterases (EAL and HD-GYP) are shown to represent the active protein domains. Production and degradation of c-di-GMP is shown in the left column. The open monomer[180] (A), the closed monomer[229] (B), the intercalated dimer[86] (C) and the extended dimer[276] (D) of c-di-GMP are shown, with their corresponding receptors (right). Color changes from green to red (GGDEF and EAL domains) indicate c-di-GMP binding to the inhibitory site. The purple GGDEF domain indicates a different c-di-GMP binding mode in the active site (cross-linking of two active sites).

domains than EAL. HD-GYP domains can indeed catalyze both phosphodiesterase steps involved in c-di-GMP degradation[211].

It is noteworthy that quite often, bacterial proteins, which contain a DGC domain, also contain a PDE domain within the same macromolecule (so-called composite proteins[94,216]). In *G. xylinus*, three out of six identified composite proteins act as DGC and the other three as PDE[235]. Only few examples of composite proteins have been found, which show both DGC and PDE activity[17,85,134,141,246]. The existence of such composite proteins indicates that a tight regulation of the whole c-di-GMP machinery is crucial, since otherwise, they could enter a futile cycle of permanent production, inhibition and degradation[1].

A multitude of GGDEF and EAL domain proteins is found within individual bacterial species[94]. This is particularly pronounced for Gram-negative bacteria but less common for Gram-positive bacteria. While *E. coli* has 29 GGDEF or EAL proteins, *C. crescentus* has 14 and various *Vibrio* species contain more than 50 such proteins[94]. It is assumed that with such a pool of proteins, oscillating c-di-GMP levels[36] (e.g. 'local spikes') for temporal and/or spatial sequestration could be created and thus would allow precise control of the related signaling network[94].

In most cases, both DGC and PDE domains are linked to signal input sensory domains, either directly or *via* a two-step phosphorylation cascade[72]. Such sensory domains are commonly found on the N-terminal side of GGDEF or EAL domains, indicating that the activity of different protein domains can be triggered by specific, sensor-dependent input signals. These sensory domains include heme- or flavin-associated PAS domains[34,48,214,234], the oxygen-binding hemerythrin domain[69], the NO-sensing H-NOX domain[149], the blue-light-sensing BLUF domain[89,175], domains, which bind small molecules (e.g. GAF[103]), or a recently discovered chemoreceptor zink binding (CZB) domain[54].

A substantial number of proteins exists, which bind c-di-GMP as an effector molecule. These include single- or multi-domain PilZ proteins (named after the type IV pilus control single-domain protein in *Pseudomonas aeruginosa*[4]), or degenerate GGDEF or EAL domains, which still contain the I-site[55,161]. Additional proteins have been found to interact with c-di-GMP in functional studies, even though none of the known binding motifs could be found in their amino acid sequences[35,139,239]. Recently, c-di-GMP has been shown to interact with two classes of structurally unrelated riboswitches[140,228,229,232], adding RNA molecules to the diversity of c-di-GMP receptors.

PilZ domains, containing the signature motifs RXXXR and DXSXXG, are so far the best studied class of c-di-GMP receptors[94]. The link between c-di-GMP and PilZ homolog domains was first discovered by *in silico* methods[5], as the presence of a PilZ domain in cellulose synthase of *G. xylinus* was observed by sequence alignment. In the same study, several PilZ domains were detected in other organisms. Subsequently, multidomain proteins containing a PilZ domain or single-domain PilZ

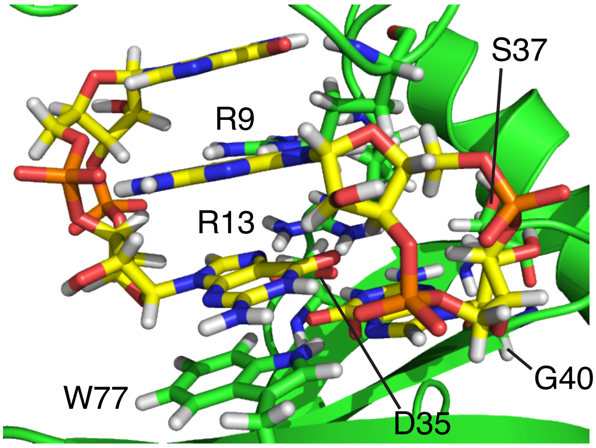


Figure 2.2 Dimeric c-di-GMP bound to the *P. aeruginosa* PilZ domain protein PA4608 (pdb code 2L74[86]). The characteristic residues of the PilZ domain motifs (RXXXR and DXSXXG) as well as the interacting tryptophane W77 are shown.

proteins have been studied experimentally[37,161,199,212] and confirmed the predicted c-di-GMP binding.

C-di-GMP adopts an intercalated, base-stacked dimer structure (Figures 2.1 C and 2.2) in complex with PilZ domains[86,129,222], which is very similar to the c-di-GMP crystal structures in complex with metal ions[61,144]. Virtually idendical c-di-GMP dimers are found in the inhibition site of diguanylate cyclases[33,46] and in a response regulator[132].

Recently, an exceptional c-di-GMP dimer structure in complex with a protein has been observed[276] (Figure 2.1 D). In this case, the two inner guanine bases stack while the outer guanine moieties interact with residues in protein cavities.

C-di-GMP-degrading EAL domains bind to the open form of monomeric c-di-GMP in their active site[14,166,180] (Figure 2.1 A). In contrast, c-di-GMP riboswitches both bind to the closed form of c-di-GMP monomers[228,229] (Figure 2.1 B). No higher c-di-GMP oligomers than dimers have been observed to be bound to any receptor so far.

The multitude of proteins that are involved in c-di-GMP interactions and the knowledge that solely bacteria exploit this molecule as secondary messenger makes it an attractive therapeutic target. Additionally, in recent time, the interest in the interplay of c-di-GMP with other bacterial second messengers and alarmones such as cAMP, cGMP, ppGpp and pppGpp as well as mechanisms such as quorum sensing has grown substantially[211].

Apart from all the described knowledge of c-di-GMP, the substance itself is subject to a variety of studies. Even though monomeric as well as dimeric c-di-GMP have been observed to bind to several proteins and much of the related signaling has been elucidated, the *in vitro* behavior of c-di-GMP, particularly in the presence of metal ions, reveals additional aspects, which are not completely understood.

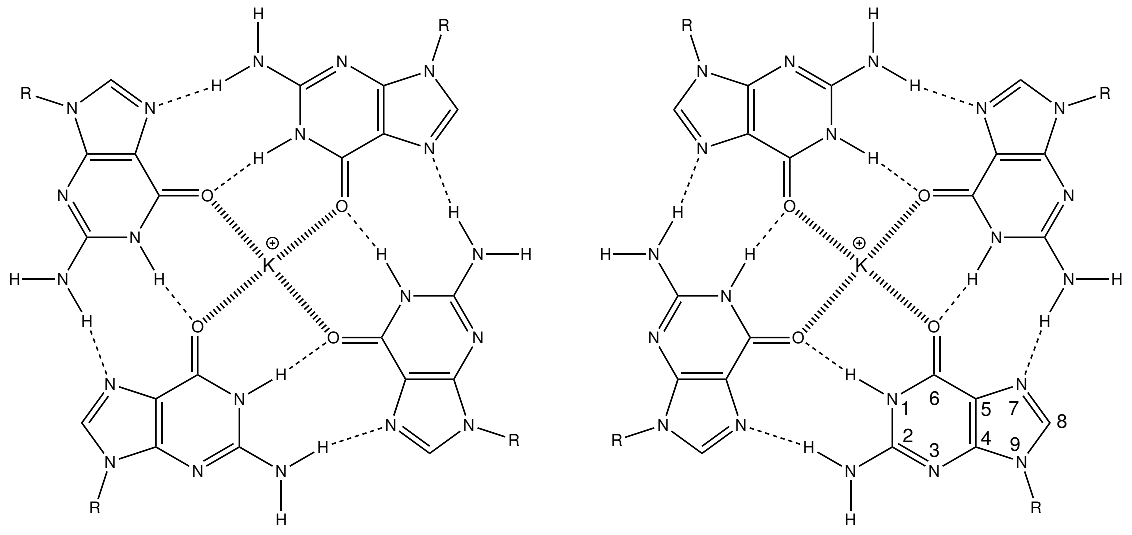


Figure 2.3 Chemical structures of guanine quartets (G-quartets) in the presence of a potassium ion. C8 atoms of the G-quartets are oriented clockwise (left) or counterclockwise (right). R denotes the ribose moiety. Atom numbering is shown for one guanine (bottom, right).

2.1.2 Polymorphism of c-di-GMP

Several studies recently reported oligomeric forms of c-di-GMP in addition to the well-known intercalated dimers, referred to as c-di-GMP polymorphism[279,280]. These oligomers (arguably tetramers and octamers) are formed in the presence of metal cations, particularly potassium, and are proposed to be based on guanine quartet[44] (G-quartet or G-quadruplex) scaffolds according to CD, UV and NMR data.

G-quadruplexes

G-quadruplexes are structures which can be formed by DNA and RNA sequences containing multiple guanine moieties[136]. In the presence of metal cations, particularly potassium, these sequences fold into compact and stable (resistant to DNAses, high melting temperature) G-quartets (Figure 2.3). Often, stacks comprising several G-quartets are formed. Sequences which show this behavior are found in telomeres and at other parts of the genome, particularly in promoters[101,102]. Therefore, G-quadruplexes have been reported to possess regulatory potential[255].

Several aspects of G-quartets are not fully understood, particularly the detailed kinetics and thermodynamics even though this may be relevant for biology[136]. It is also not clear which forces determine the topology of these structures, since only few of the 26 possible topologies have been observed *in vitro*, despite the fact that more than 90 G-quadruplex structures are deposited in the protein data bank[136].

However, it is clear that the presence of cations such as potassium, sodium or sometimes even ammonium is a prerequisite for G-quartet formation[44]. Since c-di-GMP polymorphism has also been shown to depend on cations[279,280], it seems reasonable to assume that G-quartets are the basis of c-di-GMP oligomers.

The existence of c-di-GMP oligomers based on G-quartets was first suggested when similarities of the UV spectra of c-di-GMP and a guanine-rich DNA double-strand, each in the presence of 9-aminoacridine (an aromatic DNA intercalator) were observed[144]. The authors speculated that due to the size of the intercalator, c-di-GMP could form highly symmetric G-quartet structures (c-di-GMP tetramers) with a central cavity, which mimics an intercalator binding site of DNA. However, structural details such as the preference of the glycosyl torsion angle χ (syn or anti) were not discussed.

Based on a more recent study comprising NOESY and DOSY NMR data, oligomers at 30 mM c-di-GMP concentration have been attributed to octamer and tetramer structures, respectively, for which all-syn and all-anti conformations have been proposed to coexist[280]. Nevertheless, the stoichiometry of these oligomers remains ambiguous, since no high-resolution structure of any of the proposed structures has been solved up to date.

Very recently, formation of G-quartet-based c-di-GMP oligomers has also been inferred from CD and fluorescence data at low micromolar c-di-GMP concentrations, in the presence of potassium and aromatic dyes, which have been proposed to act as templates for G-quartet formation[126,176,177]. The authors even suggested potential mechanisms how the dye would intercalate or end-stack with c-di-GMP oligomers and that both tetramers and octamers could be involved in these processes. In a recent functional study, enhancement of peroxidase activity was observed in the presence of hemin, c-di-GMP and proflavin at high potassium concentrations[178]. This effect has been attributed to intercalation and/or end-stacking of a whole complex consisting of potassium-stabilized tetrameric and/or octameric c-di-GMP, proflavin and hemin. Surprisingly, the enhancement was much stronger in the presence of lithium than in the presence of potassium, indicating that tetramers are primarily responsible for this effect[280]. At high millimolar c-di-GMP concentrations tetramers were the dominant species in the presence of lithium, whereas octamers were higher populated in the presence of potassium, with the larger potassium ions located between adjacent G-quartets and the smaller lithium or sodium ions in the center of the G-quartet plane[280].

Even though several studies surmise the existence of G-quartets in oligomers of c-di-GMP by UV, CD, fluorescence or other techniques, this remains to be proven experimentally by a high-resolution technique, such as X-ray crystallography or NMR spectroscopy[136]. Some experimental NMR evidence are the presence of H1 resonances in NMR spectra, which indicate that these protons are protected from exchange with the solvent[280]. Furthermore, some analogy is seen between separated syn and anti NMR chemical shifts of hydrogen bonded NH_2 protons of c-di-GMP octamers and the chemical shifts of a known G-quartet structure[154]. Nevertheless, the structure determination of c-di-GMP oligomers is required in order to understand the intercalation mechanisms and assess their relevance *in vivo*.

2.1.3 Oligomerization of related nucleic acid compounds

C-di-AMP as well as cAMP and cGMP have all been identified as bacterial messenger molecules but are not prone to form higher oligomers in contrast to c-di-GMP. It had been expected that structural similarities between c-di-GMP and the investigated, related compounds lead to similar stimulation of cellulose synthase activity. Therefore, the stimulatory effect of a nucleic acid derivative on cellulose synthase activity was determined by functional experiments and directly compared to the activity enhancement by c-di-GMP[209].

An NMR study of c-di-dAMP (the deoxy form of c-di-AMP) revealed that it can form similar intercalated dimers as c-di-GMP, but lacks cellulose synthase stimulation, indicating that the formation of intercalated nucleotide dimers is not sufficient to enhance cellulose synthesis[19]. Furthermore, upon heating, c-di-dAMP could be dissociated into monomers and it was argued that also an open and a closed form of these monomers exist. In the 17 mM sodium form of c-di-dAMP, no additional forms besides dimers and monomers have been observed and many different ^3J coupling constants were measured and subsequently transformed into torsion angles *via* Karplus relations. These angles describe the dimer at several temperatures indicating that it is restricted to a single conformation.

Even though c-tri-GMP has been found to modestly activate cellulose synthesis, only either one single, highly symmetric species or a rapidly averaging spectral species of symmetric and asymmetric forms is observed by NMR[168]. The authors emphasize that the inability of this structure to stimulate cellulose synthase activity is not only due to the absence of stacked dimers. They deduced requirements for cellulose synthase activator molecules, indicating that the imino proton H1 and the O6 oxygen are necessary for function, whereas a carbonyl group at position 2 prevents a stimulatory effect.

A variety of methods to synthesize c-di-GMP and its derivatives have been published. Both enzymatic[95,124,161,230,237,277] and chemical[6,71,91,123,127,151,208,275,279] synthesis routes have been developed to provide a basis for structural and functional research.
Chemically synthesized oligoethylene-oxide tethered c-di-GMP has been shown to form dimers and also the presumably G-quartet-based oligomers[280] in the presence of potassium. This observation has been attributed to the formation of the typical G-quartet hydrogen bonds[44] (Figure 2.3), despite the larger distance between adjacent G-quartets.
In contrast, no higher oligomers than dimers have been observed for the c-GMP-IMP molecule, a non-C_2 symmetric c-di-GMP analog consisting of a guanosine monophosphate and an inosine monophosphate moiety[280]. Although this molecule lacks one NH_2 group, at least the formation of single G-quartets could be envisioned.

This indicates that the NH_2 groups on both bases are indispensable for oligomer formation of c-di-GMP and its analogs.
More complex patterns of oligomers have been found for a series thiophosphate analogs of c-di-GMP, where one or more non-bridging oxygen atoms were substituted by sulphur atoms[281]. Replacement of a bridging 5'-oxygen to phosphorus (endo-S-c-di-GMP) prevented oligomer formation, indicating that also the phosphate moiety plays an important role in c-di-GMP oligomers[259]. Interestingly, the latter construct showed also biochemical differences in cleavage reactions and c-di-GMP receptor binding, attributed to conformational restriction to the open monomeric form. An additional study was performed on C8 variants of c-di-GMP[253,254]. Only the C8-Br variant formed oligomers at high concentrations of potassium. No oligomer formation has been observed for phenyl-, acetylphenyl-, MeS- or HS-C8-substituted c-di-GMP. Apart from such structural studies, some functional studies have been carried out using 2' c-di-GMP variants, indicating that only few proteins tolerate 2'-modification, while most lose their enzymatic activity[151]. However, these studies did not expect other c-di-GMP forms than monomers or dimers being responsible for catalysis.

The different properties, of c-di-GMP analogs and related compounds, particularly the oligomerization behavior, may only be completely understood, once the structure of c-di-GMP oligomers is solved. It is remarkable, that a substitution of H8 also affected the oligomerization behavior, since H8 is not involved in hydrogen bonds of G-quartets (Figure 2.3).

2.1.4 Unaddressed c-di-GMP issues

Open questions in the c-di-GMP field have been (1) whether higher oligomers (than dimers) are relevant under *in vivo* conditions and (2) the structure and the mechanism of formation of higher c-di-GMP oligomers. Investigations of c-di-GMP at low micromolar concentrations may also indicate whether c-di-GMP dimerization occurs in solution or on the protein.

Here, we have addressed the kinetics and thermodynamics of c-di-GMP oligomers in buffers resembling physiological salt concentrations at 0.6 μM to 10 mM c-di-GMP concentration range. Furthermore, we present experimental data indicating that all-syn and all-anti c-di-GMP octamers belong to a single species. These data are a first step towards the three-dimensional structure determination of these oligomers and a strategy to complete this objective is discussed.

Organization of the chapter

A study on the kinetics and equilibrium constants of c-di-GMP oligomerization in the presence of potassium cations and particularly the equilibrium between monomers and dimers is described in detail in chapter 2.2. Initial data for the solution structures of higher c-di-GMP homo-oligomers and a discussion of the importance of solving their atomic structures as well as the required steps are presented in the following section (chapter 2.3).

2.2 Oligomer Formation of the Bacterial Second Messenger c-di-GMP

Original Publication

Gentner, M., Allan, M. G., Zaehringer, F., Schirmer, T., and Grzesiek, S.

Oligomer Formation of the Bacterial Second Messenger c-di-GMP: Reaction Rates and Equilibrium Constants Indicate a Monomeric State at Physiological Concentrations

Journal of the American Chemical Society 134 (2012) 1019-1029.

JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

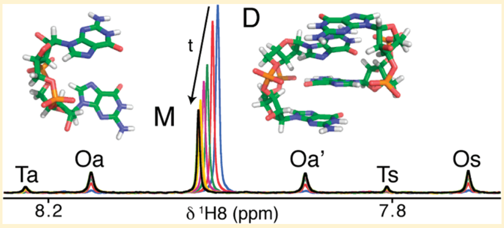
Article
pubs.acs.org/JACS

Oligomer Formation of the Bacterial Second Messenger c-di-GMP: Reaction Rates and Equilibrium Constants Indicate a Monomeric State at Physiological Concentrations

Martin Gentner, Martin G. Allan,[§] Franziska Zaehringer, Tilman Schirmer,* and Stephan Grzesiek*

Biozentrum, University of Basel, Klingelbergstrasse 50/70, 4056 Basel, Switzerland

Ⓢ *Supporting Information*



ABSTRACT: Cyclic diguanosine-monophosphate (c-di-GMP) is a bacterial signaling molecule that triggers a switch from motile to sessile bacterial lifestyles. This mechanism is of considerable pharmaceutical interest, since it is related to bacterial virulence, biofilm formation, and persistence of infection. Previously, c-di-GMP has been reported to display a rich polymorphism of various oligomeric forms at millimolar concentrations, which differ in base stacking and G-quartet interactions. Here, we have analyzed the equilibrium and exchange kinetics between these various forms by NMR spectroscopy. We find that the association of the monomer into a dimeric form is in fast exchange (<milliseconds) with an equilibrium constant of about 1 mM. At concentrations above 100 μM, higher oligomers are formed in the presence of cations. These are presumably tetramers and octamers, with octamers dominating above about 0.5 mM. Thus, at the low micromolar concentrations of the cellular environment and in the absence of additional compounds that stabilize oligomers, c-di-GMP should be predominantly monomeric. This finding has important implications for the understanding of c-di-GMP recognition by protein receptors. In contrast to the monomer/dimer exchange, formation and dissociation of higher oligomers occurs on a time scale of several hours to days. The time course can be described quantitatively by a simple kinetic model where tetramers are intermediates of octamer formation. The extremely slow oligomer dissociation may generate severe artifacts in biological experiments when c-di-GMP is diluted from concentrated stock solution. We present a simple method to quantify c-di-GMP monomers and oligomers from UV spectra and a procedure to dissolve the unwanted oligomers by an annealing step.

■ INTRODUCTION

Cyclic diguanosine-monophosphate (c-di-GMP) has emerged as a second messenger that is used by most bacteria to regulate a switch between a free-living, planktonic and a sedentary, biofilm-related lifestyle.[1] In pathogenic bacteria, these two lifestyles are correlated with virulence and persistence of infection. Presumably due to the low cellular concentration of c-di-GMP in the micromolar range,[2] the ubiquitous presence of this second messenger has only been recognized recently and indirectly from the identification of genes coding for c-di-GMP-associated enzymes in almost all bacterial genomes. In particular, diguanylate cyclases (GGDEF domain proteins) and c-di-GMP-specific phosphodiesterases (EAL and HD-GYP proteins) are commonly found as several variants. Molecular aspects of c-di-GMP signaling have been recently reviewed.[3,4]

C-di-GMP is a C_2-symmetric molecule that consists of a 12-membered macrocycle formed by the ribose and phosphate moieties and two guanine groups (Figure 1A). Several structures have been determined by X-ray crystallography, in both free (Figure 1B)[5−7] and protein-bound forms.[3] They show similar conformations for the macrocycle, with both phosphodiester torsion angles α/β in gauche(+)/gauche(+) conformation.[5] Thus, the backbone of the molecule appears rather constrained. Interestingly, in all aforementioned small-molecule structural studies, c-di-GMP is in dimeric form with intercalated bases and inter-monomeric H-bonding between the guanine N1 and the phosphate groups. Similar dimeric forms have also been reported for c-di-AMP.[8] In some cases, the c-di-GMP dimers coordinate Mg^{2+} or Co^{2+} ions by the N7 atoms of the two central bases, but this does not significantly change the structure.[5,6] The same intercalated c-di-GMP dimer has also been observed in protein complexes, where it binds to the allosteric inhibition site of diguanylate cyclases,[9,10] to PilZ receptors,[11,12] and a response regulator.[13] This prompted the notion that the biologically active species may be dimeric,[14] although for degradation c-di-GMP binds in monomeric form to EAL phosphodiesterases.[15−17] A structure of a PilZ protein with monomeric c-di-GMP has also been reported.[18]

Previous NMR studies on c-di-GMP in solution have been performed only at high (>1 mM) concentrations,[19−21] revealing a rich polymorphism, ranging from dimers to several forms of higher oligomers in a cation-dependent equilibrium. Interestingly, no such polymorphs are observed for a circular trinucleotide of GMP.[22] The higher oligomeric species have been assigned to tetramers and octamers on the basis of UV, CD, NOE, and diffusion experiments.[21] However, the detailed structures of the tetramers and octamers are unknown. For the

Received: August 16, 2011
Published: December 5, 2011

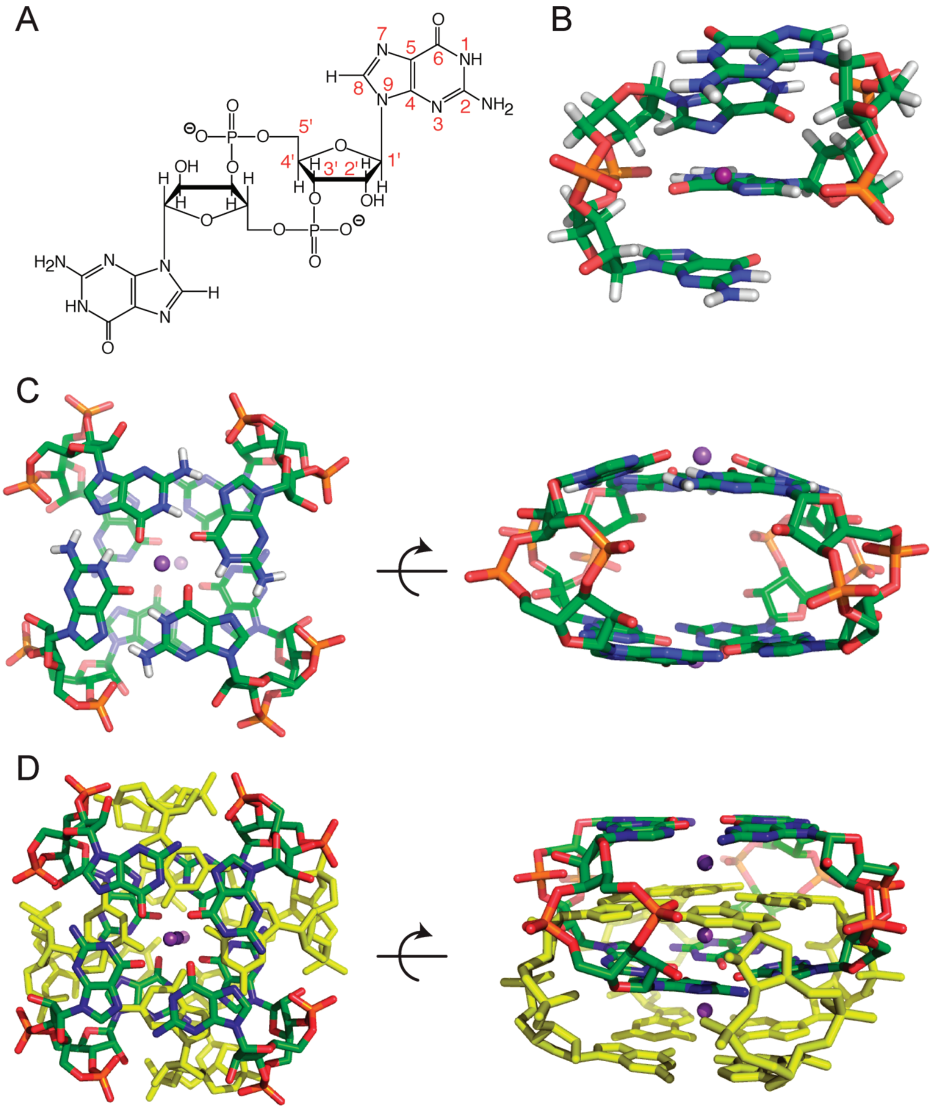


Figure 1. Chemical structure and three-dimensional models of c-di-GMP. (A) Chemical structure and atom numbering of monomeric c-di-GMP in *anti* conformation. (B) Crystal structure of dimeric c-di-GMP in the presence of magnesium, adapted from Egli.[5] (C) Modeled structure of tetrameric c-di-GMP forming two G-quartets in *anti* conformation, based on the schemes proposed by Zhang.[21] Complexed Na$^+$ cations are shown within each of the G-quartets. (D) Modeled structure of all-*anti* octameric c-di-GMP[21] in the presence of three Na$^+$ ions between adjacent G-quartets. The intercalated molecules are displayed as yellow sticks for clarity.

tetramer, a cage-like four-fold symmetric structure has been proposed[7,21] with the two guanine bases of one c-di-GMP monomer contributing to two parallel, but not stacked, quartets. Figure 1C presents an energy-minimized model structure of such a tetramer. The octamer was proposed to consist of two intercalated tetramers.[21] Modeling (Figure 1D) shows that in order to prevent steric clashes, the two intercalating tetramers need to be shifted relative to each other by about half the tetramer height and rotated by about 45° on the common four-fold symmetry axis. This results in a closely packed assembly with four stacked guanine quartets stabilized by monovalent cations situated on the symmetry axis.[23] A recent study revealed that c-di-GMP polymorphism critically depends on the phosphate linkage, since substitution of one of the 5′ oxygens by sulfur leads to a diminished propensity to form oligomers.[19] In contrast to the clear experimental evidence for higher oligomeric forms at millimolar c-di-GMP concentrations, the presence of G-quartets at micromolar concentrations has only been inferred from CD, fluorescence, and absorbance data when additional aromatic intercalators were present.[6,24,25]

The current study aimed to determine the state of c-di-GMP at physiological, i.e., micromolar, concentrations in various buffer conditions. During the study, it became apparent that c-di-GMP exhibits extremely slow kinetics of oligomer formation and dissociation in the presence of metal ions. Slow kinetics have been observed before for G-quartet formation from larger DNA oligomers,[26] but no analysis exists of c-di-GMP oligomer formation. Here, we have determined kinetic parameters and equilibrium constants for c-di-GMP monomer, dimer, tetramer, and octamer exchange in detail. The results indicate that c-di-GMP is monomeric at physiologically low concentrations. However, due to the extremely slow exchange kinetics, care must be taken to reach equilibrium conditions during *in vitro* experiments. Simple methods are presented to quantify the amount of oligomeric state by UV absorbance and to dissociate the oligomers by annealing.

■ MATERIALS AND METHODS

Modeling of c-di-GMP Oligomers. Models of c-di-GMP tetramers and octamers were built manually on the basis of the high-resolution crystal structure[23] (PDB code 352d) of a parallel-stranded Na^+/guanine tetraplex. The models were energy minimized with REFMAC5[27] employing restraints on the interbase H-bond distances (2.8 Å) and the Na^+−O6 distances (2.0 and 3.0 Å, in the case of the tetramer and octamer, respectively).

Sample Preparation. Purified c-di-GMP was produced as described previously by enzymatic synthesis from GTP.[28] In order to mimic cytosolic salt and pH conditions,[29] c-di-GMP was dissolved in 250 mM KCl, 5 mM Na_2HPO_4/NaH_2PO_4, pH 7.4, 5% D_2O in the presence of either 10 mM Na_2EDTA (buffer A) or 10 mM $MgCl_2$ (buffer B). To investigate c-di-GMP in the absence of cations, c-di-GMP was also dissolved in 5 mM TRIS/HCl (tris(hydroxymethyl)-aminomethane, HCl salt), pH 8.0, 5% D_2O (buffer C). Samples were then transferred into 5 mm standard NMR tubes (volume >400 μL) for low c-di-GMP concentrations (<2 mM) or Shigemi tubes (270 μL) at higher concentrations.

1D NMR Spectroscopy. Unless indicated otherwise, samples were heated to 60 °C for at least 1 h to facilitate the dissociation of oligomers prior to measurements and kept at room temperature for 1 h before NMR measurements. NMR experiments were carried out on a Bruker Avance DRX 600 spectrometer equipped with a triple-resonance pulse field gradient probe head (TXI) at a temperature of 24 °C. 1D proton NMR spectra were recorded with the proton carrier set on water and the excitation sculpting scheme[30] achieving water suppression by gradient dephasing. Spectra were recorded as 57 344 complex points with an acquisition time of 1.99 s and a recovery delay of 1 s. Total NMR experiment times ranged from 6 min to 15 h, depending on the c-di-GMP concentration (∼10 mM to ∼0.5 μM, respectively). Spectra were processed and evaluated using Topspin 2.1.6 (Bruker). Chemical shifts were determined relative to external trimethylsilyl-2,2,3,3-tetradeuteriopropionic acid in water (pH 7).

Determination of c-di-GMP Concentrations. Due to the hypochromic effect upon oligomerization (see below), c-di-GMP concentrations were determined not by UV absorption measurements but by comparing NMR peak intensities (normalized on the number of scans and on the proton 90° hard pulse length)[31] to the intensity of a reference GTP sample. More specifically, the peak integrals of the well-separated, aromatic H8 resonances were compared to the integral of the GTP H8 resonance (50 μM) in the identical buffer. Very similar results were obtained when H1′ resonances were used for quantification. The concentration of the GTP sample was determined by UV measurement at 253 nm using a molar extinction coefficient of 13 700 M^{-1} cm^{-1}.[32] The investigated c-di-GMP concentration ranges were 0.5 μM−6 mM (buffers A and B) and 1 μM−10 mM (buffer C). UV measurements were carried out in quartz cuvettes (Hellma) with a path length of 1 cm on an HP/Agilent 8453 spectrophotometer or a Varian Cary 50 spectrophotometer.

To exclude that oligomer formation would affect the GTP reference, we compared NMR spectra of 100 μM and 50 μM GTP dissolved in buffer A (Figure S9). Neither oligomer resonances nor any resonance shifts were observed. Repeated measurements after more than 10 days revealed less than 2% changes in both H8 and H1′ resonance integrals. Thus the effects of GTP oligomerization and H/D exchange are negligible.

Determination of c-di-GMP Dimer Dissociation Constants. C-di-GMP monomer and dimer were observed in fast chemical exchange on the chemical shift time scale, leading to a single set of resonances for monomer and dimer. We denote this combined spectral species as MD. The experimentally observed chemical shifts for MD (δ_{MD}) represent the population-weighted mean,

$$\delta_{MD} = \frac{[M]\delta_M + 2[D]\delta_D}{[M] + 2[D]} = \frac{[M]\delta_M + 2[D]\delta_D}{[MD]} \tag{1}$$

where δ_M, δ_D are the chemical shifts and [M], [D] the concentrations of monomers and dimers, respectively, and [MD] = [M] + 2[D] is the concentration of monomers plus dimers (in monomeric units). Since the dimer dissociation constant K_{MD} is defined as

$$K_{MD} = \frac{[M]^2}{[D]} \tag{2}$$

the monomer concentration can be expressed as a function of K_{MD} and [MD] as

$$[M] = -\frac{K_{MD}}{4} + \sqrt{\left(\frac{K_{MD}}{4}\right)^2 + \frac{[MD]K_{MD}}{2}} \tag{3}$$

Substitution into eq 1 then yields the calculated shift for the MD resonance (δ_{calc}) as a function of [MD], K_{MD}, δ_M, and δ_D. The unknown parameters (K_{MD} and δ_D) were determined by fitting δ_{calc} to the experimentally determined chemical shift δ_{exp} (H8 resonance) via minimization of the target function,

$$\chi^2 = \sum_i \left(\frac{\delta_{exp,i} - \delta_{calc,i}(K_{MD}, [MD])}{\sigma_i} \right)^2 \tag{4}$$

by an in-house-written Matlab (MathWorks) routine. Errors in the fit parameters were determined by a Monte Carlo procedure.

Modeling of Kinetic Parameters. The time courses of oligomer concentrations were modeled according to the kinetic scheme given in Figure 6. The coupled differential equations for the concentrations of the slowly interconverting monomers, tetramers, and octamers were assumed as

$$\begin{aligned} d[M]/dt &= -4k_{TM}[M]^4 + 4k_{MT}[T] \\ &\quad - 4k_{OT}[T][M]^4 + 4k_{TO}[O] \\ d[T]/dt &= k_{TM}[M]^4 - k_{MT}[T] - k_{OT}[T][M]^4 \\ &\quad + k_{TO}[O] \\ d[O]/dt &= k_{OT}[T][M]^4 - k_{TO}[O] \end{aligned} \tag{5}$$

with k_{BA} representing the rate of the A→B reaction.

Since, the monomer/dimer reaction is fast on the time scale of these reactions, the [M] and [D] concentrations were assumed to be in instant equilibrium according to eq 2 during the simulations. Furthermore, the equally populated but spectroscopically distinct two tetrameric species (Ts, Ta) and three octameric species (Os, Oa, Oa′) were subsumed into single tetrameric and octameric concentrations, such that [Ts] = [Ta] = [T]/2 and [Os] = [Oa] = [Oa′] = [O]/3.

The differential equations were integrated numerically using the program ProFit (Quansoft, Zürich). Non-uniform time steps were used such that the change in concentration of each species was not larger than 0.1%. Forward and backward rates of reactions M⇄T and T⇄O as well as the boundary conditions for the concentrations at time $t = 0$ were then obtained by fitting the integrated time courses to the observed data using the same program. Since the sum of the observed NMR concentrations ($\sum_{obs}$) of the identified species was not constant, but decreased (Figure 4B) or increased (Figure 5A) slightly with time (presumably due to the formation or dissociation of higher, unobservable oligomers, see below), the total c-di-GMP concentration $[\text{c-di-GMP}]_t$ was adjusted for each time point to the value obtained by fitting $\sum_{obs}$ to an exponential function. Errors in the fit parameters were determined by a Monte Carlo procedure.

Modeling of Time-Dependent UV Absorption Spectra. The behavior of the UV absorption spectrum $A(\lambda)$ of a c-di-GMP sample after dilution was modeled as a linear combination of absorption spectra from the unstacked (monomeric) form $A_{mono}(\lambda)$ and a second, distinct spectrum of a stacked form $A_{\text{stacked}}(\lambda)$, such that

$$A(\lambda) = A_{\text{mono}}(\lambda)p_{\text{mono}} + A_{\text{stacked}}(\lambda)(1 - p_{\text{mono}}) \qquad (6)$$

where the population of the monomeric form p_{mono} was derived from the NMR intensity and frequency position of the H8 proton. The component, modeled absorption spectra $A_{\text{mono}}(\lambda)$ and $A_{\text{stacked}}(\lambda)$ were then obtained by a linear fit of eq 6 to experimental absorption spectra acquired a different time points after dilution.

■ RESULTS AND DISCUSSION

1D NMR Spectra of c-di-GMP at Various Concentrations. In the presence of metal cations, complex polymorphism of c-di-GMP has been observed at high millimolar concentrations previously by NMR spectroscopy.[20,21] Here we have addressed the question of the c-di-GMP oligomerization state at conditions that resemble the physiological environment with respect to c-di-GMP concentration, salt, pH, and temperature or at conditions used routinely for the functional *in vitro* characterization of c-di-GMP-related enzymes.

C-di-GMP under Physiological Salt and pH Conditions. Figure 2A,B shows the H1, H8, and H1′ parts of c-di-GMP ^{1}H spectra at various concentrations ranging from 19 μM to 1 mM in buffer A (250 mM KCl, 5 mM Na_2HPO_4/NaH_2PO_4, 10 mM Na_2EDTA, pH 7.4), which was chosen to mimic the salt concentrations and pH of the bacterial cytosol.[29] At a concentration of 1 mM, c-di-GMP exhibits several resonances for the H1, H8, and H1′ protons as well as for the other aliphatic ribose protons (not shown). Zhang et al.[20,21] have observed similarly complex spectra at high (>30 mM) c-di-GMP concentration in the presence of monovalent cations. This spectral heterogeneity was attributed to several oligomeric species comprising an intercalated dimer, as well as several types of tetramers and octamers. The propensity of oligomer formation was much more pronounced for K^+ than for Li^+ and Na^+ cations.

It is apparent from Figure 2B that the frequencies of the tetramer and octamer resonances remain constant with varying c-di-GMP concentration, whereas their relative intensities change. Hence these species are in slow exchange on the chemical shift time scale, i.e., significantly slower than about 17 ms, which corresponds to the inverse of their ∼0.1 ppm frequency separation. In contrast, a number of resonances exhibit concentration-dependent shift changes. This is exemplified for the H8 resonance in Figure 2B. We attribute this set of resonances to c-di-GMP monomers and dimers, designated as MD, which are in fast exchange on the chemical shift scale. Since the resonances are not significantly broadened, the exchange must be considerably faster than about 34 ms, corresponding to a frequency separation of at least 0.05 ppm. It is also noted that, in contrast to tetramers and octamers, no imino proton H1 resonance is observed for the MD species. This indicates the absence of H-bonding and rapid exchange with the solvent. In a previous study,[21] this set of MD resonances had been attributed to an intercalated dimer in *anti* conformation that exchanges with an unstructured form. However, the exchange had not been characterized further.

We also investigated whether the presence of divalent cations had an influence on the observed oligomer formation. For this purpose, the experiments were repeated in the presence of magnesium in buffer B (250 mM KCl, 5 mM Na_2HPO_4/NaH_2PO_4, 10 mM $MgCl_2$, pH 7.4). A behavior very similar to that for buffer A was observed (Figure S1). Thus magnesium does not significantly influence the oligomerization process.

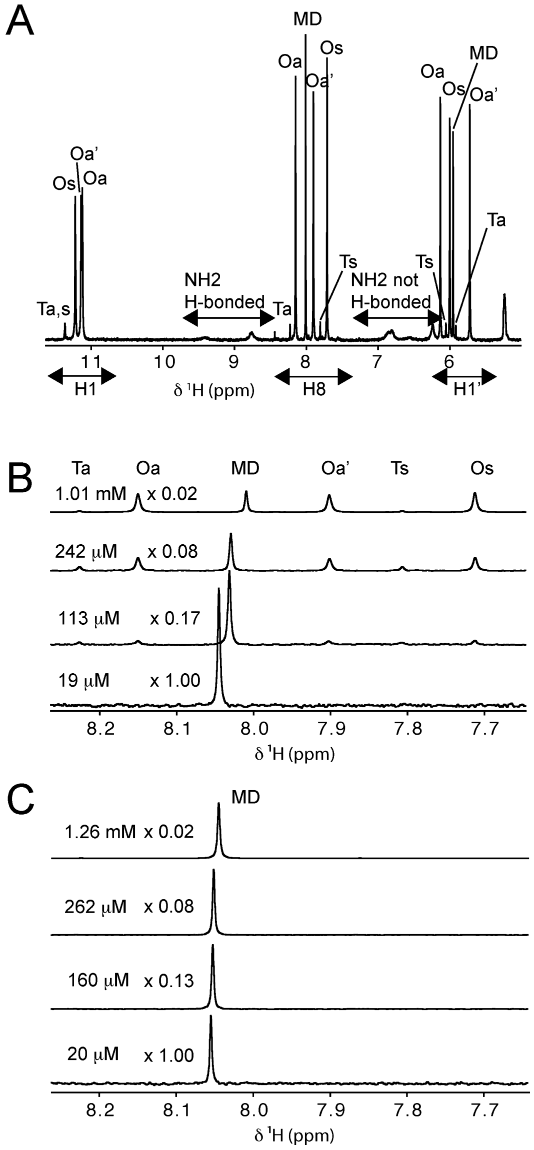


Figure 2. 1D ^{1}H NMR spectra of c-di-GMP. (A) H1, H8, and H1′ resonances of 1.0 mM c-di-GMP dissolved in buffer A. Resonance assignments for tetramers (Ta, Ts) and octamers (Oa, Oa′, Os) were taken from Zhang.[21] MD resonances represent the population-weighted mean chemical shift of rapidly exchanging monomers and dimers. (B) H8 region of the spectrum at various c-di-GMP concentrations in buffer A. Indicated c-di-GMP concentrations were obtained from the sum of the assigned H8 c-di-GMP peak intensities (see text). (C) Same as (B) but c-di-GMP was dissolved in buffer C.

c-di-GMP in TRIS Buffer. To study c-di-GMP oligomerization behavior also for conditions routinely used in enzymatic assays, the experiments were repeated in buffer C (5 mM TRIS/HCl, pH 8) containing no metal cations. Figure 2C shows the H8 part of the c-di-GMP ^{1}H spectrum at various concentrations. In contrast to buffer A (Figure 2B), only a single H8 resonance is detected in the concentration range from 20 μM to 1.26 mM and even at 10 mM c-di-GMP (not shown). Also, none of the other c-di-GMP resonances show any sign of heterogeneity. In addition, and in contrast to the tetrameric and octameric forms in the presence of cations, no imino H1 resonances are observed in buffer C even at high concentrations. This again indicates the absence of a stable H-bond and rapid exchange with the solvent. The resonances are very similar to the MD resonances in buffers A and B and also shift in a concentration-dependent manner. Hence, they were attributed to the MD species, which is in rapid exchange between monomer and dimer. We also tested the effect of phosphate on oligomerization. A solution of 1 mM c-di-GMP in 5 mM TRIS/HCl, 150 mM KCl, pH 8, devoid of phosphate showed oligomer (T, O) resonances very similar to those observed for c-di-GMP dissolved in buffers A and B (Figure S10). Thus, the lack of phosphate is not responsible for the absence of oligomers in buffer C. To assay the effect of different monovalent cations, further samples of 1 mM c-di-GMP dissolved in 5 mM TRIS/HCl were prepared in the presence of large concentrations (≥150 mM) of either KCl, NaCl, or LiCl. Only for KCl, oligomer formation was observed (Figure S10). Thus, Li^+ and Na^+ seem to have a much lower potential to induce oligomerization of c-di-GMP at low millimolar concentrations, although for higher concentrations some coordination of Li^+ and Na^+ has been reported.[20,21] Altogether, this indicates that for concentrations up to at least 1 mM and in the absence of the tested metal ions, c-di-GMP does not form oligomers beyond the dimer. This is consistent with the observation that all described G-quartet structures are stabilized by the coordination of monovalent cations.[20,21,33−35]

c-di-GMP Monomer/Dimer Equilibrium. To determine equilibrium constants and exchange rates of the different oligomeric c-di-GMP forms, the frequency positions and resonance intensities of the well-isolated H8 proton were analyzed in a quantitative manner. Figure 3A,B (left side) shows the chemical shift of the MD resonance as function of the total c-di-GMP concentration for buffers A and B, respectively. For fast monomer/dimer exchange, the shift would be expected to show a sigmoidal dependence on the concentration. However, very large scatter and multiphasic behavior were observed. This was traced back to the very slow exchange kinetics (see below) between the MD species and the higher oligomeric forms, which did not reach equilibrium after dilution from stock solutions even after extended periods of several days at room temperature.

To determine the total MD concentration in the absence of an overall equilibrium with the higher oligomers, the integral of the MD H8 resonance was used as a measure. This intensity was calibrated to an absolute concentration by comparison to the integral of the H8 resonance of GTP, determined in a separate experiment with the GTP concentration defined by UV absorbance. Figure 3A,B (right side) shows the H8 chemical shift as a function of this calibrated MD concentration. It is obvious that there is much less scatter and that the dependence becomes monophasic.

At low (<10 μM) total c-di-GMP concentration, the chemical shift value of the H8 MD resonance converges to the chemical shift of the free monomer at 8.050 and 8.054 ppm in buffers A and B, respectively. With increasing concentration the resonance shifts upfield, i.e., toward the chemical shift of the dimer. However, within the investigated concentration range of up to 6 mM, the shift does not converge to a constant, and hence the dimeric state is not fully reached. Apparently, this is due to the onset of higher oligomer formation above about 20−100 μM total c-di-GMP concentration (depending on buffer conditions), which diminishes the total available MD concentration. Fitting of the chemical shift dependence to a simple monomer/dimer equilibrium yielded dimer dissociation

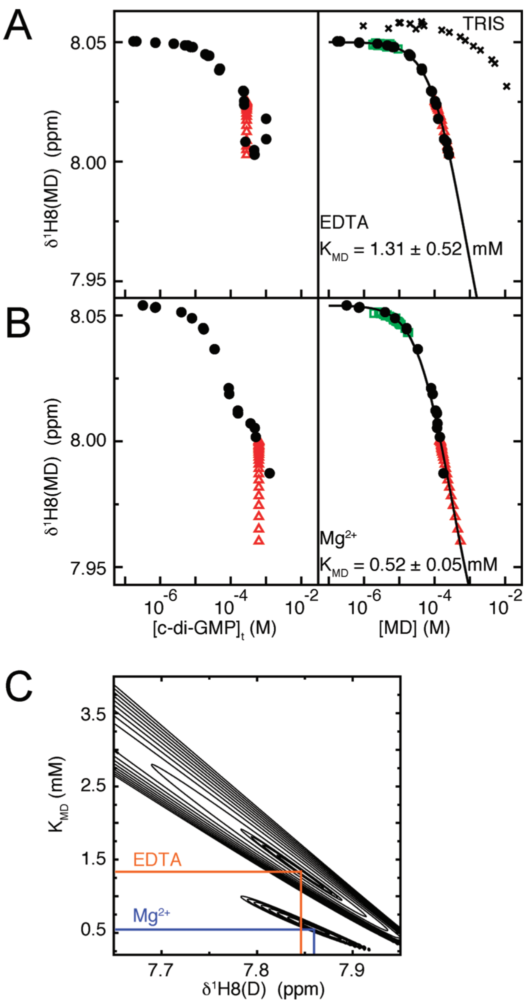


Figure 3. Concentration dependence of the c-di-GMP monomer/dimer equilibrium as evidenced from the chemical shift of the MD H8 resonance. (A) H8 chemical shift of c-di-GMP in buffer A as function of total c-di-GMP concentration $[\text{c-di-GMP}]_t$ (left) and combined monomer/dimer concentration [MD] as determined from the H8 peak intensity (right). (B) same as (A) but spectra were recorded in buffer B. Filled black circles in (A) and (B) represent data from a titration experiment, while red triangles and green squares indicate data from oligomer reassociation (Figure 4) and dilution experiments (Figure 5), respectively. Black crosses in panel A (right) depict data from a titration of c-di-GMP in buffer C. The solid lines in (A) and (B) represent fits to a two-state model according to eq 1. (C) Mean-square deviation χ^2 of fitted and experimental shifts according to eq 4 as a function of the fit parameters K_{MD}, δ_D(H8) for buffers A (EDTA) and B (Mg^{2+}).

constants of 1.31 ± 0.52 and 0.52 ± 0.05 mM in buffers A and B, respectively. Since the dimer chemical shift δ_D could not be determined experimentally, the accuracy of the resulting dissociation constant is limited, as the fit parameters K_{MD} and δ_D are strongly coupled. This is demonstrated by the elongated minima in a $\chi^2(K_{MD},\delta_D)$ plot (Figure 3C). Despite this uncertainty, the distinct minima in the χ^2 surface for buffers A and B indicate that the difference in K_{MD} is significant. Hence, the dimer is slightly more stable in the presence of magnesium. The free energy difference ΔG derived from the dissociation constants (Table 1) corresponds to 16 and 19 kJ/mol in buffers A and B, respectively. A very similar value was

Table 1. Rates and Equilibrium Constants of c-di-GMP Oligomerization

	2M ⇄ D
buffer	K_{MD} [10^{-3}M]^a
A (EDTA)	1.3 ± 0.5
B (Mg^{2+})	0.52 ± 0.01
C (TRIS)	>10

	4M ⇄ T		
	k_{TM} [10^5M^{-3} s^{-1}]	k_{MT} [10^{-6}s^{-1}]	K_{MT} [10^{-12}M^3]
A (EDTA)^b	5.6 ± 0.9	4.2 ± 0.3	6.8 ± 0.5
A (EDTA)^c	3.8	2.6	6.8^d

	T + 4M ⇄ O		
	k_{OT} [10^{12} M^{-4} s^{-1}]	k_{TO} [10^{-5}s^{-1}]	K_{TO} [10^{-17}M^4]
A (EDTA)^b	2.4 ± 0.2	5.4 ± 0.3	2.1 ± 0.1
A (EDTA)^c	3.4	7.1	2.1^d

[a] Equilibrium constants determined from MD ^{1}H8 chemical shift and peak intensities (Figure 3). [b] M ⇄ T and T ⇄ O reaction parameters derived from the fit of the kinetic model (Figure 6) to the oligomer reassociation experiment in buffer A (Figure 4). [c] M ⇄ T and T ⇄ O reaction parameters derived from the fit of the kinetic model (Figure 6) to the oligomer dissociation experiment in buffer A (Figure 5). [d] Parameter kept fixed during fit.

obtained in a recent molecular dynamics simulation (21 kJ/mol).[14]

In buffer C (TRIS), no resonances other than those belonging to the MD species are observed. Only at concentrations above about 1 mM, the H8 resonance starts to shift significantly (Figures 2C and 3A, crosses) due to the accumulation of dimer, which is in fast exchange with the monomer. However, even at 10 mM concentration, the shift change is not very pronounced, and no convergence to the dimer is achieved. Thus, under these conditions, K_{MD} is significantly higher than 10 mM. This is consistent with a recent report that c-di-GMP is monomeric at 1.0 mM concentration in the absence of metal ions.[19] However, in the latter study the possibility of fast chemical exchange had not been considered.

Under all our buffer conditions, very similar equilibrium constants were obtained when H1′ instead of H8 proton resonances were used for the quantitative evaluation (Figure S7, Table S8).

Higher c-di-GMP Oligomers. Consistent with the previously observed metal-ion-dependent polymorphism[21] at millimolar c-di-GMP concentrations, several sets of resonances become visible at concentrations above about 100 μM (Figure 2A,B). Two sets of these resonances were previously attributed to two separate tetramers of two parallel, but not stacked, G-quartets, each (Figure 1C) with the nucleotides in either all-*syn*

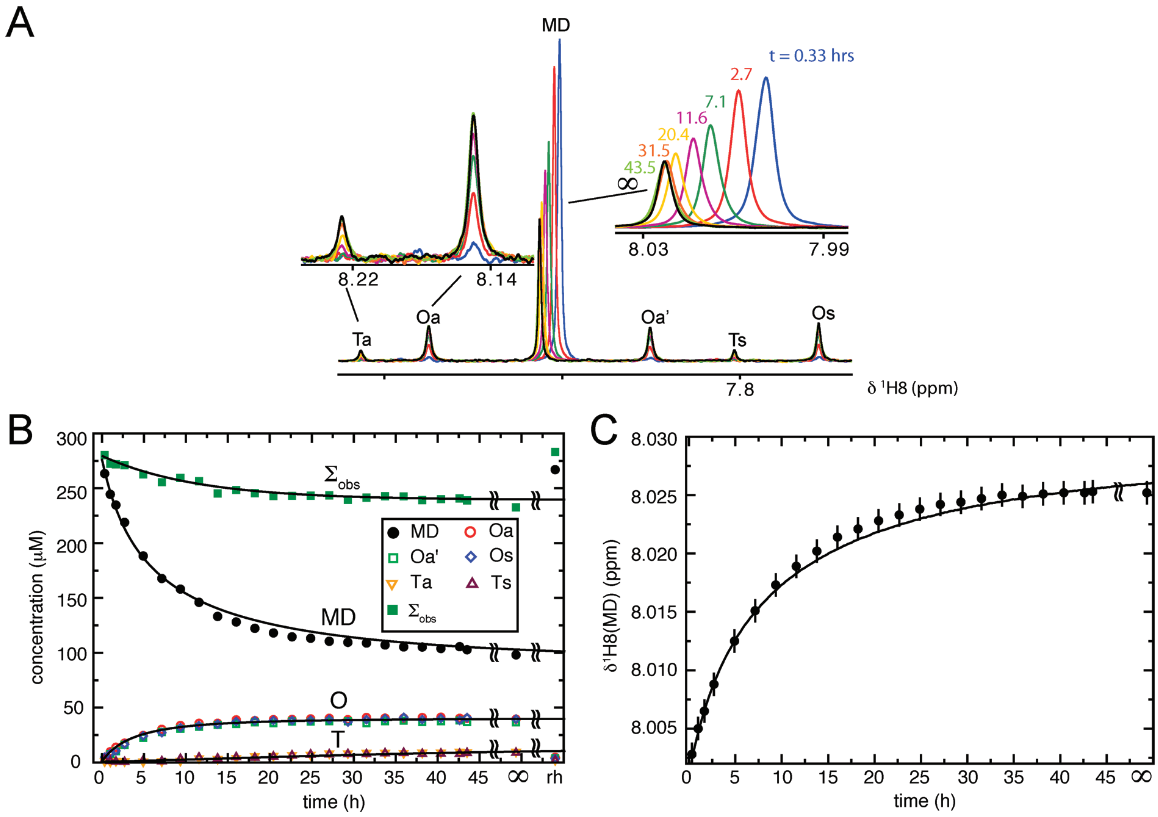


Figure 4. C-di-GMP oligomer reassociation after heat dissociation. (A) Time series of ^{1}H NMR spectra of a 280 μM c-di-GMP sample in buffer A at 24 °C after heat treatment at 60 °C for 1 h. Insets show enlargements of some of the peaks. (B) H8 proton concentrations of the various oligomeric species (symbols) obtained from the corresponding 1D NMR resonance intensities in (A) as a function of time after heating. Continuous lines are fits according to the kinetic model described in the text. An additional point acquired after reheating of the sample to 60 °C for 1.5 h is labeled as “rh”. (C) Chemical shift values of the H8 MD resonance (symbols) as a function of time after heating. The continuous line indicates the fit according to the kinetic model shown in Figure 6. For kinetic parameters, see Table 1.

(Ts) or all-*anti* (Ta) conformation based on H1′ to H8 NOE crosspeaks.[21] Throughout the examined concentration range and also during the two independent kinetic experiments (see below, Figures 4 and 5), the intensities of these resonances remained virtually identical between the two sets (Figure S2A). Virtually equal intensities were also observed during titration and kinetic experiments among the three sets of resonances previously attributed to several octamer structures (Figures 2A,B, 4, 5). On the basis of NOE data, they were identified as all-*syn* (Os) and all-*anti* (Oa/Oa′) octamer conformations, respectively.[21] The very similar intensities for the tetramer and octamer subspecies under all our experimental conditions may indicate subconformations of the same tetramer and octamer complexes rather than completely distinct molecular complexes. In contrast to our observations, unequal intensities of oligomer resonances (Ts vs Ta and Os vs Oa/Oa′) were observed in a recent study on thiophosphate analogues of c-di-GMP under different buffer conditions.[36] In the absence of fully determined tetramer and octamer structures, this issue is unresolved. Besides the tetramer and octamer resonances, additional sets of peaks appear at concentrations above 3 mM (Figure S3). Very likely, these correspond to even higher oligomers. In contrast, at very low concentrations (<∼20 μM) and under all buffer conditions, only the set of MD peaks is observed, shifted almost completely to the monomeric form.

Oligomer Reassociation after Heat Dissociation. The higher oligomers are heat-sensitive and dissociate at an elevated temperature of 60 °C (Figure S4). This provides a convenient means to study the association kinetics of oligomers in a temperature-jump experiment. For this, a 280 μM c-di-GMP sample dissolved in buffer A was heated in a water bath to 60 °C for 1 h and subsequently inserted into the magnet at 24 °C. Figure 4A shows the following slow buildup of the tetramer and octamer peaks in the 1D NMR spectra. At the same time, the MD peak decreases in intensity and shifts downfield, i.e., toward the monomer. The equilibrium is reached only after more than 24 h (Figure 4B,C). Notably, formation of octamers is considerably faster than the formation of tetramers.

Interestingly, the total sum of the H8 peak integrals of the identified species, $\sum_{obs}$, decreased by about 10% over several days and then remained constant for over 6 months (Figure 4B). No precipitation was observed. The effect was not due to an underestimation of the higher oligomer peak intensities resulting from their larger T_1 times, since an extended recovery delay of 62 s (instead of 3 s) yielded identical results. Furthermore, the effect could be reversed by heating to 60 °C (Figure 4B). This is consistent with the reversible formation of even higher oligomeric species, which were below the detection limit of the NMR spectra.

In addition, we checked the changes of oligomer concentrations upon heating an equilibrated sample (280 μM, buffer A) from 297 to 308 K. Consistent with the observation of oligomer dissociation at higher temperatures, the octamer intensities decrease, whereas tetramer intensities first increase and then decrease. The MD resonance gains intensity and shifts toward the dimer, in agreement with the dissociation of the higher oligomeric species. Even after more than 2 h, equilibrium was not reached after the temperature jump. These results indicate that, at higher physiological temperatures in living systems, such as 310 K, c-di-GMP oligomers are even less abundant than at room temperature.

Oligomer Dissociation upon Dilution. Often dilutions from high-concentration stock solutions are used for *in vitro* c-di-GMP enzymatic or biological assays, and it is tacitly assumed that equilibrium is attained immediately. In order to monitor the kinetics of oligomer dissociation under such conditions, a 25 μM c-di-GMP sample in buffer A was prepared by dilution from an equilibrated 1.25 mM stock solution (297 K), and a series of 1D NMR spectra was acquired as a function of time after dilution (Figure 5A,B). As expected, after dilution the MD peak intensity increases and its position shifts toward the dimer, indicative of an increase in total MD concentration and relative dimer population. However, the increase in MD concentration is rather slow and only saturates after about 100 h. Simultaneously, the octamer population decreases continuously, while the tetramer population first increases and then after a period of about 12 h decreases. Several days later, both tetramer and octamer peaks have disappeared completely. This kinetic behavior suggests that tetramers are intermediates of the octamer dissociation reaction. Again, as for the temperature-jump experiment, the total sum of identified peak intensities $\sum_{obs}$ does not remain constant, but in this case increases slightly over time. This is again consistent with invisible higher aggregates at higher concentration that dissociate into observable species upon dilution.

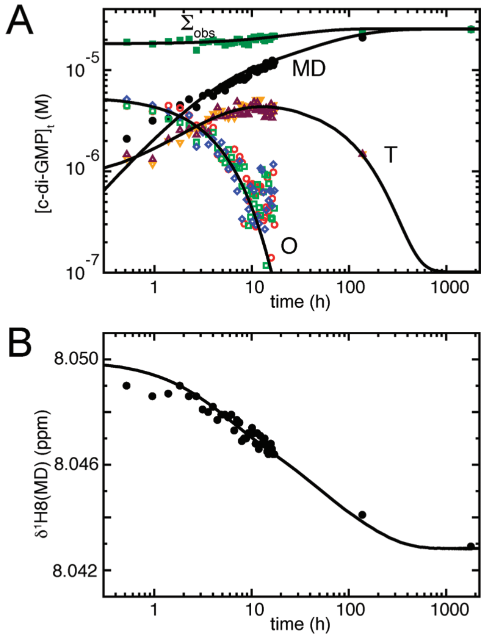


Figure 5. C-di-GMP oligomer dissociation upon dilution. A times series of ^{1}H NMR spectra was recorded on a sample of 25 μM c-di-GMP after dilution from a 1.25 mM stock solution in buffer A. (A) Data points represent experimentally determined concentrations of various c-di-GMP species as a function of time (symbols as in Figure 4). The continuous lines represent the fit according to the kinetic model. (B) Chemical shift of the H8 MD resonance (symbols) with fitted values (continuous line). Experimental chemical shift errors are not larger than 1 ppb.

The slow kinetics upon dilution at ambient temperatures constitutes a problem for biological assays, if care is not taken to reach equilibrium. Procedures for annealing have been proposed, such as a heating−cooling cycle of 95 °C (5 min) to room temperature (15 min) followed by storage in the refrigerator at 4 °C (12 h).[24] Since oligomers are stabilized at low temperatures, we rather propose to keep the sample at the measurement temperature after a suitable high-temperature annealing step. The exact duration and temperature may depend on the buffer conditions and c-di-GMP concentration, and equilibration should be assessed experimentally, e.g., by NMR or UV (see below). For a typical situation, we found that incubation at 60 °C (buffer A) of a 25 μM sample dissolved from a 1.3 mM stock solution dissociates 85% of the higher oligomers to monomers and dimers within 1 h, and any higher oligomers are below the detection limit after 2 h (data not shown).

Kinetic Model of c-di-GMP Self-Association. The experimental results can be integrated into a quantitative framework with predictive power by a simple kinetic model of c-di-GMP oligomerization (Figure 6). As shown above, monomers and dimers are rapidly interconverting. Very likely, the dimer is composed of self-intercalated monomers with stacked bases, as observed in c-di-GMP crystals and in complex with proteins (Figure 1B). Assembly of the hollow-formed tetramer with two nonstacked G-quartets (Figure 1C) would proceed via sequential alignment of monomers in an overall fourth-order reaction. The intercalated dimer is incompatible with such a proposed tetramer structure, since the intercalation would block G-quartet formation. For the assembly of the putative intercalated octamer (Figure 1D), we assume that monomers have to insert sequentially into a preformed tetramer. Note that the intercalated dimer is also not directly compatible with the intercalated octamer structure, since the relative orientation of the monomer subunits changes by about 90° from the dimer to the octamer. Thus, the intercalated dimer may not be directly involved in the formation of the proposed tetramers and octamers. For these reasons, our kinetic model (Figure 6) involves only three reactions ($M \rightleftarrows D$, $M \rightleftarrows T$, $T \rightleftarrows O$), each with a forward and a backward rate (k_{BA} and k_{AB}, respectively) from which the corresponding dissociation constants $K_{AB} = k_{AB}/k_{BA}$ can be derived. Fitting the experimental data to the kinetic model yielded the kinetic parameters for the oligomer equilibria. Since no experimental kinetic data exist for the fast monomer/dimer exchange, the $M \rightleftarrows D$ reaction was treated in the fit as in instant equilibrium according to the dimer dissociation constant K_{MD} determined from the concentration-dependent shift of the MD peak position (see above). The results of the fit to the time-dependent oligomer concentrations of the temperature-jump reassociation experiment are shown in Figure 4B. Taking into account the complication from the decrease in total identified species concentration, $\sum_{\text{obs}}$, the fit of the data is quite satisfactory. Further confidence in the kinetic model and the kinetic parameters was obtained from an equally good fit to the dilution experiment data (Figure 5), which yielded very similar rate constants.

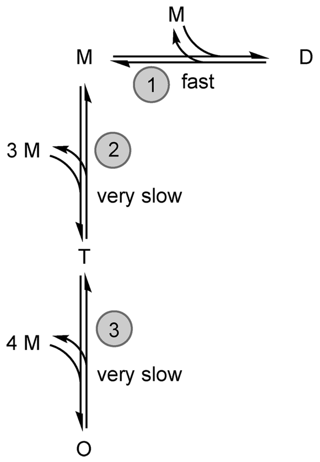


Figure 6. Kinetic scheme of c-di-GMP oligomer formation. M, monomer; D, dimer; T, tetramer; O, octamer.

The resulting kinetic parameters and equilibrium constants are listed in Table 1. It is apparent that the octamer dissociation rate k_{TO} is below 10^{-4} s^{-1}, i.e., extremely slow. This may be explained by the tight integration of each monomer within the octamer, since each of the two guanine bases of one monomer forms four H-bonds with neighbors in the G-quartets. The same is true for tetramer dissociation, for which an even slower dissociation rate ($k_{MT} < 10^{-5}$ s^{-1}) is found.

With a quantitative model of c-di-GMP oligomerization established that describes the observed data, the concentration of the various species can be calculated as a function of total c-di-GMP concentration. Figure 7 shows that, up to a concentration of about 10 μM, virtually only monomers are present. On a further increase of concentration, dimers start to build up, but only to about 10% of the total population, before octamer formation sets in at about 100 μM. Tetramers, as intermediates to octamer formation, are populated only to a low extent.

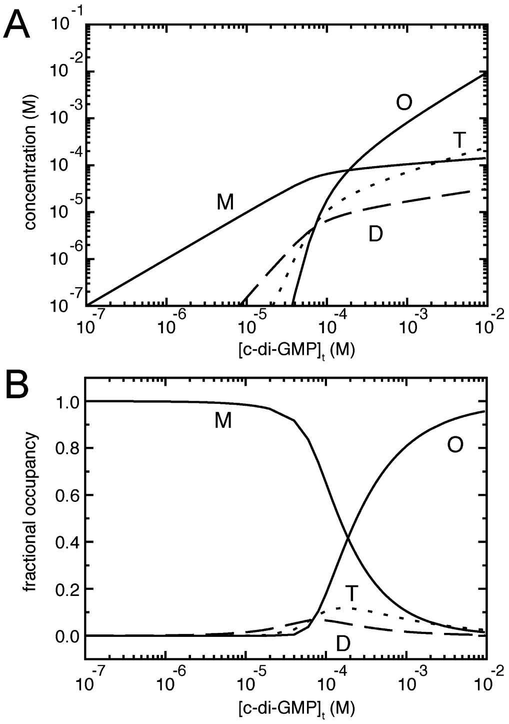


Figure 7. Concentrations of the various oligomeric species of c-di-GMP as a function of total c-di-GMP concentration. (A) Concentrations of the indicated c-di-GMP species (in monomeric units of c-di-GMP) as a function of total c-di-GMP concentration in buffer A. The data were simulated according to the kinetic model shown in Figure 6 with parameters of Table 1. (B) Fractional occupancy of the different species as a function of total c-di-GMP concentration in buffer A.

It has also been reported that aromatic compounds may induce oligomer formation.[6,24,25] This conclusion was obtained on the basis of changes in CD, fluorescence, and absorbance spectra. However, no NMR analysis was carried out. To test for such an induced quadruplex formation, we have repeated some of our measurements in the presence of acriflavin/proflavin (Figure S11), which were also used by Sintim and colleagues.[25] After addition of the dye (16 μM), indeed a significant decrease in c-di-GMP (25 μM) MD intensity (H8, H1′) is observed. However, no new signals can be detected that would match the described dimer, tetramer or octamer resonances. In particular, also no H1 resonances are detected that would indicate H-bond formation. We interpret this observation as a possible stacking reaction between dye and c-di-GMP that, however, does not lead to the canonical H-bonded c-di-GMP dimers, tetramers, or octamers.

General considerations[1] and measurements[2] indicate that the overall concentration of c-di-GMP in the bacterial cell is in the 50 nM to a few μM range. Local c-di-GMP pools with higher concentrations have been discussed[1,37] but so far remain hypothetical. Under our buffer conditions, such low concentrations would imply a monomeric form of c-di-GMP. Clearly, intracellular stacking complexes with aromatic compounds cannot be excluded. However, in the absence of such heteromeric stacking reactions, c-di-GMP should be monomeric under *in vivo* conditions. In this respect it is important to note that c-di-GMP adopts a dimeric, self-intercalated form in complex with several proteins, such as diguanylate cyclase,[9,10] a response regulator,[13] and the PilZ receptor.[12] It may be possible that *in vivo* such complexes are formed by the consecutive binding of two monomeric ligands, i.e., that dimerization occurs only on the protein. The conserved intercalation of aromatic and arginine amino acids into the c-di-GMP dimer in the protein complexes[12] may act as a template for the formation of such dimers at low c-di-GMP concentrations.

Assessment of c-di-GMP Association from UV Spectra. Concentration- and temperature-dependent changes in the absorptivity of c-di-GMP have been reported and attributed to the hypochromic effect caused by base stacking.[38] For convenient absorption measurement ($OD_{253} < 1$), c-di-GMP is usually diluted from millimolar stock solutions to few tens of micromolar concentration. However, due to the very slow kinetics of oligomer dissociation this procedure may leave the solution in an undefined state with varying absorbance properties. As described above, the problem can be overcome by annealing. Once the higher oligomers are dissociated, base stacking in the dimer will only minimally influence the absorption measurement below about 50 μM, since the monomer/dimer dissociation constants are above 0.5 mM, even in the presence of metal ions. No such problems from oligomerization are expected in the absence of metal ions (e.g., TRIS, buffer C), where no higher oligomers are present up to a concentration of at least 10 mM and where the monomer/dimer dissociation constant is even larger. Indeed, Figure S5 demonstrates that Lambert−Beer's law is obeyed perfectly in buffer C, and a molar extinction coefficient was determined at the absorption maximum of 253 nm as ε_{253}(c-di-GMP) = $(28.6 \pm 1.0) \times 10^3$ M^{-1} cm^{-1} by a combination of absorption and NMR intensity measurements that were calibrated to a reference GTP sample (see Materials and Methods). The form of the absorption spectrum of c-di-GMP under these conditions is virtually identical to that of GTP, and not surprisingly ε_{253}(c-di-GMP) is very close to twice the value for GTP[32] (13.7×10^3 M^{-1} cm^{-1}), indicating that the two guanine bases in monomeric c-di-GMP are not strongly interacting.

In contrast to the absorption spectrum of monomeric c-di-GMP, the spectra of its oligomeric forms differ from that of GTP (Figure 8). The base stacking causes both a decrease in the amplitude (hypochromicity) and a red shift (bathochromicity) of the guanine spectrum. Figure 8A shows this effect in time-dependent absorption spectra of 25 μM c-di-GMP recorded at 24 °C after dilution into buffer A from a 1.25 mM stock solution. Initially, the spectrum has a main maximum at 256 nm and a second shoulder at 295 nm. With increasing time, the absorption shifts to the blue and increases to converge toward the GTP spectrum with a main maximum at 253 nm and the second shoulder at 280 nm. This behavior can be described quantitatively to a good approximation by a linear superposition of a monomeric (unstacked) and a stacked spectral species, where the time-dependent population of the monomeric form is determined from the intensity and position

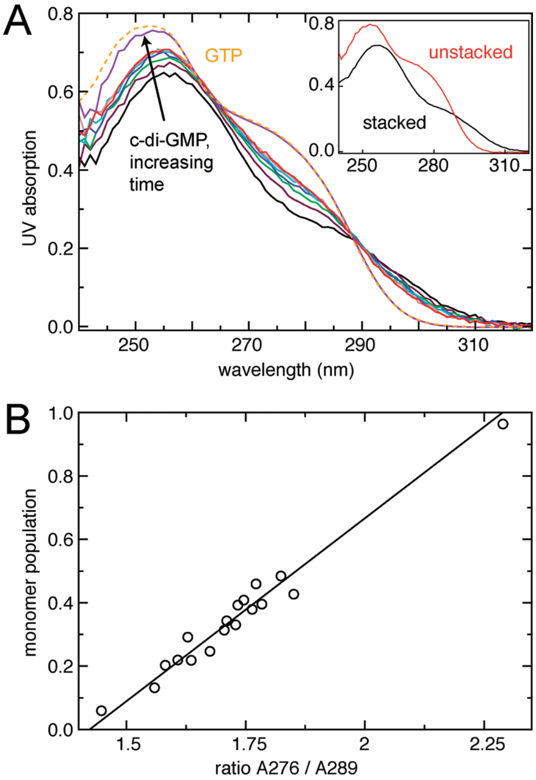


Figure 8. UV analysis of the aggregation state of c-di-GMP. (A) Time course of UV absorption spectra of 25 μM c-di-GMP after dilution into buffer A. Colors present spectra recorded at different times: 10 min (black), 190 min (maroon), 370 min (green), 550 min (blue), 730 min (cyan), 910 min (brown), 970 min (red), and 75 days after dilution (violet). For comparison, the absorption spectrum of GTP (scaled according to its NMR peak intensity relative to that of c-di-GMP) is shown (orange, dashed). The inset shows the deconvoluted spectra for unstacked (black) and stacked (red) species. (B) Population of monomeric c-di-GMP p_{mono} as a function of the UV absorption ratio of 276/289 nm (isosbestic point, 288.8 nm). The monomeric populations were obtained from the NMR dilution experiment (Figure 5). The solid line indicates a linear fit of the form $p_{\text{mono}} = 1.15\,(A_{276}/A_{289}) - 1.64$.

of the H8 resonance in a ^{1}H spectrum recorded on the same sample. A linear fit according to eq 6 yields predicted time-dependent spectra, which are almost indistinguishable from the experimental data (Figure S6). The decomposed monomeric (unstacked) and stacked spectra according to the fit are shown in Figure 8A. The spectrum of the unstacked form is identical to the GTP spectrum with a maximum at 253 nm, whereas the spectrum of the stacked form has an about 16% lower maximal extinction at 256 nm. The isosbestic point is located at 288.8 nm. For accurate determination of the total concentration in the presence of stacking, it is thus advisible to use the absorbance at the isosbestic point.

Figure 8A shows that the maximal difference between monomeric and stacked species occurs at 276 nm. The ratio of the absorbance at this wavelength to the absorbance at the isosbestic point (A_{276}/A_{289}) may be taken as a measure for the relative monomer population p_{mono}. Indeed, an excellent linear correlation is observed (Figure 8B). Thus the absorbance ratio A_{276}/A_{289} provides an easy measure of the dissociation state.

CONCLUSION

Due to the propensity of base-stacking and G-quartet formation, c-di-GMP can display a rich polymorphism of higher oligomeric forms in the presence of metal ions. Besides monomers and dimers, at least two forms of guanosines in putative tetramer and three forms in putative octamer structures are distinguishable in ^{1}H proton NMR spectra at low millimolar concentrations. In this study, we have determined kinetic parameters and equilibrium constants of the interchange between these different forms.

Dissociation constants of the dimer are on the order of 1 mM in the presence of metal ions but significantly larger in their absence. Irrespective of the presence of metal ions, c-di-GMP exchanges rapidly between monomeric and dimeric forms on the chemical shift time scale of milliseconds. Higher oligomer formation occurs only in the presence of monovalent (particularly K^+) metal ions. The two forms of guanosines in tetramers and the three forms in octamers are always found in equal concentration presumably because these oligomers contain distinct guanosine subconformations within the same molecule. In contrast to the dimerization reaction, the kinetics of higher oligomer formation and dissociation is extremely slow and can require weeks to reach equilibrium. The kinetics of monomer, dimer, tetramer, and octamer exchange can be described to good quantitative agreement by a simplified model where the tetramer is an intermediate of octamer formation.

The extremely slow dissociation kinetics of the higher oligomers may lead to artifactual low concentrations of the active monomeric form when c-di-GMP is diluted from concentrated stock solutions in biological assays. The problem can be overcome by an annealing step, e.g., incubation at 60 °C for more than 2 h. Since the base stacking associated with oligomer formation causes a red shift of the UV spectra, the proper monomer content can be verified easily from the UV absorption ratio (A_{276}/A_{289}).

Under all tested buffer conditions, c-di-GMP was monomeric at low micromolar concentrations, in particular also at very high potassium concentrations that favor higher homooligomer formation. Although it cannot be excluded that aromatic compounds may serve as templates for other types of higher oligomers in the complex environment of a biological cell, it is unlikely that homooligomeric structures in free form are populated to a significant extent at the low cellular c-di-GMP concentrations.

ASSOCIATED CONTENT

Supporting Information

1D NMR spectra of c-di-GMP under various buffer and temperature conditions, correlation of UV data with c-di-GMP NMR concentrations in TRIS buffer, fit of the time-dependent UV spectra by two-state model, evaluation of NMR H1′ data, and NMR spectra of GTP. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

stephan.grzesiek@unibas.ch; tilman.schirmer@unibas.ch

Present Address

§Harvard Medical School, 240 Longwood Ave., Boston, MA 02115

ACKNOWLEDGMENTS

We thank U. Jenal, D. Häussinger, H.-J. Sass, and D. Samoray for helpful discussions. This work was supported by SNF grants 31-132857 to S.G. and 31-105587 to T.S.

REFERENCES

(1) Hengge, R. *Nat. Rev. Microbiol.* **2009**, *7*, 263.

(2) Simm, R.; Morr, M.; Rerriminghorst, U.; Andersson, M.; Romling, U. *Anal. Biochem.* **2009**, *386*, 53.

(3) Schirmer, T.; Jenal, U. *Nat. Rev. Microbiol.* **2009**, *7*, 724.

(4) Kim, D.; Hunt, J. F.; Schirmer, T. In *The Second Messenger Cyclic Di-GMP*; Wolfe, A. J., Visick, K. L., Eds.; John Wiley & Sons: New York, 2010; p 76.

(5) Egli, M.; Gessner, R. V.; Williams, L. D.; Quigley, G. J.; Vandermarel, G. A.; Vanboom, J. H.; Rich, A.; Frederick, C. A. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 3235.

(6) Liaw, Y. C.; Gao, Y. G.; Robinson, H.; Sheldrick, G. M.; Sliedregt, L.; Vandermarel, G. A.; Vanboom, J. H.; Wang, A. H. J. *FEBS Lett.* **1990**, *264*, 223.

(7) Guan, Y.; Gao, Y. G.; Liaw, Y. C.; Robinson, H.; Wang, A. H. J. *J. Biomol. Struct. Dyn.* **1993**, *11*, 253.

(8) Blommers, M. J. J.; Haasnoot, C. A. G.; Walters, J.; Vandermarel, G. A.; Vanboom, J. H.; Hilbers, C. W. *Biochemistry* **1988**, *27*, 8361.

(9) Chan, C.; Paul, R.; Samoray, D.; Amiot, N. C.; Giese, B.; Jenal, U.; Schirmer, T. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17084.

(10) De, N.; Pirruccello, M.; Krasteva, P. V.; Bae, N.; Raghavan, R. V.; Sondermann, H. *PLoS Biol.* **2008**, *6*, 601.

(11) Ko, J.; Ryu, K. S.; Kim, H.; Shin, J. S.; Lee, J. O.; Cheong, C.; Choi, B. S. *J. Mol. Biol.* **2010**, *398*, 97.

(12) Habazettl, J.; Allan, M. G.; Jenal, U.; Grzesiek, S. *J. Biol. Chem.* **2011**, *286*, 14304.

(13) Krasteva, P. V.; Fong, J. C. N.; Shikuma, N. J.; Beyhan, S.; Navarro, M.; Yildiz, F. H.; Sondermann, H. *Science* **2010**, *327*, 866.

(14) Zhang, L. X.; Meuwly, M. *ChemPhysChem* **2011**, *12*, 295.

(15) Minasov, G.; Padavattan, S.; Shuvalova, L.; Brunzelle, J. S.; Miller, D. J.; Basle, A.; Massa, C.; Collart, F. R.; Schirmer, T.; Anderson, W. F. *J. Biol. Chem.* **2009**, *284*, 13174.

(16) Barends, T. R. M.; Hartmann, E.; Griese, J. J.; Beitlich, T.; Kirienko, N. V.; Ryjenkov, D. A.; Reinstein, J.; Shoeman, R. L.; Gomelsky, M.; Schlichting, I. *Nature* **2009**, *459*, 1015.

(17) Navarro, M. V. A. S.; De, N.; Bae, N.; Wang, Q.; Sondermann, H. *Structure* **2009**, *17*, 1104.

(18) Benach, J.; Swaminathan, S. S.; Tamayo, R.; Handelman, S. K.; Folta-Stogniew, E.; Ramos, J. E.; Forouhar, F.; Neely, H.; Seetharaman, J.; Camilli, A.; Hunt, J. F. *EMBO J.* **2007**, *26*, 5153.

(19) Wang, J.; Zhou, J.; Donaldson, G. P.; Nakayama, S.; Yan, L.; Lam, Y.-f.; Lee, V. T.; Sintim, H. O. *J. Am. Chem. Soc.* **2011**, *133*, 9320.

(20) Zhang, Z.; Gaffney, B. L.; Jones, R. A. *J. Am. Chem. Soc.* **2004**, *126*, 16700.

(21) Zhang, Z.; Kim, S.; Gaffney, B. L.; Jones, R. A. *J. Am. Chem. Soc.* **2006**, *128*, 7015.

(22) Mooren, M. M. W.; Wijmenga, S. S.; Vandermarel, G. A.; Vanboom, J. H.; Hilbers, C. W. *Nucleic Acids Res.* **1994**, *22*, 2658.

(23) Phillips, K.; Dauter, Z.; Murchie, A. I.; Lilley, D. M.; Luisi, B. *J. Mol. Biol.* **1997**, *273*, 171.

(24) Nakayama, S.; Kelsey, I.; Wang, J. X.; Roelofs, K.; Stefane, B.; Luo, Y. L.; Lee, V. T.; Sintim, H. O. *J. Am. Chem. Soc.* **2011**, *133*, 4856.

(25) Nakayama, S.; Kelsey, I.; Wang, J. X.; Sintim, H. O. *Chem. Commun.* **2011**, *47*, 4766.

(26) Lane, A. N.; Chaires, J. B.; Gray, R. D.; Trent, J. O. *Nucleic Acids Res.* **2008**, *36*, 5482.

(27) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. *Acta Crystallogr. D* **1997**, *53*, 240.

(28) Zaehringer, F.; Massa, C.; Schirmer, T. *Appl. Biochem. Biotechnol.* **2011**, *163*, 71.

(29) Sundararaj, S.; Guo, A.; Habibi-Nazhad, B.; Rouani, M.; Stothard, P.; Ellison, M.; Wishart, D. S. *Nucleic Acids Res.* **2004**, *32*, D293.

(30) Hwang, T. L.; Shaka, A. J. *J. Magn. Reson. Ser. A* **1995**, *112*, 275.

(31) Wider, G.; Dreier, L. *J. Am. Chem. Soc.* **2006**, *128*, 2571.

(32) Bock, R. M.; Ling, N. S.; Morell, S. A.; Lipton, S. H. *Arch. Biochem. Biophys.* **1956**, *62*, 253.

(33) Gilbert, D. E.; Feigon, J. *Curr. Opin. Struct. Biol.* **1999**, *9*, 305.

(34) Ida, R.; Wu, G. *J. Am. Chem. Soc.* **2008**, *130*, 3590.

(35) Ida, R.; Kwan, I. C. M.; Wu, G. *Chem. Commun.* **2007**, 795.

(36) Zhao, J. W.; Veliath, E.; Kim, S.; Gaffney, B. L.; Jones, R. A. *Nucleosides Nucleotides Nucleic Acids* **2009**, *28*, 352.

(37) Römling, U.; Gomelsky, M.; Galperin, M. Y. *Mol. Microbiol.* **2005**, *57*, 629.

(38) Danilov, V. I. *FEBS Lett.* **1974**, *47*, 155.

2.3 Structures of C-di-GMP Oligomers

Abstract

In the growing field of c-di-GMP signaling, all protein and RNA c-di-GMP receptors that have been solved so far bind either to monomeric, dimeric or both forms of c-di-GMP. No protein structure in complex with higher c-di-GMP oligomers than dimers has been solved up to date and the existence of such higher oligomers *in vivo* is currently matter of debate. Polymorphism of c-di-GMP[279,280] *in vitro* has been proposed to be based on guanine quartets (G-quartets)[44] due to the salt dependence (potassium), the slow kinetics as well as the similarities of UV and CD spectra compared to known G-quartet structures of DNA[144]. However, despite several proposed structures for c-di-GMP tetramers and octamers, their detailed molecular structures remain elusive. While the formation of oligomers (> dimers) has been confirmed by several techniques at high c-di-GMP concentrations, their existence at physiologically low ($\leq$ 1 μM) concentrations[225] is still unclear. Recently, we have shown that c-di-GMP homo-oligomers are not significantly populated at physiological concentrations in the absence of G-quartet templates[75]. The reported formation of these particular c-di-GMP oligomers at low micromolar concentrations in the presence of small aromatic compounds remains to be proven by a high-resolution method. Even though c-di-GMP NMR resonances were reduced after addition of acriflavin/proflavin, which had been described to act as G-quartet templates and interact with c-di-GMP tetramers[144,177], octamers or even both[178], no appearance of these specific oligomers has been observed by NMR[75]. Without detailed knowledge of the three dimensional structures of these c-di-GMP oligomers, all described results remain qualitative and lack complete understanding at the atomic level. Only with high-resolution structures at hand, the mechanism of interactions between c-di-GMP oligomers and these aromatic compounds that may serve as G-quartet templates can be addressed unambiguously.

Here, NMR data are presented that indicate that the putative octamers, which have been divided into separate all-syn (Os) and all-anti (Oa/Oa') structures[280], are in fact not distinct and rather represent subconformations of a single species. Results of c-di-GMP oligomers dissolved in H_2O as well as D_2O are compared and some strategies to solve their atomic resolution structures by NMR using isotope labeling are proposed.

2.3.1 Introduction

C-di-GMP (other names: cyclic diguanylate, cyclic diguanylic acid, Bis-(3'-5')-cyclic dimeric GMP, cyclic diguanosine-monophosphate) is a ubiquitous bacterial second messenger regulating many cellular processes upon binding to diverse targets[165]. The most important function is probably the control of the switch of bacterial lifestyles from the motile, planktonic to the sessile, biofilm-related lifestyle[94]. C-di-GMP targets several effectors modulating downstream processes, including biosynthesis of exopolysaccharide components leading to biofilm formation[160–162,167,179,193], bacterial motility[76], cell cycle progression[55], synthesis of secondary metabolites[66] and virulence of pathogens[22,119,183].

C-di-GMP is synthesized from two GTP molecules by diguanylate cyclases (DGC) which contain the characteristic GGDEF domain and it is degraded by enzymes with EAL or HD-GYP domains which have phosphodiesterase (PDE) activity. Generally, proteins with either GGDEF or EAL domains, or even both (composite GGDEF-EAL proteins), possess N-terminal sensory domains which often also contain transmembrane helices and thus may sense environmental conditions in the periplasm[94]. The function of composite GGDEF-EAL proteins still remains unknown, since for this protein family, the majority has only either DGC or PDE activity[235]. Few bifunctional composite proteins have been discovered only recently[17,85,134,141,246]. The interplay of the many GGDEF, EAL/HD-GYP and composite proteins in bacteria may be able to generate oscillating c-di-GMP levels[36], which then can bind to target molecules depending on the respective binding strengths. The full c-di-GMP machinery, comprising synthesis, degradation and binding to its multiple receptors allows regulation on the transcriptional, the translational and the post-translational level[165].

In addition to the GGDEF and EAL/HD-GYP domains which are involved in the formation and dissociation of c-di-GMP, several protein domains with c-di-GMP binding properties have been identified[216]. Most diguanylate cyclases contain an inhibitory site (I-site), apart from the catalytical site, enabling feedback-inhibition[32,94,216,264].

In most complexes with macromolecules, c-di-GMP binds as a dimer[165] (Figure 2.4 D) and exhibits a similar intercalated structure as has been reported in the crystal structure of dimeric c-di-GMP in the presence of metal ions[61,144]. Representative examples are c-di-GMP dimers in PilZ domains[86,129,222], in the inhibition site of diguanylate cyclases[33,46] and in a response regulator[132]. An exceptional c-di-GMP dimer conformation in complex with a protein has been discovered recently[276] (Figure 2.4 E), where the two inner guanines stack while the others interact with the protein.

Furthermore, some protein and RNA structures in complex with monomeric c-di-GMP have been solved. These include c-di-GMP-degrading EAL domains which bind to the open form of c-di-GMP[14,166,180] (Figure 2.4 B), and two classes of structurally unrelated c-di-GMP riboswitches, which both were found to bind to the closed form of c-di-GMP monomers[228,229] (Figure 2.4 C). C-di-GMP binding to proteins with unknown structures, which do not contain any of the known binding sites has also been observed and may represent new classes of c-di-GMP effector proteins[35,139,239].

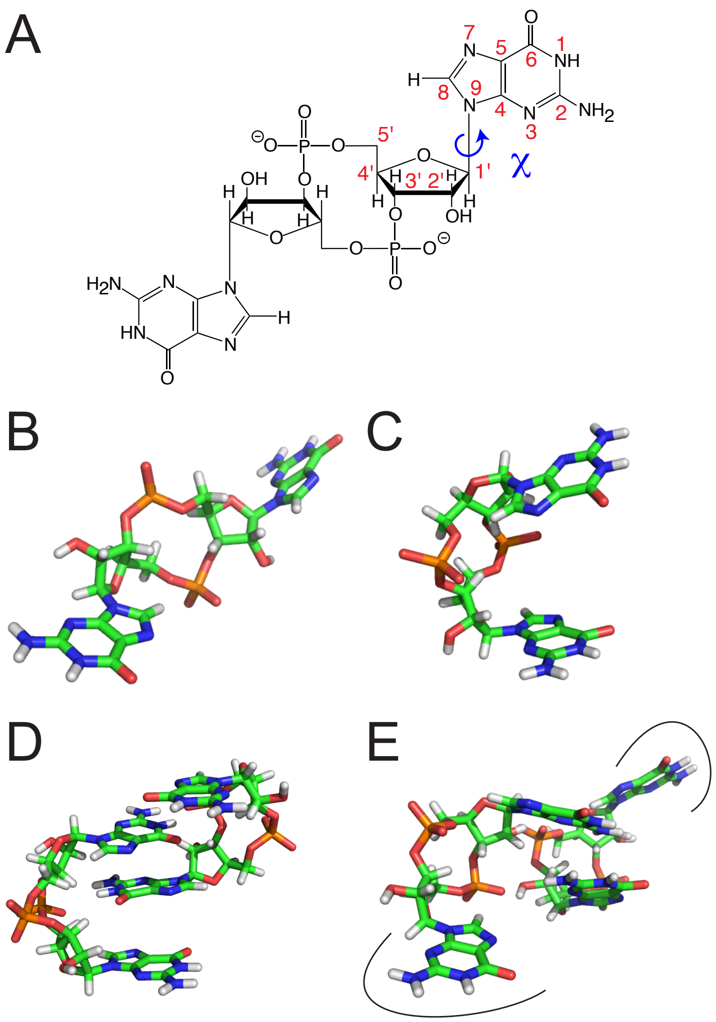


Figure 2.4 Structures of various c-di-GMP forms found in complex with RNA or proteins. **A:** Chemical structure of a c-di-GMP monomer with atom numbering and the dihedral angle χ indicated according to IUPAC recommendations[153]. **B:** An open form c-di-GMP monomer bound to the protein FimX from *P. aeruginosa* (pdb code 3HV8[180]). **C:** Monomeric c-di-GMP in the closed conformation bound to a c-di-GMP-II riboswitch from *C. acetobutylicum* (pdb code 3Q3Z[229]). In the open form of monomers, the χ angles of both moieties deviate more than 50 degrees from the typically observed anti conformation. **D:** The predominantly found form of intercalated dimeric c-di-GMP in complex with the *P. aeruginosa* PilZ domain protein PA4608 (pdb code 2L74[86]), which is very similar to the X-ray structure in the presence of magnesium[61] or cobalt[144]. **E:** A novel form of dimeric c-di-GMP bound to the asymmetric homodimer of the *X. campestris* protein XCC4471 (pdb code 3QYY[276]). The black curves indicate the binding sites of the outer two bases, while only the inner two are stacked. It is remarkable that mostly the anti conformation is observed in complexes. Colors indicate atom types as follows: C (green), H (white), N (blue), O (red), P (orange).

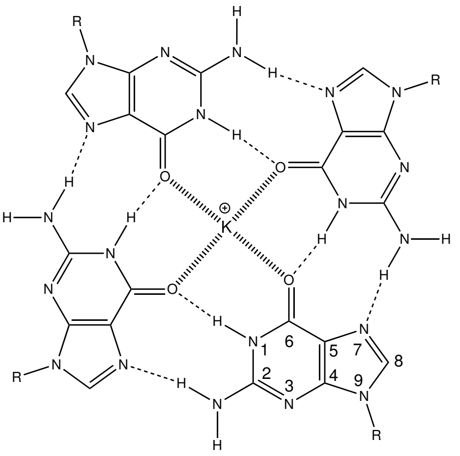


Figure 2.5 Chemical structure of a guanine quartet (G-quartet) in the presence of a potassium ion. R denotes the ribose moiety. Atom numbering is shown for one guanine (bottom, right).

However, no higher c-di-GMP oligomers than dimers have been observed to be bound to protein or RNA.

Liaw *et al.*[144] proposed a schematic structure of a c-di-GMP tetramer, which incorporates an aromatic dye in the central cavity formed by two parallel G-quartets (about 6.8 Å apart), which mimic an intercalator binding site of double helical DNA. This hypothesis was based on similarities of UV spectra of c-di-GMP and a double-strand DNA, each in the presence of 9-aminoacridine. The authors proposed a cage-like structure, composed of head-to-tail arranged guanine bases as observed in an X-ray study on poly(G)[8].

G-quartets are characterized by the circular arrangement of four guanine bases in a plane, where each base forms a total of four hydrogen bonds (H-bonds), two to each adjacent guanine[44] (Figure 2.5). In these hydrogen bonds, NH1 and the $N2H_2$ groups act as H-bond donors and O6 and N7 act as H-bond acceptors. The O6 atom of each of the four guanine moieties coordinates to a central metal ion (mostly sodium or potassium), which resides either in the center of or between adjacent G-quartets.

C-di-GMP oligomers in addition to the known dimers have only been identified recently at high millimolar c-di-GMP concentrations in the presence of an excess of potassium ions[279,280]. These oligomers have been termed tetramers and octamers, and structures have been proposed, each based on multiple parallel G-quartet stacks (Figure 2.6). However, the exact stoichiometry of these c-di-GMP polymorphs based on NMR DOSY data remained arguable. From several NOE experiments, it had been concluded that two forms of tetramers (all-syn, Ts and all-anti, Ta) and two forms of octamers (all-syn, Os and all-anti, Oa/Oa') exist. Syn and anti refer to the χ angle (O4'-C1'-N9-C4, Figure 2.4 A) of the nucleobase relative to the sugar, where syn and anti denote χ angles of about 0 and 180 degrees, respectively[153]. However, the underlying experimental NOE data was inconclusive, since some observable cross-resonances between Os and Oa, had apparently not been noticed (see below).

Models of the proposed all-anti tetramer and all-anti octamer structures[280] are depicted in Figure 2.6. It remains unclear why mixed anti-syn tetramers or octamers (Figure 2.6 C) could not exist as well.

Furthermore, when these octamer structures were proposed, the distinction of Oa and Oa' has been rationalized by separate resonances for inner and outer G-quartets which would not be distinguishable for Os[280]. Based on structural considerations the latter statement remains obscure as it is clear that an all-syn octamer would also contain inner and outer quartets and thus should also show two sets of resonances. Additionally, the relative intensities of the resonances have not been analyzed in detail. Surprisingly, in the published spectra, these intensities are not consistent among the various resonances of the involved species, indicating resonance overlap. In our recent study[75], the intensities of the resonances assigned to the various octamer subspecies were nearly identical and also showed a very similar kinetic behavior upon oligomer formation and dissociation. This has lead to the conclusion that these resonances belong to subconformations of a single species. At low c-di-GMP concentrations (< 1 mM), these specific oligomers have only been found in the presence of potassium while at very low concentrations of about 20 μM c-di-GMP, neither tetramers nor octamers have been observed by NMR[75].

Even more complex polymorphism has been observed for thiophosphate analogs of c-di-GMP[281], while replacement of one bridging O5' oxygen by sulphur prevented the formation of higher aggregates[259]. In contrast to c-di-GMP, no such polymorphism has been reported for related compounds such as cAMP, cGMP, ppGpp (guanosine tetraphosphate), c-tri-GMP[168], or c-di-dAMP, for which both monomers and intercalated dimers have been characterized by NMR[19].

Interactions between c-di-GMP and aromatic dyes have been reported and attributed to intercalation or end-stacking of the dye molecule to G-quartet-based c-di-GMP oligomers[126,144,176–178]. The formation of such oligomers, in the presence of aromatics, has only been inferred from CD, fluorescence or UV data, which are all non-high-resolution techniques. In a recent review of G-quadruplex structures[136], it has been argued that in all investigations of G-quartets, a high-resolution method such as X-ray cristallography or NMR should be used, to definitely prove the formation of G-quartets.

Observations in UV spectroscopy attributed to the presence of G-quartet-based, stacked c-di-GMP oligomers are a red shift (bathochromicity) of the absorption maximum as well as a reduction in UV absorption magnitude (hypochromicity)[279,280]. In our detailed time-dependent study of oligomer dissociation, the isosbestic point has been determined as 289 nm at high (250 mM) potassium concentrations[75] (Figure 2.7). Similar UV spectra showing a virtually identical isosbestic point have been observed in a comparison of UV spectra of the quadruplex formed in the human telomere sequence 5'AGGG(TTAGGG)$_3$ in the presence of sodium at 20 °C and 95 °C[136].

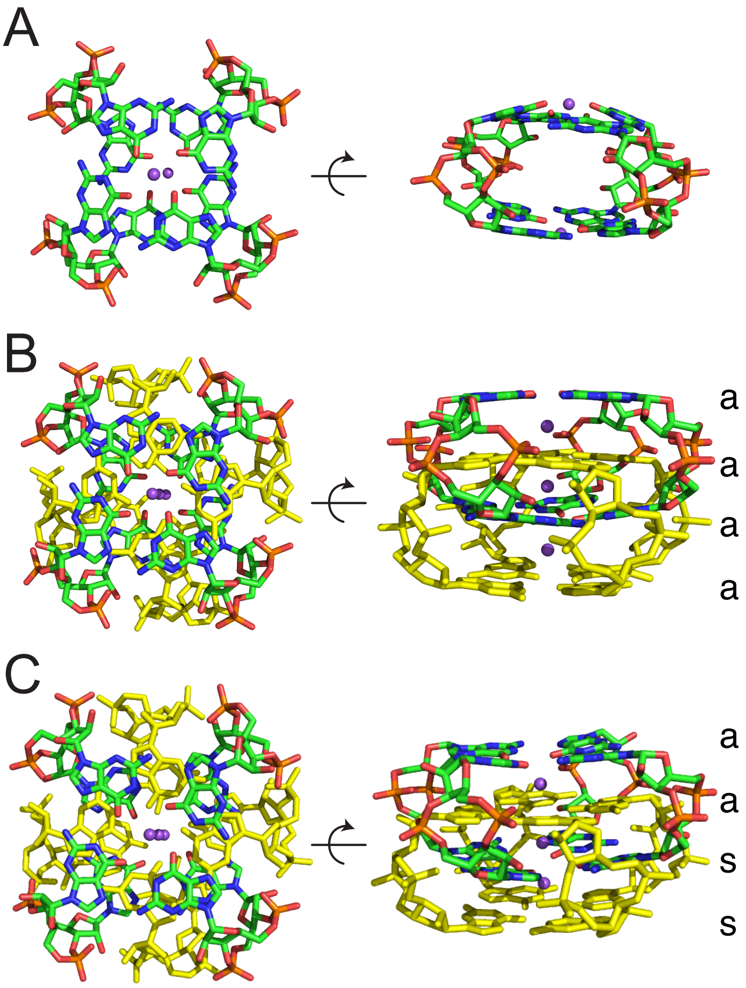


Figure 2.6 Computed models of various proposed c-di-GMP oligomers in the presence of sodium ions (purple). Color coding is identical to Figure 2.4. **A:** Model of the proposed all-anti tetramer[280] with sodium ions located in the center of each G-quartet. **B:** Model of the proposed all-anti octamer[280] with sodium ions located between adjacent G-quartets. **C:** Model of a theoretical mixed anti-syn octamer with sodium ions located between adjacent G-quartets. In the displayed orientation, the C8 atoms are oriented counterclockwise in the upper two and clockwise in the lower two G-quartets in the all-anti (a-a-a-a) octamer, whereas the C8 atoms in the mixed anti-syn (a-a-s-s) octamer G-quartets are all oriented oriented counterclockwise. The annotation a-a-s-s refers to the dihedral angle χ (Figure 2.4), ordered from the top G-quartet (left) to the bottom (right). Note that the hypothetical mixed anti-syn octamer is composed of eight anti-syn c-di-GMP monomers. For the depicted structures, each individual G-quartet is composed of either all-syn or all-anti conformers. A set of four monomers in each c-di-GMP octamer has been colored yellow to visualize the intercalation. Details of the structure modeling are described in Gentner *et al.*, 2012[75].

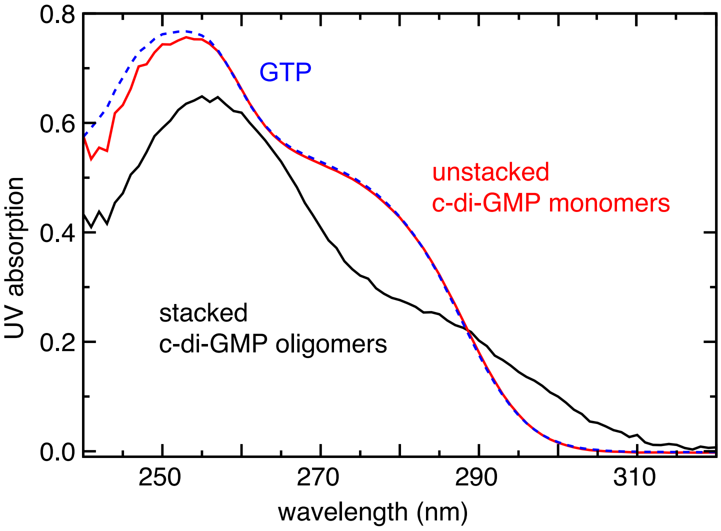


Figure 2.7 UV spectra of 25 μM c-di-GMP in the presence (black) and the absence (red) of oligomers at 20 °C. For comparison, a scaled spectrum of GTP is shown in blue (dashed line). The similarity of the UV spectra of GTP and the oligomer-free c-di-GMP (< 5% dimers) indicates the absence of stacking effects. Figure adapted from[75].

In addition to UV, CD spectroscopy has been used to assess base stacking and G-quartet formation in c-di-GMP, mostly referring to a positive band around 300 nm[126,176,177,279,280]. Such a band has indeed been found for the human telomere quadruplex sequence 5'AGGG(TTAGGG)$_3$ at 20 °C[136]. Even though some *ab initio* calculations are found for CD spectra of proteins and nucleic acid duplexes[231,269], such data is rather sparse for G-quartets[78,107]. Furthermore, the interpretation of optical properties as well as the shape and sign of CD bands is still controversial, since no simple relation between G-quadruplex fold and the CD spectrum could be established so far[43,136].

NMR spectroscopy gives a more detailed insight, since it provides atomic resolution. Several NMR studies of potassium-stabilized G-quartets have been performed elucidating their topologies in telomeres[260–262], in chloroform solution[154] or topologies and interactions in the presence of other cations such as ammonium and sodium[42,99,198,217]. Even the hydrogen bond network could be characterized by NMR scalar couplings across hydrogen bonds and the temperature stability of a G-quadruplex could be monitored[50,51]. Only few reports show NOESY data of oligomers of c-di-GMP or its analogs[280,281]. More recent NMR studies of c-di-GMP have focused on the kinetics and thermodynamics of c-di-GMP by 1D NMR[75] or a combination of 1D and DOSY spectra[259]. Common to all of these studies is the presence of imino H1 resonances in the corresponding NMR spectra[279,280], which are not observed for the rapidly interconverting c-di-GMP monomers and dimers[75].

This phenomenon results from the rapid chemical exchange of the unprotected imino protons with solvent protons. Hydrogen bonds, as found in G-quartets, protect these protons from chemical exchange and enable their detection by NMR. Based on the presence of H1 resonances in NMR spectra of c-di-GMP oligomers at high potassium concentrations, the existence of G-quartets seems reasonable, even though their spatial arrangement cannot be deduced from these 1D NMR data[75,279,280].
In an earlier study on the cation-dependent self-assembly of a lipophilic deoxyguanosine derivative, only one all-syn and one all-anti G-quartet, arranged in a 'head' to 'tail' fashion with potassium located between the two adjacent G-quartets, had been observed[154]. The structural arrangement was based on simple patterns of NOE cross-peaks from only two sets of resonances, which excluded other topologies. In contrast, the situation is more challenging for c-di-GMP due to the presence of three sets of putative octamer resonances, which lead to more complicated cross-peak patterns and thus prevent a straightforward structural analysis. Nevertheless, NOESY cross-peaks from H-bonded and non-H-bonded syn-NH_2 and anti-NH_2 (syn and anti were assigned based on their upfield or downfield shift, respectively, Figure 2.8) to other c-di-GMP protons also indicate the presence of G-quartets in c-di-GMP oligomers[280], even though their topology could not be readily deduced.

In addition to the spectroscopic data, also the slow kinetics of c-di-GMP oligomer formation[75] is an argument for the formation of G-quartets, since this process has also been found to be very slow[87,136,273]. C-di-GMP oligomers dissociate within many hours after dilution from a 1.25 mM stock solution in the presence of 250 mM potassium ions and also re-associate on the time scale of hours after heat dissociation (at 280 μM c-di-GMP concentration)[75].

Interactions between c-di-GMP and acriflavin/proflavin have been confirmed by NMR, even though no new signals were observed which could be attributed to any of the aforementioned tetramers or octamers[75]. The conclusion was that, at least acriflavin/proflavin, does not lead to a significant formation of these oligomers due to the absence of resonances in the H1 region of the NMR spectrum, which would appear if the H1 proton was protected from chemical exchange due to H-bond formation.
Possibly, larger assemblies of several c-di-GMP and dye molecules form under these conditions and are undetectable by NMR due to very fast relaxation. Whether these interactions are directly related to the proposed tetramers and/or octamers remains to be investigated further.

Template-assisted oligomerization may also help to resolve the current issue of *in vivo* relevance of other c-di-GMP oligomers than dimers. Initially, these higher oligomers seemed to be artifacts of high micromolar *in vitro* c-di-GMP concentrations in the presence of potassium ions, while lately also the possibility of their formation *in vivo*, assisted by G-quartet templates as proposed more than twenty years ago[144] has been re-discovered[126,176–178]. While the earlier suggestion of c-di-GMP G-quartet formation was based on CD observations of c-di-GMP in the presence of aromatic compounds and the biological role of c-di-GMP was suspected to be acting as a host or carrier for small aromatic compounds (mimicking an intercalator DNA binding pocket), the latter studies all focused on the aromatic molecules serving as template for G-quartet formation. Knowledge of the detailed structures of these particular oligomers is essential for the verification of the proposed mechanisms of c-di-GMP interactions with aromatic compounds and for the proof that G-quartets are involved.

To determine the structure of these specific c-di-GMP oligomers, we measured ^{1}H-^{1}H NOE and ROE data of unlabeled c-di-GMP at millimolar concentrations in the presence of potassium ions. We have observed several NOEs and ROES from Os to Oa as well as from Oa to Oa', which cannot be attributed to exchange processes, indicating that these resonances belong to subconformations of a single octameric form. It has been speculated that such c-di-GMP octamers may represent an inactivated state and could serve as second messenger supply *in vivo*[280]. Understanding the structure and formation of these oligomers will make it possible to address this question. This may open new ways of biofilm-targeting drugs, which could force the transition from sessile to motile bacterial lifestyle by inactivation of c-di-GMP *via* oligomerization.

2.3.2 Results and Discussion

Structural data of c-di-GMP octamers

We recorded NMR ^{1}H-^{1}H NOESY and ROESY data to initiate the structural analysis of c-di-GMP oligomers. Since these structures are still not solved, we keep the same resonance nomenclature as in the first detailed study of these oligomeric forms[280] as shown in Figure 2.8. The true stoichiometry of c-di-GMP oligomers is still unclear. In order to facilitate particularly the formation of c-di-GMP octamers, we added 150 mM KCl to 1 mM c-di-GMP dissolved in 5 mM TRIS/HCl at pH 8. After incubation at room temperature for 7 days, the ^{1}H NMR spectrum showed the characteristic resonances as seen in all known NMR studies of millimolar c-di-GMP in the presence of potassium ions[75,259,279,280] (Figure 2.8) including the monomer/dimer (fast exchange) resonance and the desired tetramer and octamer resonances. Due to the much lower concentration of tetramers and the absence of intermolecular NOE cross-peaks (as expected due to the approximately 6.8 Å separation of G-quartets, based on the proposed structure), we focus particularly on the structure of the proposed c-di-GMP octamers.

The subsequently recorded ^{1}H-^{1}H NOESY spectra showed several, mostly strong cross-peaks between Oa and Oa' resonances, indicating close spatial proximity, consistent with earlier observations[280]. NOE cross-peaks from H1 to H-bonded and non-H-bonded NH_2 groups confirm the assignment of syn and anti conformation due to the chemical shifts of the NH_2 resonances. This is also corroborated by strong Os-H8 to Os-H1' cross-resonances, indicating a close spatial proximity of these protons.

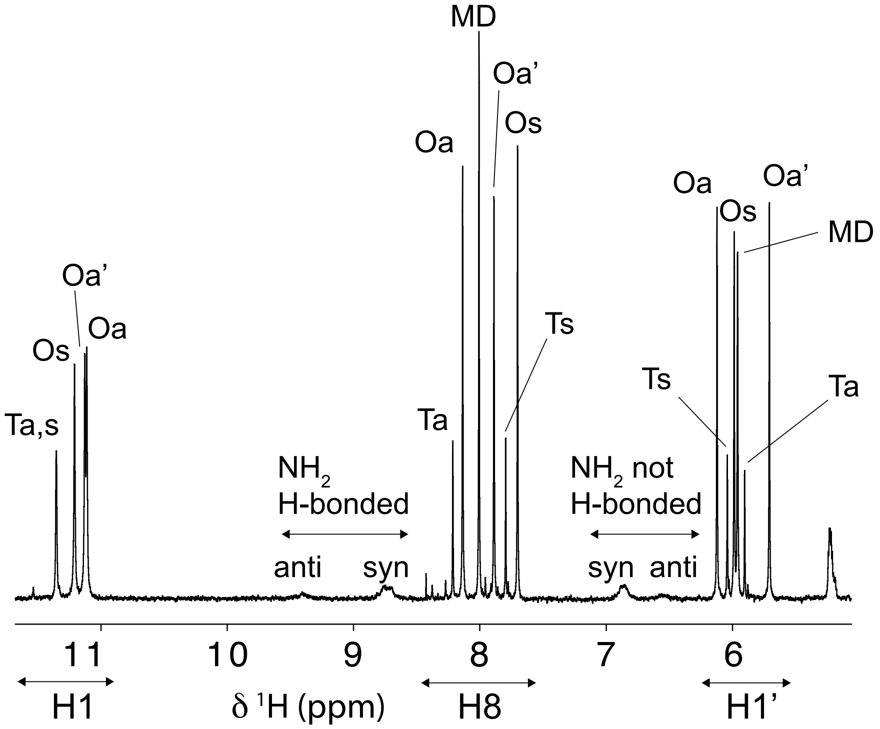


Figure 2.8 1D ^{1}H NMR spectrum of c-di-GMP dissolved in 150 mM KCl, 5 mM TRIS, pH 8 at 25 °C. The assignments were taken from[280] and are indicated for the H1, H8 and H1' protons. All-syn and all-anti tetramers (Ts, Ta) as well as octamers (Os, Oa/Oa') are shown. MD denotes the population-weighted average resonance of monomers and dimers, which are in fast exchange on the chemical shift time scale[75].

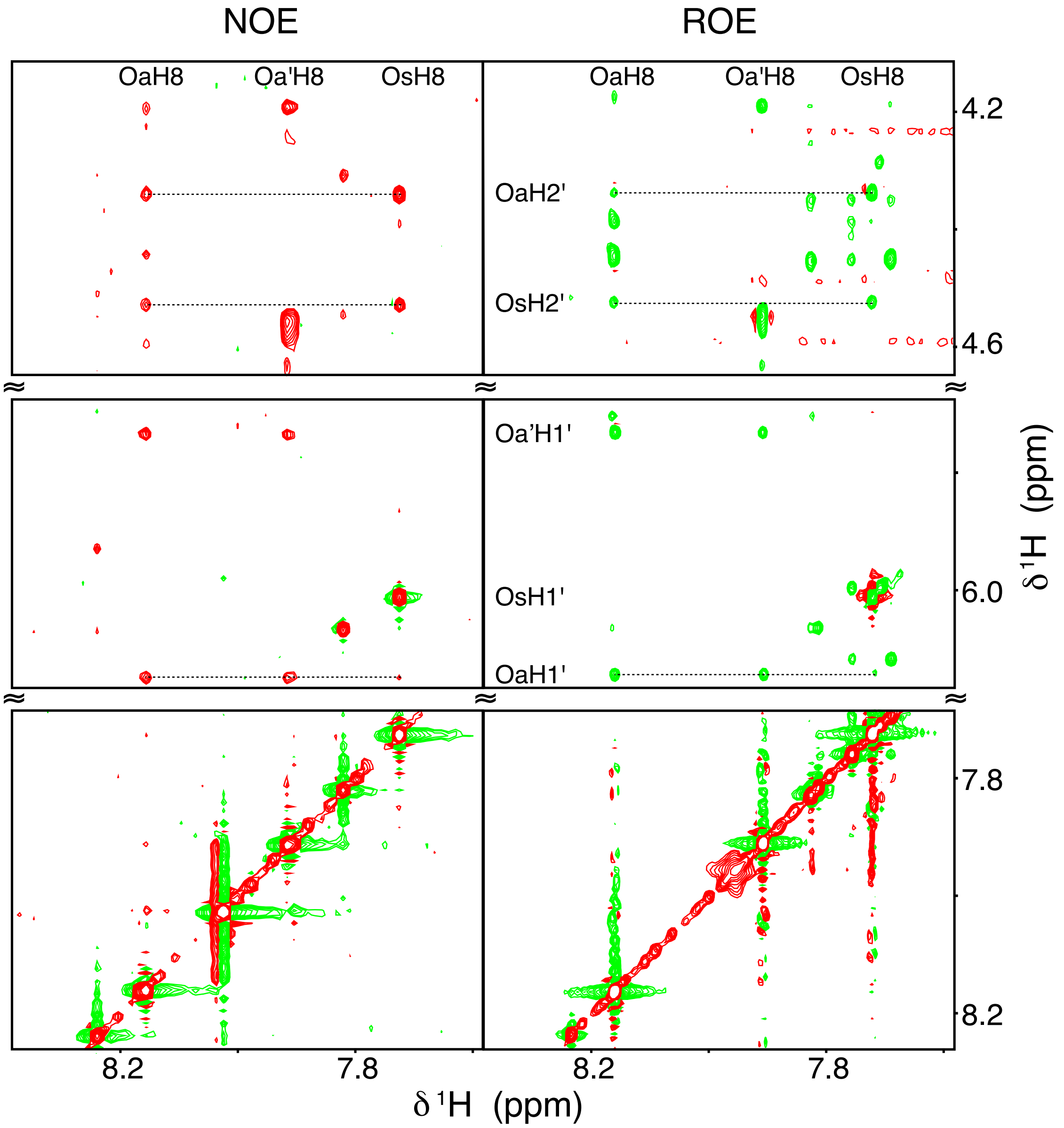


Figure 2.9 Sections of ^{1}H-^{1}H NOESY (left) and ^{1}H-^{1}H ROESY spectra (right) of c-di-GMP oligomers. Cross-peaks to aromatic H8 resonances are shown for sugar resonances other than H1' (top) and to H1' (middle). Diagonal H8 peaks are displayed at the bottom. Positive and negative resonances are colored red and green, respectively. Exchange processes and spin-diffusion can be ruled out as source for the cross-peaks due to their inverted sign in the ROE compared to the NOE experiment. Dashed lines indicate cross-peaks between Os and Oa resonances. From the present data, the H8 atom of Oa seems to be located between Oa' and Os, since Oa shows cross-peaks to both, but not a single cross-peak from Oa' to Os has been observed. The NOESY spectrum was recorded with a mixing time of 300 ms on 1 mM c-di-GMP in the presence of 150 mM KCl, while the ROESY spectrum was recorded with a spin-lock pulse of 75 ms duration on 10 mM c-di-GMP in the presence of 250 mM KCl. The ROESY spectrum was slightly shifted (< 0.03 ppm) to facilitate the comparison of the spectra.

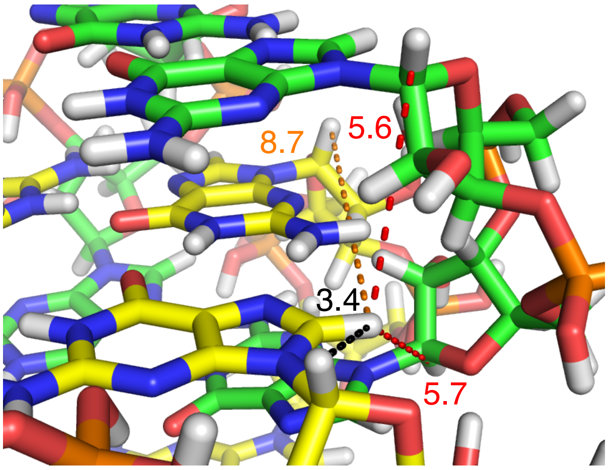


Figure 2.10 H8-H1' distances in a model of an all-anti c-di-GMP octamer. Dashed lines indicate intramolecular (black), intra-quartet (orange) and inter-quartet (red) H8-H1' distances. All annotated distances are in Å. Carbon atoms of the upper octamer half are colored green, whereas those of the lower half are shown in yellow. The structure was modeled as described[75].

species	H1'	H2'	H3'	H4'	H5' *	H5" *
Os	6.01	4.53	5.26	4.38	4.66	4.17
Oa	6.15	(4.34)	(4.81)	(4.46)	(4.61)	(4.21)
Oa'	5.73	(4.19)	(4.57)	(4.55)	(4.64)	(4.26)

Table 2.1 Assignments of sugar proton resonances of c-di-GMP octamers. All values are chemical shifts in ppm. Values in parenthesis are ambiguously assigned. No stereospecific assignments have been obtained for resonances marked by an asterisk.

NOE cross-peaks of comparable (medium) intensity are observed for Oa-H1' to Oa-H8, Oa-H1' to Oa'-H8, Oa'-H1' to Oa'-H8 and Oa'-H1' to Oa-H8, indicative of very similar distances between these nuclei (Figure 2.9). These lower cross-peak intensities compared to the Os-H8 to Os-H1' cross-resonance indicate anti conformations. However, modeling of an all-anti octamer, where Oa and Oa' denote inner and outer G-quartets (or *vice versa*), respectively, reveals that the H8-H1' distance between stacked G-quartets is larger than 5 Å (Figure 2.10). Within a G-quartet, this distance is even larger than 8 Å, which would lead to significantly weaker cross-peaks than observed.

A more detailed discussion of Oa to Oa' cross-resonances is not possible, since apart from H1', no other sugar proton resonances have been unambiguously assigned, yet (Table 2.1). Nevertheless, from the current NOE data, it is obvious that H1' and H8 of both Oa and Oa' show some cross-peaks to the same sugar proton resonances. Therefore, the proposed all-anti octamer structure cannot explain all experimental observations. The large number of cross-peaks between Oa and Oa'

indicates that there is a close connection between these two spectral species.
Additionally, several cross-peaks between Os and Oa, but not Os and Oa' are observed, which indicates that Oa is located between Oa' and Os. This should not be the case if only distinct all-syn and all-anti octamers would exist. To verify, whether these observed anti to syn cross-peaks are really due to through-space magnetization transfer and do not arise from exchange processes or spin-diffusion, a ^{1}H-^{1}H ROESY spectrum was recorded. Even though the ROE spectrum suffered from the limitation in the spin-lock time as well as severe distortions from the solvent signal, some of the cross-peaks between syn and anti forms (Oa-H8 to Os-H1, Oa-H1' to Os-H8, (Figure 2.9, middle), Os-H3' to Oa-H8) and mutual cross-peaks to the sugar region (Oa-H8 and Os-H8 to 4.53 ppm (Os-H2'), 4.34 ppm (Oa-H2', ambiguously assigned), Figure 2.9, top) were still observed and had an opposite sign than the diagonal resonances, indicative of real non-exchange contacts. Interestingly, the similar intensities of the Oa-H8 to Os-H1 and the Os-H8 to Os-H1 cross-peaks (not shown) indicate very similar distances between the respective atoms.
Unfortunately, not more than the described cross-peaks between Os and Oa could be found in the ROE experiment. Removal of spectral artifacts and an improved baseline could possibly reveal more of such real ROE cross-peaks. At least the same resonances as in the NOE experiment should be observed and the sign of the respective cross-resonances would again allow determination of their origin. The much reduced number of cross-resonances can be attributed to the limitation in the applied spin-lock time (power). This limitation could be circumvented by using another (non-cryo) probe. It clearly has to be mentioned that more ^{1}H-^{1}H ROE cross-peaks are observed between Oa and Oa' than between Oa and Os. This corresponds to the observation that the cross-peaks between the former two spectral species are also quite strong in NOE experiments, indicating a shorter distance. Thus, the bases of Oa and Oa' are in closer spatial proximity than the Os base to any of the other two.
NOE cross-peaks between syn and anti resonances of octamers are evidence that these are not distinct all-syn and all-anti structures. Furthermore, we measured T_2 relaxation times of the H8 protons. The octamer resonances show very similar values around 100 ms (Oa: 95 ms, Oa': 102 ms, Os: 103 ms), which is significantly shorter than the tetramer T_2 relaxation times (Ta: 173 ms, Ts: 136 ms). This supports the finding that the assigned octamers belong indeed to larger assemblies. Higher oligomers than octamers, which we have reported previously[75], have T_2 times between 45 and 75 ms (unpublished). This indicates that the investigated oligomers in this study have the ability to further increase in size.

A closer inspection of the already published ^{1}H-^{1}H NOESY data (Figure 6, Zhang *et al.*, 2006[280]) reveals the presence of several Os to Oa cross-peaks (e.g. Oa-H8 to Os-H1, Os-H8 to Oa-H1, syn-NH_2 to Oa-H8, syn-NH_2 to Oa-H1' or Os-H8 to Oa-H1'), which apparently had been overlooked. This has lead to the putative conclusion of distinct all-anti and all-syn c-di-GMP octamer structures[280].

The stoichiometry of the c-di-GMP octamers is ambiguous, since DOSY NMR data were only used for qualitative ranking of the molecular size (octamers diffuse more slowly than tetramers). The diffusion constants were not discussed in detail but their values are very similar (1.6×10^{-6} $\mathrm{cm^2/s}$ for octamers and 1.9×10^{-6} $\mathrm{cm^2/s}$ for tetramers)[280]. According to the Stokes-Einstein equation (eq. 2.1) for spherical particles, the diffusion constant (D) is dependent on the radius r of the molecular sphere, while all other parameters (temperature T, solvent viscosity η) depend on experimental conditions and k_B is the Boltzmann constant.

$$\mathrm{D} = \frac{\mathrm{k_B T}}{6\pi\eta \mathrm{r}} \tag{2.1}$$

Assuming the proposed octamer and tetramer structures were correct, this finding could be explained by the compact form of a single octamer, which would be only moderately larger than a tetramer due to intercalation, resulting in similar radii and comparable diffusion constants. Nevertheless, this experimental observable on its own is clearly not sufficient to properly determine the stoichiometry in this case.

Taken together, our results indicate that the base of Oa is located between those of Oa' and Os based on NOE cross-peaks originating from short distances between the involved protons. In order to determine the 3D structure of these oligomers or the single oligomer, if the described resonances belong to a single species, intermolecular and intramolecular contacts (or inter- to intra-nucleotide) need to be distinguished.
^{1}H-^{1}H TOCSY experiments yield through-bond information, which would be ideal in this case to address the question of intra-sugar resonances. All protons, particularly those of the sugar moieties, would be accessible *via* a cascade of magnetization transfer since each of the sugar protons has at least one other sugar proton within three bonds. However, no ^{1}H-^{1}H TOCSY cross-peaks could be observed for any of the H1' protons for both octamers and tetramers. Therefore, it is highly likely, that the dihedral angle H1'-C1'-C2'-H2' is about 90 degrees. This would correspond to a ^3J coupling constant of approximately 0 Hz between H1' and H2'. It is noteworthy, that this finding would be compatible with the modeled structures (Figure 2.6) based on the proposed structural scaffold[280].

Clearly, the nearly identical kinetics and thermodynamics of oligomer formation and dissociation as well as the similar relaxation behavior indicate that the dominant spectral species (referred to as octamers) belong indeed to subconformations of the same molecular assembly as proposed due to NOE contacts between Os and Oa.

C-di-GMP oligomers in D_2O

Since most of the sugar proton resonances are very close to the H_2O solvent resonance, these are obscured or even removed by solvent suppression. In order to observe these sugar resonances, solvent suppression with the conventional WATERGATE[195] scheme cannot be used. Therefore, we changed the solvent from H_2O to D_2O.

In D_2O, initially, the same characteristic c-di-GMP oligomer resonances as in H_2O, are observed in the presence of K^+ ions (Figure 2.11). The resonance frequencies in D_2O differ non-uniformly from those in H_2O (maximal ± 0.03 ppm).
It turned out that the H1 resonances showed a time-dependent, non-uniform reduction of intensities, indicating proton to deuterium exchange processes. The H1 resonance of Os diminished much faster than the Oa and Oa' resonances, while Oa' remained the most intense of these three. This indicates that the Oa'-H1 proton is better protected from chemical exchange with the solvent than the others. Interestingly, the overlapped Ta and Ts H1 protons exchanged noticeably slower than the H1 proton of Oa and Oa'. Based on the proposed c-di-GMP tetramer structures this is unexpected, since their H1 protons would be more solvent-accessible than the inner G-quartets in octamers.
In contrast to the non-uniform disappearance of H1 resonances, H8 and H1' resonance intensities of all three octameric forms remained very similar. After the equilibrium was reached, the integrals of all non-exchanged oligomer resonances were reduced in D_2O compared to H_2O solvent. Consequently, the MD resonances (monomer/dimer fast exchange resonances) gained intensity accompanied by an upfield shift, due to a change in the chemical equilibria among the various c-di-GMP species[75] (Figure 2.11).
These observations lead to the conclusion, that c-di-GMP oligomers are slightly less stable in D_2O as compared to H_2O, even though virtually complete exchange of H1 protons to deuterium is observed. Possible explanations are that deuterium bonds are weaker than hydrogen bonds under these conditions or a change of the aromatic pi-stacking interaction energies in D_2O as compared to H_2O.
It is also remarkable that the tetramer to octamer ratio decreased with time after D_2O addition (Figure 2.11). This observation combined with the simultaneous upfield shift of the MD resonance indicate that monomers, dimers and octamers are formed at the expense of tetramers under these conditions. This is counterintuitive, since dissociation of the putative octamers in H_2O has been shown to lead to an increase in monomers and dimers (which causes the upfield shift of the MD resonance) as well as an initial increase of tetramer resonance integrals[75]. The structure determination of these c-di-GMP oligomers is essential to completely understand the thermodynamic and kinetic behavior in both H_2O and D_2O.

NOESY spectra of c-di-GMP in D_2O solvent showed significantly more sugar resonances apart from H1' due to the reduced distortion by the solvent signal and solvent suppression artifacts (see above). The combination of information from NOESY spectra of c-di-GMP in aqueous solutions as well as NOESY and TOCSY spectra recorded in D_2O solution facilitated the assignment of the Os sugar protons, even though no stereospecific assignments of H5' and H5" could be achieved (Table 2.1). In this case, the well-isolated resonance at about 5.26 ppm, identified as Os-H3' was essential. Due to the central position of the H3' in the sugar, cross-peaks to all other sugar protons were expected. Indeed the NOE spectrum showed relatively strong cross-peaks to this proton and the TOCSY data were sufficient to determine

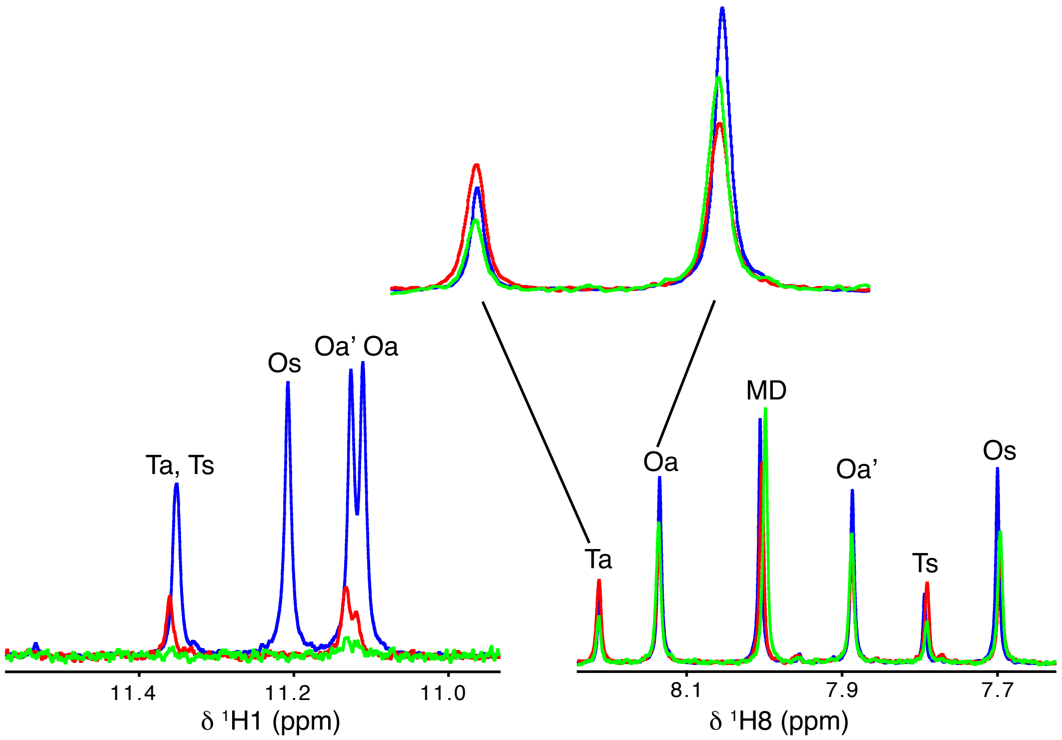


Figure 2.11 H1 (left) and H8 (right) region of 1D NMR spectra of 1 mM c-di-GMP in the presence of potassium ions (150 mM). Spectra in D_2O are shown 4 hours (red) and 22 hours (green) after solvent exchange, respectively. For comparison, a spectrum recorded in H_2O, scaled to match the H8 integral sum, is superimposed (blue). The H8 spectra in D_2O were shifted by -0.02 ppm relative to the referenced spectrum in H_2O for clarity. Resonance labels are identical to Figure 2.8. The inset shows the H8 resonances of the putative all-anti tetramer (Ta) and all-anti octamer (Oa).

the through-bond connectivity, which subsequently lead to the assignments.
Unfortunately, the situation is again much more complicated for Oa' and particularly Oa due to the large number of cross-peaks to sugar resonances of all three spectral species. The TOCSY spectrum could also not solve these problems, since none of these resonances is isolated enough to be assigned unambiguously and spectral artifacts obscured the region of interest. The currently assigned sugar resonances (based solely on NOE intensities) of the putative c-di-GMP octamers are shown in Table 2.1. Compared to chemical shift estimations of monomeric c-di-GMP or NMR data from dimeric c-di-GMP in complex with the PilZ domain protein PA4608 (BMRB entry 17496[86]), it is noticeable that the chemical shifts of c-di-GMP octamers deviate quite strongly, indicating a different local environment.

Even though, the current data is not sufficient for the structure determination, it provides a reasonable foundation for further experiments. In order to unambiguosly establish through-bond connectivities, (^{15}N,^{13}C) isotope-labeled c-di-GMP could be used. The correlation of proton and heteronuclear resonances would directly indicate which resonances belong to the same molecule. Hydrogen bond experiments could finally elucidate the G-quartet topology. Applications of heteronuclear NMR experiments would also help to solve overlap problems, since the resonances are more dispersed (particularly ^{13}C). Determination of ^3J scalar couplings between a variety of nuclei would allow to determine torsion angles with the known Karplus coefficients for nucleotides[19,137,138,168]. This would also reveal all sugar conformations. With these experiments in addition to the already present data, the three dimensional structure determination should be possible.

2.3.3 Conclusions and Perspective

Based on our NOE and ROE data, it is highly likely, that the previously proposed structures and the interpretation as related but distinct c-di-GMP all-syn (Os) and all-anti (Oa/Oa') octamer structures[280] are not completely correct. It appears that these structures may represent subconformations of a single species due to the observation of nearly identical kinetics of their formation and dissociation[75]. Furthermore, the very similar H8 T_2 relaxation data between these spectral species corroborate this interpretation.
In the presence of several aromatic compounds, the formation of c-di-GMP oligomers based on G-quartets has been surmised recently, even at low micromolar c-di-GMP concentrations[126,176–178]. Models for both intercalation and end-stacking have been proposed for the interaction of aromatic compounds with c-di-GMP G-quartets at micromolar c-di-GMP concentrations[126,176–178]. One such complex presumably involves a hemin moiety, a c-di-GMP octamer or tetramer and a proflavin molecule and is held responsible for the enhancement of peroxidase activity in the presence of c-di-GMP at high metal cation concentrations[178].
Interestingly, the activity in the presence of lithium was higher than in the presence of potassium. Since tetramers are reported to be stabilized by lithium as compared

to potassium[279,280], this may indicate that tetramers are involved in the peroxidase reaction. Thus, these authors expect oligomeric c-di-GMP structures (> dimers) also to exist *in vivo*.

No incompatibilities of the proposed tetramer structures with experimental data has been observed so far. However, data concerning these tetramers is sparse since they are not highly populated at millimolar c-di-GMP concentrations in the presence of 250 mM potassium and the G-quartets would be 6.8 Å apart according to the models[75] of the proposed structures[280]. This would prevent observation of inter-quartet NOEs, which would make it quite difficult to assess whether two G-quartets belong to a single or two distinct tetramers. The former is again supported by similar kinetics and resonance integrals for Ta and Ts[75], while the latter may seem reasonable due to the difference in the H8 T_2 relaxation.

Knowledge of the c-di-GMP oligomer structures would help to improve and understand the mechanism of peroxidase activity enhancement[178]. The structures of octamers could also shed light on the structure of tetramers, since we have shown before that tetramers are intermediates towards octamer formation[75]. These oligmers must provide a basis for further assembling c-di-GMP monomers into higher polymorphs, since at high concentrations, such oligomers have been observed in the presence of potassium[75].

Even if these specific c-di-GMP oligomers turn out not to be involved in the stacking processes of aromatic compounds, insight into their structure may still provide a new view of structural diversity, which may be linked to signaling. Furthermore, the model for the kinetics and thermodynamics could be improved and new applications to target biofilms may be envisioned.

2.3.4 Materials and Methods

NMR sample preparation

C-di-GMP was produced enzymatically and purified as described[277]. NMR samples were prepared from lyophilized c-di-GMP by addition of the desired buffer. For higher oligomer (> octamer) samples, 10 mM c-di-GMP was dissolved in 250 mM KCl, 10 mM $MgCl_2$, 5 mM NaH_2PO_4, pH 7.4, 5% D_2O[75]. C-di-GMP octamers were studied at a concentration of 1 mM c-di-GMP dissolved in 150 mM KCl, 5 mM TRIS/HCl, pH 8.0, 5% D_2O. Solvent exchange to 100% D_2O was achieved by lyophilization and addition of 100% D_2O (three times), under dry atmosphere in a glovebox.

NMR measurements and data analysis

All NMR experiments were recorded on a Bruker Avance DRX 600 spectrometer, equipped with a triple-resonance, triple-axis gradient (TXI) probe and on a Bruker DRX 800 spectrometer equipped with a cryogenic triple-resonance, Z-gradient (TCI) probe. 1D spectra were recorded using the excitation sculpting scheme[104], with acquisition times of 1.99 s and a recycle delay of 1.0 s. Chemical shifts in aqueous solution were referenced to external TSP. All experiments were performed at 25 °C (298 K). ^{1}H-^{1}H NOESY experiments in H_2O were recorded with an in-house written pulse sequence using WATERGATE[195] and water flip-back pulses[81] for suppression of solvent signals. Acquisition times of 1000 ms (t_2) and 80.4 ms (t_1) as well as a mixing time of 300 ms and a recycle delay of 1.0 s were applied. The total experiment time was 31.5 h. For 100% D_2O samples, WATERGATE and water flip-back pulses were removed from the ^{1}H-^{1}H NOE pulse program and the delays were compensated accordingly. The ^{1}H-^{1}H ROESY experiments were recorded with an in-house written pulse sequence, similar to the NOESY experiment, and acquisition times of 1000 ms (t_2) and 70.5 ms (t_1) and a ROESY spin-lock time of 75 ms. The total experiment time was 11.5 h. The ^{1}H-^{1}H TOCSY spectrum was recorded in D_2O with acquisition times of 820 ms (t_2) and 82.5 ms (t_1) and a mixing time of 58 ms. The total experimental time was 3.5 hours. 1D NMR datasets were processed and analyzed with Topspin 2.1.6 (Bruker). Multidimensional NMR data were processed with NMRPipe[47] and analyzed with Sparky[77].

Bibliography

[1] S. Abel, P. Chien, P. Wassmann, T. Schirmer, V. Kaever, M. T. Laub, T. A. Baker, and U. Jenal. Regulatory cohesion of cell cycle and cell differentiation through interlinked phosphorylation and second messenger networks. *Mol Cell*, 43(4):550–60, Aug. 2011.

[2] M. Affolter, T. Marty, M. A. Vigano, and A. Jazwinska. Nuclear interpretation of dpp signaling in drosophila. *EMBO J*, 20(13):3298–305, July 2001.

[3] K. Akasaka. Highly fluctuating protein structures revealed by variable-pressure nuclear magnetic resonance. *Biochemistry*, 42(37):10875–85, Sept. 2003.

[4] R. A. Alm, A. J. Bodero, P. D. Free, and J. S. Mattick. Identification of a novel gene, pilz, essential for type 4 fimbrial biogenesis in pseudomonas aeruginosa. *J Bacteriol*, 178(1):46–53, Jan. 1996.

[5] D. Amikam and M. Y. Galperin. Pilz domain is part of the bacterial c-di-gmp binding protein. *Bioinformatics*, 22(1):3–6, Jan. 2006.

[6] N. Amiot, K. Heintz, and B. Giese. New approach for the synthesis of c-di-gmp and its analogues. *Synthesis-Journal of Synthetic Organic Chemistry*, (24):4230–4236, 2006.

[7] M. R. Arnold, W. Kremer, H. D. Lüdemann, and H. R. Kalbitzer. 1h-nmr parameters of common amino acid residues measured in aqueous solutions of the linear tetrapeptides gly-gly-x-ala at pressures between 0.1 and 200 mpa. *Biophys Chem*, 96(2-3):129–40, May 2002.

[8] S. Arnott, R. Chandrasekaran, and C. M. Marttila. Structures for polyinosinic acid and polyguanylic acid. *Biochem J*, 141(2):537–43, Aug. 1974.

[9] Y. Bai, H. Zhou, and Y. Zhou. Critical nucleation size in the folding of small apparently two-state proteins. *Protein Sci*, 13(5):1173–81, May 2004.

[10] J. Balbach. Compaction during protein folding studied by real-time nmr diffusion experiments. *JOURNAL-AMERICAN CHEMICAL SOCIETY*, 122(24):5887–5888, 2000.

[11] J. Balbach, V. Forge, W. S. Lau, J. A. Jones, N. A. van Nuland, and C. M. Dobson. Detection of residue contacts in a protein folding intermediate. *Proc Natl Acad Sci U S A*, 94(14):7182–5, July 1997.

[12] J. Balbach, V. Forge, W. S. Lau, N. A. van Nuland, K. Brew, and C. M. Dobson. Protein folding monitored at individual residues during a two-dimensional nmr experiment. *Science*, 274(5290):1161–3, Nov. 1996.

[13] J. Balbach, V. Forge, N. A. van Nuland, S. L. Winder, P. J. Hore, and C. M. Dobson. Following protein folding in real time using nmr spectroscopy. *Nat Struct Biol*, 2(10):865–70, Oct. 1995.

[14] T. R. M. Barends, E. Hartmann, J. J. Griese, T. Beitlich, N. V. Kirienko, D. A. Ryjenkov, J. Reinstein, R. L. Shoeman, M. Gomelsky, and I. Schlichting. Structure and mechanism of a bacterial light-regulated cyclic nucleotide phosphodiesterase. *Nature*, 459(7249):1015–8, June 2009.

[15] J. Barette, A. Velyvis, T. L. Religa, D. M. Korzhnev, and L. E. Kay. Cross-validation of the structure of a transiently formed and low populated ff domain folding intermediate determined by relaxation dispersion nmr and cs-rosetta. *J Phys Chem B*, (in press), Dec. 2011.

[16] A. Bax. Weak alignment offers new nmr opportunities to study protein structure and dynamics. *Protein Sci*, 12(1):1–16, Jan. 2003.

[17] B. K. Bharati, I. M. Sharma, S. Kasetty, M. Kumar, R. Mukherjee, and D. Chatterji. A full length bifunctional protein involved in c-di-gmp turnover is required for long term survival under nutrient starvation in mycobacterium smegmatis. *Microbiology*, Feb. 2012.

[18] E. A. Bienkiewicz and K. J. Lumb. Random-coil chemical shifts of phosphorylated amino acids. *J Biomol NMR*, 15(3):203–6, Nov. 1999.

[19] M. J. Blommers, C. A. Haasnoot, J. A. Walters, G. A. van der Marel, J. H. van Boom, and C. W. Hilbers. Solution structure of the 3'-5' cyclic dinucleotide d(papa). a combined nmr, uv melting, and molecular mechanics study. *Biochemistry*, 27(22):8361–9, Nov. 1988.

[20] A. Boehm, S. Steiner, F. Zaehringer, A. Casanova, F. Hamburger, D. Ritz, W. Keck, M. Ackermann, T. Schirmer, and U. Jenal. Second messenger signalling governs escherichia coli biofilm induction upon ribosomal stress. *Mol Microbiol*, 72(6):1500–16, June 2009.

[21] D. Braun, G. Wider, and K. Wüthrich. Sequence-corrected 15n" random coil" chemical shifts. *Journal of the American Chemical Society*, 116(19):8466–8469, 1994.

[22] E. Brouillette, M. Hyodo, Y. Hayakawa, D. K. R. Karaolis, and F. Malouin. 3',5'-cyclic diguanylic acid reduces the virulence of biofilm-forming staphylococcus aureus strains in a mouse model of mastitis infection. *Antimicrob Agents Chemother*, 49(8):3109–13, Aug. 2005.

[23] P. N. Bryan and J. Orban. Proteins that switch folds. *Curr Opin Struct Biol*, 20(4):482–8, Aug. 2010.

[24] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins*, 21(3):167–95, Mar. 1995.

[25] A. Buevich and J. Baum. Dynamics of unfolded proteins: Incorporation of distributions of correlation times in the model free analysis of nmr relaxation data. *Journal of the American Chemical Society*, 121(37):8671–8672, 1999.

[26] A. Buevich, U. Shinde, M. Inouye, and J. Baum. Backbone dynamics of the natively unfolded pro-peptide of subtilisin by heteronuclear nmr relaxation studies. *Journal of biomolecular NMR*, 20(3):233–249, 2001.

[27] G. Campbell and A. Tomlinson. Transducing the dpp morphogen gradient in the wing of drosophila: regulation of dpp targets by brinker. *Cell*, 96(4):553–62, Feb. 1999.

[28] H. Y. Carr and E. M. Purcell. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Phys. Rev.*, 94:630–638, May 1954.

[29] J. P. Carver and R. E. Richards. A general two-site solution for the chemical exchange produced dependence of t2 upon the carr-purcell pulse separation. *J Magn Reson*, 6(1):89–105, 1972.

[30] A. Cavalli, X. Salvatella, C. M. Dobson, and M. Vendruscolo. Protein structure determination from nmr chemical shifts. *Proc Natl Acad Sci U S A*, 104(23):9615–20, June 2007.

[31] J. Cavanagh, W. J. Fairbrother, A. G. Palmer, M. Rance, and N. J. Skelton. *Protein NMR Spectroscopy: Principles and Practice*. Elsevier, second edition edition, 2007.

[32] C. Chan, R. Paul, D. Samoray, N. C. Amiot, B. Giese, U. Jenal, and T. Schirmer. Structural basis of activity and allosteric control of diguanylate cyclase. *Proc Natl Acad Sci U S A*, 101(49):17084–9, Dec. 2004.

[33] C. Chan, R. Paul, D. Samoray, N. C. Amiot, B. Giese, U. Jenal, and T. Schirmer. Structural basis of activity and allosteric control of diguanylate cyclase. *Proc Natl Acad Sci U S A*, 101(49):17084–9, Dec. 2004.

[34] A. L. Chang, J. R. Tuckerman, G. Gonzalez, R. Mayer, H. Weinhouse, G. Volman, D. Amikam, M. Benziman, and M. A. Gilles-Gonzalez. Phosphodiesterase a1, a regulator of cellulose synthesis in acetobacter xylinum, is a heme-based sensor. *Biochemistry*, 40(12):3420–6, Mar. 2001.

[35] K.-H. Chin, Y.-C. Lee, Z.-L. Tu, C.-H. Chen, Y.-H. Tseng, J.-M. Yang, R. P. Ryan, Y. McCarthy, J. M. Dow, A. H.-J. Wang, and S.-H. Chou. The camp receptor-like protein clp is a novel c-di-gmp receptor linking cell-cell signaling to virulence gene expression in xanthomonas campestris. *J Mol Biol*, 396(3):646–62, Feb. 2010.

[36] B. Christen, M. Christen, R. Paul, F. Schmid, M. Folcher, P. Jenoe, M. Meuwly, and U. Jenal. Allosteric control of cyclic di-gmp signaling. *J Biol Chem*, 281(42):32015–24, Oct. 2006.

[37] M. Christen, B. Christen, M. G. Allan, M. Folcher, P. Jenö, S. Grzesiek, and U. Jenal. Dgra is a member of a new family of cyclic diguanosine monophosphate receptors and controls flagellar motor function in caulobacter crescentus. *Proc Natl Acad Sci U S A*, 104(10):4112–7, Mar. 2007.

[38] T. D. W. Claridge. *High-resolution NMR techniques in organic chemistry*, volume 19. Pergamon Pr, 1999.

[39] P. L. Clark. Protein folding in the cell: reshaping the folding funnel. *Trends Biochem Sci*, 29:527–534, 2004.

[40] G. M. Clore and A. M. Gronenborn. Nmr structure determination of proteins and protein complexes larger than 20 kda. *Curr Opin Chem Biol*, 2(5):564–70, Oct. 1998.

[41] F. Cordier, B. Hartmann, M. Rogowski, M. Affolter, and S. Grzesiek. Dna recognition by the brinker repressor-an extreme case of coupling between binding and folding. *J Mol Biol*, 361(4):659–72, Aug. 2006.

[42] M. Crnugelj, N. V. Hud, and J. Plavec. The solution structure of d(g(4)t(4)g(3))(2): a bimolecular g-quadruplex with a novel fold. *J Mol Biol*, 320(5):911–24, July 2002.

[43] V. Dapic, V. Abdomerovic, R. Marrington, J. Peberdy, A. Rodger, J. O. Trent, and P. J. Bates. Biophysical and biological properties of quadruplex oligodeoxyribonucleotides. *Nucleic Acids Res*, 31(8):2097–107, Apr. 2003.

[44] J. T. Davis. G-quartets 40 years later: from 5'-gmp to molecular biology and supramolecular chemistry. *Angew Chem Int Ed Engl*, 43(6):668–98, Jan. 2004.

[45] N. De, M. V. A. S. Navarro, R. V. Raghavan, and H. Sondermann. Determinants for the activation and autoinhibition of the diguanylate cyclase response regulator wspr. *J Mol Biol*, 393(3):619–33, Oct. 2009.

[46] N. De, M. Pirruccello, P. V. Krasteva, N. Bae, R. V. Raghavan, and H. Sondermann. Phosphorylation-independent regulation of the diguanylate cyclase wspr. *PLoS Biol*, 6(3):e67, Mar. 2008.

[47] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax. Nmrpipe: a multidimensional spectral processing system based on unix pipes. *J Biomol NMR*, 6(3):277–93, Nov. 1995.

[48] V. M. Delgado-Nixon, G. Gonzalez, and M. A. Gilles-Gonzalez. Dos, a heme-binding pas protein from escherichia coli, is a direct oxygen sensor. *Biochemistry*, 39(10):2685–91, Mar. 2000.

[49] K. A. Dill, S. B. Ozkan, M. S. Shell, and T. R. Weikl. The protein folding problem. *Annu Rev Biophys*, 37:289–316, 2008.

[50] A. J. Dingley, J. E. Masse, J. Feigon, and S. Grzesiek. Characterization of the hydrogen bond network in guanosine quartets by internucleotide 3hj(nc)' and 2hj(nn) scalar couplings. *J Biomol NMR*, 16(4):279–89, Apr. 2000.

[51] A. J. Dingley, R. D. Peterson, S. Grzesiek, and J. Feigon. Characterization of the cation and temperature dependence of dna quadruplex hydrogen bond properties using high-resolution nmr. *J Am Chem Soc*, 127(41):14466–72, Oct. 2005.

[52] C. Dobson, A. Sali, and M. Karplus. Protein folding: a perspective from theory and experiment. *Angewandte Chemie International Edition*, 37(7):868–893, 1998.

[53] C. M. Dobson. Protein folding and misfolding. *Nature*, 426(6968):884–90, Dec. 2003.

[54] J. Draper, K. Karplus, and K. M. Ottemann. Identification of a chemoreceptor zinc-binding domain common to cytoplasmic bacterial chemoreceptors. *J Bacteriol*, 193(17):4338–45, Sept. 2011.

[55] A. Duerig, S. Abel, M. Folcher, M. Nicollier, T. Schwede, N. Amiot, B. Giese, and U. Jenal. Second messenger-mediated spatiotemporal control of protein degradation regulates bacterial cell cycle progression. *Genes Dev*, 23(1):93–104, Jan. 2009.

[56] A. K. Dunker, C. J. Brown, and Z. Obradovic. Identification and functions of usefully disordered proteins. *Adv Protein Chem*, 62:25–49, 2002.

[57] A. K. Dunker, M. S. Cortese, P. Romero, L. M. Iakoucheva, and V. N. Uversky. Flexible nets. the roles of intrinsic disorder in protein interaction networks. *FEBS J*, 272(20):5129–48, Oct. 2005.

[58] H. J. Dyson and P. E. Wright. Unfolded proteins and protein folding studied by nmr. *Chem Rev*, 104(8):3607–22, Aug. 2004.

[59] H. J. Dyson and P. E. Wright. Elucidation of the protein folding landscape by nmr. *Methods Enzymol*, 394:299–321, 2005.

[60] H. J. Dyson and P. E. Wright. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol*, 6(3):197–208, Mar. 2005.

[61] M. Egli, R. V. Gessner, L. D. Williams, G. J. Quigley, G. A. van der Marel, J. H. van Boom, A. Rich, and C. A. Frederick. Atomic-resolution structure of the cellulose synthase regulator cyclic diguanylic acid. *Proc Natl Acad Sci U S A*, 87(8):3235–9, Apr. 1990.

[62] D. Eliezer. Biophysical characterization of intrinsically disordered proteins. *Curr Opin Struct Biol*, 19(1):23–30, Feb. 2009.

[63] D. Eliezer, J. Chung, H. J. Dyson, and P. E. Wright. Native and non-native secondary structure and dynamics in the ph 4 intermediate of apomyoglobin. *Biochemistry*, 39(11):2894–901, Mar. 2000.

[64] N. A. Farrow, O. Zhang, A. Szabo, D. A. Torchia, and L. E. Kay. Spectral density function mapping using 15n relaxation data exclusively. *J Biomol NMR*, 6(2):153–62, Sept. 1995.

[65] N. G. Faux, S. P. Bottomley, A. M. Lesk, J. A. Irving, J. R. Morrison, M. G. de la Banda, and J. C. Whisstock. Functional insights from the distribution and role of homopeptide repeat-containing proteins. *Genome Res*, 15(4):537–51, Apr. 2005.

[66] P. C. Fineran, N. R. Williamson, K. S. Lilley, and G. P. C. Salmond. Virulence and prodigiosin antibiotic biosynthesis in serratia are regulated pleiotropically by the ggdef/eal domain protein, pigx. *J Bacteriol*, 189(21):7653–62, Nov. 2007.

[67] N. C. Fitzkee, P. J. Fleming, and G. D. Rose. The protein coil library: a structural database of nonhelix, nonstrand fragments derived from the pdb. *Proteins*, 58(4):852–4, Mar. 2005.

[68] E. Fraenkel and C. O. Pabo. Comparison of x-ray and nmr structures for the antennapedia homeodomain-dna complex. *Nat Struct Biol*, 5(8):692–7, Aug. 1998.

[69] C. French, J. Bell, and F. Ward. Diversity and distribution of hemerythrin-like proteins in prokaryotes. *FEMS microbiology letters*, 279(2):131–145, 2008.

[70] M. Fuxreiter, I. Simon, P. Friedrich, and P. Tompa. Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J Mol Biol*, 338(5):1015–26, May 2004.

[71] B. L. Gaffney and R. A. Jones. One-flask synthesis of cyclic diguanosine monophosphate (c-di-gmp). *Curr Protoc Nucleic Acid Chem*, Chapter 14:Unit 14.8.1–7, Mar. 2012.

[72] M. Y. Galperin. Structural classification of bacterial response regulators: diversity of output domains and domain combinations. *J Bacteriol*, 188(12):4169–82, June 2006.

[73] D. Garrett, R. Powers, A. Gronenborn, and G. Clore. A common sense approach to peak picking in two-, three-, and four-dimensional spectra using automatic computer analysis of contour diagrams. *Journal of magnetic resonance*, 95(1):214–220, 1991.

[74] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. Wilkins, R. Appel, and A. Bairoch. Protein identification and analysis tools on the expasy server. *The proteomics protocols handbook*, pages 571–607, 2005.

[75] M. Gentner, M. G. Allan, F. Zaehringer, T. Schirmer, and S. Grzesiek. Oligomer formation of the bacterial second messenger c-di-gmp: Reaction rates and equilibrium constants indicate a monomeric state at physiological concentrations. *J Am Chem Soc*, 134(2):1019–29, Jan. 2012.

[76] H. S. Girgis, Y. Liu, W. S. Ryu, and S. Tavazoie. A comprehensive genetic characterization of bacterial motility. *PLoS Genet*, 3(9):1644–60, Sept. 2007.

[77] T. D. Goddard and D. G. Kneller. Sparky 3. *University of California, San Francisco*, 2006.

[78] D. M. Gray, J.-D. Wen, C. W. Gray, R. Repges, C. Repges, G. Raabe, and J. Fleischhauer. Measured and calculated cd spectra of g-quartets stacked with the same or opposite polarities. *Chirality*, 20(3-4):431–40, Mar. 2008.

[79] N. J. Greenfield. Analysis of circular dichroism data. *Methods Enzymol*, 383:282–317, 2004.

[80] M. Grimmler, Y. Wang, T. Mund, Z. Cilensek, E.-M. Keidel, M. B. Waddell, H. Jäkel, M. Kullmann, R. W. Kriwacki, and L. Hengst. Cdk-inhibitory activity and stability of p27kip1 are directly regulated by oncogenic tyrosine kinases. *Cell*, 128(2):269–80, Jan. 2007.

[81] S. Grzesiek and A. Bax. The importance of not saturating water in protein nmr. application to sensitivity enhancement and noe measurements. *Journal of the American Chemical Society*, 115(26):12593–12594, 1993.

[82] S. Grzesiek, A. Bax, J. S. Hu, J. Kaufman, I. Palmer, S. J. Stahl, N. Tjandra, and P. T. Wingfield. Refined solution structure and backbone dynamics of hiv-1 nef. *Protein Sci*, 6(6):1248–63, June 1997.

[83] S. Grzesiek, F. Cordier, and A. J. Dingley. Scalar couplings across hydrogen bonds. *Methods Enzymol*, 338:111–33, 2001.

[84] J. Gsponer, M. E. Futschik, S. A. Teichmann, and M. M. Babu. Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. *Science*, 322(5906):1365–8, Nov. 2008.

[85] K. Gupta, P. Kumar, and D. Chatterji. Identification, activity and disulfide connectivity of c-di-gmp regulating proteins in mycobacterium tuberculosis. *PLoS One*, 5(11):e15072, 2010.

[86] J. Habazettl, M. G. Allan, U. Jenal, and S. Grzesiek. Solution structure of the pilz domain protein pa4608 complex with cyclic di-gmp identifies charge clustering as molecular readout. *J Biol Chem*, 286(16):14304–14, Apr. 2011.

[87] H. Han, C. L. Cliff, and L. H. Hurley. Accelerated assembly of g-quadruplex structures by a small molecule. *Biochemistry*, 38(22):6981–6, June 1999.

[88] D. F. Hansen, P. Vallurupalli, and L. E. Kay. Using relaxation dispersion nmr spectroscopy to determine structures of excited, invisible protein states. *J Biomol NMR*, 41(3):113–20, July 2008.

[89] K. Hasegawa, S. Masuda, and T. aki Ono. Light induced structural changes of a full-length protein and its bluf domain in ycgf(blrp), a blue-light sensing protein that uses fad (bluf). *Biochemistry*, 45(11):3785–93, Mar. 2006.

[90] P. Hasson, B. Müller, K. Basler, and Z. Paroush. Brinker requires two corepressors for maximal and versatile repression in dpp signalling. *EMBO J*, 20(20):5725–36, Oct. 2001.

[91] Y. Hayakawa, R. Nagata, A. Hirata, M. Hyodo, and R. Kawai. A facile synthesis of cyclic bis (3'–> 5') diguanylic acid. *Tetrahedron*, 59(34):6465–6471, 2003.

[92] W. Haynes. *Handbook of chemistry and physics, 92nd ed.* CRC Press: Boca Raton, FL, 2011.

[93] B. He, K. Wang, Y. Liu, B. Xue, V. N. Uversky, and A. K. Dunker. Predicting intrinsic disorder in proteins: an overview. *Cell Res*, 19(8):929–49, Aug. 2009.

[94] R. Hengge. Principles of c-di-gmp signalling in bacteria. *Nat Rev Microbiol*, 7(4):263–73, Apr. 2009.

[95] J. W. Hickman and C. S. Harwood. Identification of fleq from pseudomonas aeruginosa as a c-di-gmp-responsive transcription factor. *Mol Microbiol*, 69(2):376–89, July 2008.

[96] F. Hofmeister. Zur lehre von der wirkung der salze. *Arch Exp Pathol Pharmakol*, 24:247–60, 1888.

[97] O. Howarth and D. Lilley. Carbon-13-nmr of peptides and proteins. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 12(1):1–40, 1978.

[98] Y. Huang and Z. Liu. Kinetic advantage of intrinsically disordered proteins in coupled folding-binding process: a critical assessment of the "fly-casting" mechanism. *J Mol Biol*, 393(5):1143–59, Nov. 2009.

[99] N. V. Hud, P. Schultze, V. Sklenar, and J. Feigon. Binding sites and dynamics of ammonium ions in a telomere repeat dna quadruplex. *J Mol Biol*, 285(1):233–43, Jan. 1999.

[100] M. A. Huntley and G. B. Golding. Simple sequences are rare in the protein data bank. *Proteins*, 48(1):134–40, July 2002.

[101] J. L. Huppert and S. Balasubramanian. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res*, 33(9):2908–16, 2005.

[102] J. L. Huppert and S. Balasubramanian. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res*, 35(2):406–13, 2007.

[103] J. H. Hurley. Gaf domains: cyclic nucleotides come full circle. *Sci STKE*, 2003(164):PE1, Jan. 2003.

[104] T. Hwang and A. Shaka. Water suppression that works. excitation sculpting using arbitrary wave-forms and pulsed-field gradients. *Journal of Magnetic Resonance, Series A*, 112(2):275–279, 1995.

[105] T. I. Igumenova, U. Brath, M. Akke, and A. G. Palmer. Characterization of chemical exchange using residual dipolar coupling. *J Am Chem Soc*, 129(44):13396–7, Nov. 2007.

[106] R. Ishima and D. A. Torchia. Extending the range of amide proton relaxation dispersion experiments in proteins using a constant-time relaxation-compensated cpmg approach. *J Biomol NMR*, 25(3):243–8, Mar. 2003.

[107] H. Ito, S. Tanaka, and M. Miyasaka. Circular dichroism spectra of dna quadruplexes [d(g(5)t(5))](4) as formed with g(4) and t(4) tetrads and [d(g(5)t(5)). d(a(5)c(5))]2 as formed with watson-crick-like (g-c)(2) and (t-a)(2) tetrads. *Biopolymers*, 65(2):61–80, Oct. 2002.

[108] J. Iwahara and G. M. Clore. Detecting transient intermediates in macromolecular binding by paramagnetic nmr. *Nature*, 440(7088):1227–30, Apr. 2006.

[109] J. Iwahara and G. M. Clore. Direct observation of enhanced translocation of a homeodomain between dna cognate sites by nmr exchange spectroscopy. *J Am Chem Soc*, 128(2):404–5, Jan. 2006.

[110] J. Iwahara, M. Zweckstetter, and G. M. Clore. Nmr structural and kinetic characterization of a homeodomain diffusing and hopping on nonspecific dna. *Proc Natl Acad Sci U S A*, 103(41):15062–7, Oct. 2006.

[111] S. E. Jackson. How do small single-domain proteins fold? *Fold Des*, 3(4):R81–91, 1998.

[112] W. Jahnke, R. M. Grotzfeld, X. Pella, A. Strauss, G. Fendrich, S. W. Cowan-Jacob, S. Cotesta, D. Fabbro, P. Furet, J. Mestan, and A. L. Marzinzik. Binding or bending: distinction of allosteric abl kinase agonists from antagonists by an nmr-based conformational assay. *J Am Chem Soc*, 132(20):7043–8, May 2010.

[113] A. Jazwinska, N. Kirov, E. Wieschaus, S. Roth, and C. Rushlow. The drosophila gene brinker reveals a novel mechanism of dpp target gene regulation. *Cell*, 96(4):563–73, Feb. 1999.

[114] A. Jazwinska, C. Rushlow, and S. Roth. The role of brinker in mediating the graded response to dpp in early drosophila embryos. *Development*, 126(15):3323–34, Aug. 1999.

[115] U. Jenal and J. Malone. Mechanisms of cyclic-di-gmp signaling in bacteria. *Annu Rev Genet*, 40:385–407, 2006.

[116] M. Jimenez, J. Nieto, M. Rico, J. Santoro, J. Herranz, and F. Bermejo. A study of the nh nmr signals of gly-gly-x-ala tetrapeptides in h2o at low temperature. *Journal of Molecular Structure*, 143:435–438, 1986.

[117] B. Johnson and R. Blevins. Nmr view: A computer program for the visualization and analysis of nmr data. *Journal of Biomolecular NMR*, 4(5):603–614, 1994.

[118] W. Jorgensen, J. Chandrasekhar, J. Madura, R. Impey, M. Klein, et al. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983.

[119] D. K. R. Karaolis, T. K. Means, D. Yang, M. Takahashi, T. Yoshimura, E. Muraille, D. Philpott, J. T. Schroeder, M. Hyodo, Y. Hayakawa, B. G. Talbot, E. Brouillette, and F. Malouin. Bacterial c-di-gmp is an immunostimulatory molecule. *J Immunol*, 178(4):2171–81, Feb. 2007.

[120] M. Karplus. Contact electron-spin coupling of nuclear magnetic moments. *J Chem Phys*, 30:11–15, 1959.

[121] M. Karplus. Vicinal proton coupling in nuclear magnetic resonance. *J Am Chem Soc*, 85(18):2870–2871, 1963.

[122] M. Karplus. Behind the folding funnel diagram. *Nat Chem Biol*, 7(7):401–4, July 2011.

[123] R. Kawai, R. Nagata, A. Hirata, and Y. Hayakawa. A new synthetic approach to cyclic bis(3'–>5')diguanylic acid. *Nucleic Acids Res Suppl*, (3):103–4, 2003.

[124] B. I. Kazmierczak, M. B. Lebron, and T. S. Murray. Analysis of fimx, a phosphodiesterase that governs twitching motility in pseudomonas aeruginosa. *Mol Microbiol*, 60(4):1026–43, May 2006.

[125] P. Keim, R. A. Vigna, A. M. Nigen, J. S. Morrow, and F. R. Gurd. Carbon 13 nuclear magnetic resonance of pentapeptides of glycine containing central residues of methionine, proline, arginine, and lysine. *J Biol Chem*, 249(13):4149–56, July 1974.

[126] I. Kelsey, S. Nakayama, and H. O. Sintim. Diamidinium and iminium aromatics as new aggregators of the bacterial signaling molecule, c-di-gmp. *Bioorg Med Chem Lett*, 22(2):881–5, Jan. 2012.

[127] I. Kiburu, A. Shurer, L. Yan, and H. O. Sintim. A simple solid-phase synthesis of the ubiquitous bacterial signaling molecule, c-di-gmp and analogues. *Mol Biosyst*, 4(6):518–20, June 2008.

[128] T. Kiefhaber, A. Bachmann, and K. S. Jensen. Dynamics and mechanisms of coupled protein folding and binding reactions. *Curr Opin Struct Biol*, 22(1):21–9, Feb. 2012.

[129] J. Ko, K.-S. Ryu, H. Kim, J.-S. Shin, J.-O. Lee, C. Cheong, and B.-S. Choi. Structure of pp4397 reveals the molecular basis for different c-di-gmp binding modes by pilz domain proteins. *J Mol Biol*, 398(1):97–110, Apr. 2010.

[130] D. M. Korzhnev, T. L. Religa, W. Banachewicz, A. R. Fersht, and L. E. Kay. A transient and low-populated protein-folding intermediate at atomic resolution. *Science*, 329(5997):1312–6, Sept. 2010.

[131] D. M. Korzhnev, X. Salvatella, M. Vendruscolo, A. A. Di Nardo, A. R. Davidson, C. M. Dobson, and L. E. Kay. Low-populated folding intermediates of fyn sh3 characterized by relaxation dispersion nmr. *Nature*, 430(6999):586–90, July 2004.

[132] P. V. Krasteva, J. C. N. Fong, N. J. Shikuma, S. Beyhan, M. V. A. S. Navarro, F. H. Yildiz, and H. Sondermann. Vibrio cholerae vpst regulates matrix production and motility by directly sensing cyclic di-gmp. *Science*, 327(5967):866–8, Feb. 2010.

[133] A. Krezel, R. Latajka, G. D. Bujacz, and W. Bal. Coordination properties of tris(2-carboxyethyl)phosphine, a newly introduced thiol reductant, and its oxide. *Inorg Chem*, 42(6):1994–2003, Mar. 2003.

[134] M. Kumar and D. Chatterji. Cyclic di-gmp: a second messenger required for long-term survival, but not for biofilm formation, in mycobacterium smegmatis. *Microbiology*, 154(Pt 10):2942–55, Oct. 2008.

[135] Y. Kuroda, S. Kidokoro, and A. Wada. Thermodynamic characterization of cytochrome c at low ph. observation of the molten globule state and of the cold denaturation process. *J Mol Biol*, 223(4):1139–53, Feb. 1992.

[136] A. N. Lane, J. B. Chaires, R. D. Gray, and J. O. Trent. Stability and kinetics of g-quadruplex structures. *Nucleic Acids Res*, 36(17):5482–515, Oct. 2008.

[137] P. P. Lankhorst, C. A. Haasnoot, C. Erkelens, and C. Altona. Carbon-13 nmr in conformational analysis of nucleic acid fragments. 2. a reparametrization of the karplus equation for vicinal nmr coupling constants in ccop and hcop fragments. *J Biomol Struct Dyn*, 1(6):1387–405, June 1984.

[138] P. P. Lankhorst, C. A. Haasnoot, C. Erkelens, and C. Altona. Carbon-13 nmr in conformational analysis of nucleic acid fragments. 3. the magnitude of torsional angle epsilon in d(tpa) from ccop and hcop nmr coupling constants. *Nucleic Acids Res*, 12(13):5419–28, July 1984.

[139] J. L. Leduc and G. P. Roberts. Cyclic di-gmp allosterically inhibits the crp-like protein (clp) of xanthomonas axonopodis pv. citri. *J Bacteriol*, 191(22):7121–2, Nov. 2009.

[140] E. R. Lee, J. L. Baker, Z. Weinberg, N. Sudarsan, and R. R. Breaker. An allosteric self-splicing ribozyme triggered by a bacterial second messenger. *Science*, 329(5993):845–8, Aug. 2010.

[141] M. Levet-Paulo, J.-C. Lazzaroni, C. Gilbert, D. Atlan, P. Doublet, and A. Vianney. The atypical two-component sensor kinase lpl0330 from legionella pneumophila controls the bifunctional diguanylate cyclase-phosphodiesterase lpl0329 to modulate bis-(3'-5')-cyclic dimeric gmp synthesis. *J Biol Chem*, 286(36):31136–44, Sept. 2011.

[142] C. Levinthal. How to fold graciously. In *Mossbauer Spectroscopy in Biological Systems Proceedings of a meeting held at Allerton House*, pages 22–24. University of Illinois Press, 1969.

[143] L.-Y. Liang and G. Roberts, editors. *Protein NMR Spectroscopy: Practical Techniques and Applications.* John Wiley & Sons Ltd, first edition edition, 2011.

[144] Y. C. Liaw, Y. G. Gao, H. Robinson, G. M. Sheldrick, L. A. Sliedregt, G. A. van der Marel, J. H. van Boom, and A. H. Wang. Cyclic diguanylic acid behaves as a host molecule for planar intercalators. *FEBS Lett*, 264(2):223–7, May 1990.

[145] K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw. How fast-folding proteins fold. *Science*, 334(6055):517–20, Oct. 2011.

[146] G. Lipari and A. Szabo. Model-free approach to the interpretation of nuclear magnetic relaxation in macromolecules. 2. analysis of experimental results. *J. Am. Chem. Soc*, 104:4559–4570, 1982.

[147] G. Lipari and A. Szabo. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. theory and range of validity. *Journal of the American Chemical Society*, 104(17):4546–4559, 1982.

[148] J. Liu, N. B. Perumal, C. J. Oldfield, E. W. Su, V. N. Uversky, and A. K. Dunker. Intrinsic disorder in transcription factors. *Biochemistry*, 45(22):6873–88, June 2006.

[149] N. Liu, Y. Xu, S. Hossain, N. Huang, D. Coursolle, J. A. Gralnick, and E. M. Boon. Nitric oxide regulation of cyclic di-gmp synthesis and hydrolysis in shewanella woodyi. *Biochemistry*, 51(10):2087–99, Mar. 2012.

[150] J. Loria, M. Rance, and A. Palmer III. A relaxation-compensated carr-purcell-meiboom-gill sequence for characterizing chemical exchange by nmr spectroscopy. *Journal of the American Chemical Society*, 121(10):2331–2332, 1999.

[151] Y. Luo, J. Zhou, S. K. Watt, V. T. Lee, T. K. Dayie, and H. O. Sintim. Differential binding of 2'-biotinylated analogs of c-di-gmp with c-di-gmp riboswitches and binding proteins. *Mol Biosyst*, 8(3):772–8, Mar. 2012.

[152] A. MacKerell Jr, D. Bashford, M. Bellott, R. Dunbrack Jr, J. Evanseck, M. Field, S. Fischer, J. Gao, H. Guo, S. Ha, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B*, 102(18):3586–3616, 1998.

[153] J. L. Markley, A. Bax, Y. Arata, C. W. Hilbers, R. Kaptein, B. D. Sykes, P. E. Wright, and K. Wüthrich. Recommendations for the presentation of nmr structures of proteins and nucleic acids–iupac-iubmb-iupab inter-union task group on the standardization of data bases of protein and nucleic acid structures determined by nmr spectroscopy. *Eur J Biochem*, 256(1):1–15, Aug. 1998.

[154] A. Marlow, E. Mezzina, G. Spada, S. Masiero, J. Davis, and G. Gottarelli. Cation-templated self-assembly of a lipophilic deoxyguanosine: Solution structure of a k+-dg8 octamer. *The Journal of Organic Chemistry*, 64(14):5116–5123, 1999.

[155] T. Marty, B. Müller, K. Basler, and M. Affolter. Schnurri mediates dpp-dependent repression of brinker transcription. *Nat Cell Biol*, 2(10):745–9, Oct. 2000.

[156] S. Meiboom and D. Gill. Modified spin-echo method for measuring nuclear relaxation times. *Review of Scientific Instruments*, 29(8):688–691, 1958.

[157] S. Meier, M. Blackledge, and S. Grzesiek. Conformational distributions of unfolded polypeptides from novel nmr techniques. *J Chem Phys*, 128(5):052204, Feb. 2008.

[158] S. Meier, S. Grzesiek, and M. Blackledge. Mapping the conformational landscape of urea-denatured ubiquitin using residual dipolar couplings. *J Am Chem Soc*, 129(31):9799–807, Aug. 2007.

[159] S. Meier, M. Strohmeier, M. Blackledge, and S. Grzesiek. Direct observation of dipolar couplings and hydrogen bonds across a beta-hairpin in 8 m urea. *J Am Chem Soc*, 129(4):754–5, Jan. 2007.

[160] A. Meissner, V. Wild, R. Simm, M. Rohde, C. Erck, F. Bredenbruch, M. Morr, U. Römling, and S. Häussler. Pseudomonas aeruginosa cupa-encoded fimbriae expression is regulated by a ggdef and eal domain-dependent modulation of the intracellular level of cyclic diguanylate. *Environ Microbiol*, 9(10):2475–85, Oct. 2007.

[161] M. Merighi, V. T. Lee, M. Hyodo, Y. Hayakawa, and S. Lory. The second messenger bis-(3'-5')-cyclic-gmp and its pilz domain-containing receptor alg44 are required for alginate biosynthesis in pseudomonas aeruginosa. *Mol Microbiol*, 65(4):876–95, Aug. 2007.

[162] J. H. Merritt, K. M. Brothers, S. L. Kuchma, and G. A. O'Toole. Sadc reciprocally influences biofilm formation and swarming motility via modulation of exopolysaccharide production and flagellar function. *J Bacteriol*, 189(22):8154–64, Nov. 2007.

[163] G. Merutka, H. J. Dyson, and P. E. Wright. 'random coil' 1h chemical shifts obtained as a function of temperature and trifluoroethanol concentration for the peptide series ggxgg. *J Biomol NMR*, 5(1):14–24, Jan. 1995.

[164] O. Millet, J. Loria, C. Kroenke, M. Pons, and A. Palmer III. The static magnetic field dependence of chemical exchange linebroadening defines the nmr chemical shift time scale. *Journal of the American Chemical Society*, 122(12):2867–2877, 2000.

[165] E. Mills, I. S. Pultz, H. D. Kulasekara, and S. I. Miller. The bacterial second messenger c-di-gmp: mechanisms of signalling. *Cell Microbiol*, 13(8):1122–9, Aug. 2011.

[166] G. Minasov, S. Padavattan, L. Shuvalova, J. S. Brunzelle, D. J. Miller, A. Basle, C. Massa, F. R. Collart, T. Schirmer, and W. F. Anderson. Crystal structures of ykui and its complex with second messenger cyclic di-gmp suggest catalytic mechanism of phosphodiester bond cleavage by eal domains. *J Biol Chem*, 284(19):13174–84, May 2009.

[167] R. D. Monds, P. D. Newell, R. H. Gross, and G. A. O'Toole. Phosphate-dependent modulation of c-di-gmp levels regulates pseudomonas fluorescens pf0-1 biofilm formation by controlling secretion of the adhesin lapa. *Mol Microbiol*, 63(3):656–79, Feb. 2007.

[168] M. M. Mooren, S. S. Wijmenga, G. A. van der Marel, J. H. van Boom, and C. W. Hilbers. The solution structure of the circular trinucleotide cr(gpgpgp) determined by nmr and molecular mechanics calculation. *Nucleic Acids Res*, 22(13):2658–66, July 1994.

[169] J. A. Moscoso, H. Mikkelsen, S. Heeb, P. Williams, and A. Filloux. The pseudomonas aeruginosa sensor rets switches type iii and type vi secretion via c-di-gmp signalling. *Environ Microbiol*, 13(12):3128–38, Dec. 2011.

[170] M. Moser and G. Campbell. Generating and interpreting the brinker gradient in the drosophila wing. *Dev Biol*, 286(2):647–58, Oct. 2005.

[171] J. Moult. A decade of casp: progress, bottlenecks and prognosis in protein structure prediction. *Curr Opin Struct Biol*, 15(3):285–9, June 2005.

[172] F. A. Mulder, N. R. Skrynnikov, B. Hon, F. W. Dahlquist, and L. E. Kay. Measurement of slow (micros-ms) time scale dynamics in protein side chains by (15)n relaxation dispersion nmr spectroscopy: application to asn and gln residues in a cavity mutant of t4 lysozyme. *J Am Chem Soc*, 123(5):967–75, Feb. 2001.

[173] F. A. A. Mulder, B. Hon, A. Mittermaier, F. W. Dahlquist, and L. E. Kay. Slow internal dynamics in proteins: application of nmr relaxation dispersion spectroscopy to methyl groups in a cavity mutant of t4 lysozyme. *J Am Chem Soc*, 124(7):1443–51, Feb. 2002.

[174] B. Müller, B. Hartmann, G. Pyrowolakis, M. Affolter, and K. Basler. Conversion of an extracellular dpp/bmp morphogen gradient into an inverse transcriptional gradient. *Cell*, 113(2):221–33, Apr. 2003.

[175] Y. Nakasone, T. aki Ono, A. Ishii, S. Masuda, and M. Terazima. Transient dimerization and conformational change of a bluf protein: Ycgf. *J Am Chem Soc*, 129(22):7028–35, June 2007.

[176] S. Nakayama, I. Kelsey, J. Wang, K. Roelofs, B. Stefane, Y. Luo, V. T. Lee, and H. O. Sintim. Thiazole orange-induced c-di-gmp quadruplex formation facilitates a simple fluorescent detection of this ubiquitous biofilm regulating molecule. *J Am Chem Soc*, 133(13):4856–64, Apr. 2011.

[177] S. Nakayama, I. Kelsey, J. Wang, and H. O. Sintim. c-di-gmp can form remarkably stable g-quadruplexes at physiological conditions in the presence of some planar intercalators. *Chem Commun (Camb)*, 47(16):4766–8, Apr. 2011.

[178] S. Nakayama, K. Roelofs, V. T. Lee, and H. O. Sintim. A c-di-gmp-proflavine-hemin supramolecular complex has peroxidase activity-implication for a simple colorimetric detection. *Mol Biosyst*, 8(3):726–9, Mar. 2012.

[179] A. Nakhamchik, C. Wilde, and D. A. Rowe-Magnus. Cyclic-di-gmp regulates extracellular polysaccharide production, biofilm formation, and rugose colony development by vibrio vulnificus. *Appl Environ Microbiol*, 74(13):4199–209, July 2008.

[180] M. V. A. S. Navarro, N. De, N. Bae, Q. Wang, and H. Sondermann. Structural analysis of the ggdef-eal domain-containing c-di-gmp receptor fimx. *Structure*, 17(8):1104–16, Aug. 2009.

[181] R. Nelson, M. R. Sawaya, M. Balbirnie, A. O. Madsen, C. Riekel, R. Grothe, and D. Eisenberg. Structure of the cross-beta spine of amyloid-like fibrils. *Nature*, 435(7043):773–8, June 2005.

[182] D. Neri, M. Billeter, G. Wider, and K. Wüthrich. Nmr determination of residual structure in a urea-denatured protein, the 434-repressor. *Science*, 257(5076):1559–63, Sept. 1992.

[183] A. D. Ogunniyi, J. C. Paton, A. C. Kirby, J. A. McCullers, J. Cook, M. Hyodo, Y. Hayakawa, and D. K. R. Karaolis. c-di-gmp is an effective immunomodulator and vaccine adjuvant against pneumococcal infection. *Vaccine*, 26(36):4676–85, Aug. 2008.

[184] J. N. Onuchic, P. G. Wolynes, Z. Luthey-Schulten, and N. D. Socci. Toward an outline of the topography of a realistic protein-folding funnel. *Proc Natl Acad Sci U S A*, 92(8):3626–30, Apr. 1995.

[185] V. Y. Orekhov, D. M. Korzhnev, and L. E. Kay. Double- and zero-quantum nmr relaxation dispersion experiments sampling millisecond time scale dynamics in proteins. *J Am Chem Soc*, 126(6):1886–91, Feb. 2004.

[186] G. Otting, Y. Q. Qian, M. Billeter, M. Müller, M. Affolter, W. J. Gehring, and K. Wüthrich. Protein–dna contacts in the structure of a homeodomain–dna complex determined by nuclear magnetic resonance spectroscopy in solution. *EMBO J*, 9(10):3085–92, Oct. 1990.

[187] A. G. Palmer, C. D. Kroenke, and J. P. Loria. Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol*, 339:204–38, 2001.

[188] L. M. Pegram and M. T. Record. Thermodynamic origin of hofmeister ion effects. *J Phys Chem B*, 112(31):9428–36, Aug. 2008.

[189] L. M. Pegram, T. Wendorff, R. Erdmann, I. Shkel, D. Bellissimo, D. J. Felitsky, and M. T. Record. Why hofmeister effects of many salts favor protein folding but not dna helix formation. *Proc Natl Acad Sci U S A*, 107(17):7716–21, Apr. 2010.

[190] J. Peng and G. Wagner. Mapping of spectral density functions using heteronuclear nmr relaxation measurements. *J Magn Reson*, (98):204–238, 1992.

[191] D. Perez-Mendoza, S. J. Coulthurst, S. Humphris, E. Campbell, M. Welch, I. K. Toth, and G. P. C. Salmond. A multi-repeat adhesin of the phytopathogen, pectobacterium atrosepticum, is secreted by a type i pathway and is subject to complex regulation involving a non-canonical diguanylate cyclase. *Mol Microbiol*, 82(3):719–33, Nov. 2011.

[192] K. Pervushin, A. Ono, C. Fernandez, T. Szyperski, M. Kainosho, and K. Wüthrich. Nmr scalar couplings across watson-crick base pair hydrogen bonds in dna observed by transverse relaxation-optimized spectroscopy. *Proc Natl Acad Sci U S A*, 95(24):14147–51, Nov. 1998.

[193] C. Pesavento, G. Becker, N. Sommerfeldt, A. Possling, N. Tschowri, A. Mehlis, and R. Hengge. Inverse regulatory coordination of motility and curli-mediated adhesion in escherichia coli. *Genes Dev*, 22(17):2434–46, Sept. 2008.

[194] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten. Scalable molecular dynamics with namd. *J Comput Chem*, 26(16):1781–802, Dec. 2005.

[195] M. Piotto, V. Saudek, and V. Sklenar. Gradient-tailored excitation for single-quantum nmr spectroscopy of aqueous solutions. *J Biomol NMR*, 2(6):661–5, Nov. 1992.

[196] K. W. Plaxco and C. M. Dobson. Time-resolved biophysical methods in the study of protein folding. *Curr Opin Struct Biol*, 6(5):630–6, Oct. 1996.

[197] K. W. Plaxco, C. J. Morton, S. B. Grimshaw, J. A. Jones, M. Pitkeathly, I. D. Campbell, and C. M. Dobson. The effects of guanidine hydrochloride on the 'random coil' conformations and nmr chemical shifts of the peptide series ggxgg. *J Biomol NMR*, 10(3):221–30, Oct. 1997.

[198] P. Podbevsek, N. V. Hud, and J. Plavec. Nmr evaluation of ammonium ion movement within a unimolecular g-quadruplex in solution. *Nucleic Acids Res*, 35(8):2554–63, 2007.

[199] J. T. Pratt, R. Tamayo, A. D. Tischler, and A. Camilli. Pilz domain proteins bind cyclic diguanylate and regulate diverse processes in vibrio cholerae. *J Biol Chem*, 282(17):12860–70, Apr. 2007.

[200] G. Pyrowolakis, B. Hartmann, B. Müller, K. Basler, and M. Affolter. A simple molecular complex mediates widespread bmp-induced repression during drosophila development. *Dev Cell*, 7(2):229–40, Aug. 2004.

[201] F. Qin, Y. Jiang, Y. Chen, M. Wu, G. Yan, W. Ye, Y. Li, J. Zhang, and H.-F. Chen. Conformational selection or induced fit for brinker and dna recognition. *Phys Chem Chem Phys*, 13(4):1407–12, Jan. 2011.

[202] E. Rennella, T. Cutuil, P. Schanda, I. Ayala, V. Forge, and B. Brutscher. Real-time nmr characterization of structure and dynamics in a transiently populated protein folding intermediate. *J Am Chem Soc*, 134(19):8066–9, May 2012.

[203] R. Richarz and K. Wüthrich. Carbon-13 nmr chemical shifts of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides h-gly-gly-x-l-ala-oh. *Biopolymers*, 17(9):2133–2141, 1978.

[204] H. Roder. Structural characterization of protein folding intermediates by proton magnetic resonance and hydrogen exchange. *Methods Enzymol*, 176:446–73, 1989.

[205] H. Roder, K. Maki, and H. Cheng. Early events in protein folding explored by rapid mixing methods. *Chem Rev*, 106(5):1836–61, May 2006.

[206] U. Römling, M. Gomelsky, and M. Y. Galperin. C-di-gmp: the dawning of a novel bacterial signalling system. *Mol Microbiol*, 57(3):629–39, Aug. 2005.

[207] J. rong Huang, F. Gabel, M. R. Jensen, S. Grzesiek, and M. Blackledge. Sequence-specific mapping of the interaction between urea and unfolded ubiquitin from ensemble analysis of nmr and small angle scattering data. *J Am Chem Soc*, 134(9):4429–36, Mar. 2012.

[208] P. Ross, R. Mayer, H. Weinhouse, D. Amikam, Y. Huggirat, M. Benziman, E. De Vroom, A. Fidder, P. De Paus, and L. Sliedregt. The cyclic diguanylic acid regulatory system of cellulose synthesis in acetobacter xylinum. chemical synthesis and biological activity of cyclic nucleotide dimer, trimer, and phosphothioate derivatives. *Journal of Biological Chemistry*, 265(31):18933–18943, 1990.

[209] P. Ross, H. Weinhouse, Y. Aloni, D. Michaeli, P. Weinberger-Ohana, R. Mayer, S. Braun, E. de Vroom, G. A. van der Marel, J. H. van Boom, and M. Benziman.

Regulation of cellulose synthesis in acetobacter xylinum by cyclic diguanylic acid. *Nature*, 325(6101):279–81, Jan. 1987.

[210] L. M. Ruiz, M. Castro, A. Barriga, C. A. Jerez, and N. Guiliani. The extremophile acidithiobacillus ferrooxidans possesses a c-di-gmp signalling pathway that could play a significant role during bioleaching of minerals. *Lett Appl Microbiol*, 54(2):133–9, Feb. 2012.

[211] R. P. Ryan, Y. McCarthy, M. Andrade, C. S. Farah, J. P. Armitage, and J. M. Dow. Cell-cell signal-dependent dynamic interactions between hd-gyp and ggdef domain proteins mediate virulence in xanthomonas campestris. *Proc Natl Acad Sci U S A*, 107(13):5989–94, Mar. 2010.

[212] D. A. Ryjenkov, R. Simm, U. Römling, and M. Gomelsky. The pilz domain is a receptor for the second messenger c-di-gmp: the pilz domain protein ycgr controls motility in enterobacteria. *J Biol Chem*, 281(41):30310–4, Oct. 2006.

[213] E. Saller and M. Bienz. Direct competition between brinker and drosophila mad in dpp target gene transcription. *EMBO Rep*, 2(4):298–305, Apr. 2001.

[214] Y. Sasakura, S. Hirata, S. Sugiyama, S. Suzuki, S. Taguchi, M. Watanabe, T. Matsui, I. Sagami, and T. Shimizu. Characterization of a direct oxygen sensor heme protein from escherichia coli. effects of the heme redox states and mutations at the heme-binding site on catalysis and structure. *J Biol Chem*, 277(26):23821–7, June 2002.

[215] H. J. Sass, G. Musco, S. J. Stahl, P. T. Wingfield, and S. Grzesiek. Solution nmr of proteins within polyacrylamide gels: diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes. *J Biomol NMR*, 18(4):303–9, Dec. 2000.

[216] T. Schirmer and U. Jenal. Structural and mechanistic determinants of c-di-gmp signalling. *Nat Rev Microbiol*, 7(10):724–35, Oct. 2009.

[217] P. Schultze, N. V. Hud, F. W. Smith, and J. Feigon. The effect of sodium, potassium and ammonium ions on the conformation of the dimeric quadruplex formed by the oxytricha nova telomere repeat oligonucleotide d(g(4)t(4)g(4)). *Nucleic Acids Res*, 27(15):3018–28, Aug. 1999.

[218] S. Schwarzinger, G. J. Kroon, T. R. Foss, J. Chung, P. E. Wright, and H. J. Dyson. Sequence-dependent correction of random coil nmr chemical shifts. *J Am Chem Soc*, 123(13):2970–8, Apr. 2001.

[219] S. Schwarzinger, G. J. Kroon, T. R. Foss, P. E. Wright, and H. J. Dyson. Random coil chemical shifts in acidic 8 m urea: implementation of random coil shift data in nmrview. *J Biomol NMR*, 18(1):43–8, Sept. 2000.

[220] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, and W. Wriggers. Atomic-level characterization of the structural dynamics of proteins. *Science*, 330(6002):341–6, Oct. 2010.

[221] Y. Shen, O. Lange, F. Delaglio, P. Rossi, J. M. Aramini, G. Liu, A. Eletsky, Y. Wu, K. K. Singarapu, A. Lemak, A. Ignatchenko, C. H. Arrowsmith, T. Szyperski, G. T. Montelione, D. Baker, and A. Bax. Consistent blind protein structure generation from nmr chemical shift data. *Proc Natl Acad Sci U S A*, 105(12):4685–90, Mar. 2008.

[222] J.-S. Shin, K.-S. Ryu, J. Ko, A. Lee, and B.-S. Choi. Structural characterization reveals that a pilz domain protein undergoes substantial conformational change upon binding to cyclic dimeric guanosine monophosphate. *Protein Sci*, 20(2):270–7, Feb. 2011.

[223] B. A. Shoemaker, J. J. Portman, and P. G. Wolynes. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc Natl Acad Sci U S A*, 97(16):8868–73, Aug. 2000.

[224] D. Shortle and M. S. Ackerman. Persistence of native-like topology in a denatured protein in 8 m urea. *Science*, 293(5529):487–9, July 2001.

[225] R. Simm, M. Morr, U. Remminghorst, M. Andersson, and U. Römling. Quantitative determination of cyclic diguanosine monophosphate concentrations in nucleotide extracts of bacteria by matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry. *Anal Biochem*, 386(1):53–8, Mar. 2009.

[226] N. R. Skrynnikov, F. W. Dahlquist, and L. E. Kay. Reconstructing nmr spectra of "invisible" excited protein states using hsqc and hmqc experiments. *J Am Chem Soc*, 124(41):12352–60, Oct. 2002.

[227] N. R. Skrynnikov, F. A. Mulder, B. Hon, F. W. Dahlquist, and L. E. Kay. Probing slow time scale dynamics at methyl-containing side chains in proteins by relaxation dispersion nmr measurements: application to methionine residues in a cavity mutant of t4 lysozyme. *J Am Chem Soc*, 123(19):4556–66, May 2001.

[228] K. D. Smith, S. V. Lipchock, T. D. Ames, J. Wang, R. R. Breaker, and S. A. Strobel. Structural basis of ligand binding by a c-di-gmp riboswitch. *Nat Struct Mol Biol*, 16(12):1218–23, Dec. 2009.

[229] K. D. Smith, C. A. Shanahan, E. L. Moore, A. C. Simon, and S. A. Strobel. Structural basis of differential ligand recognition by two classes of bis-(3'-5')-cyclic dimeric guanosine monophosphate-binding riboswitches. *Proc Natl Acad Sci U S A*, 108(19):7757–62, May 2011.

[230] V. Spehr, R. Warrass, K. Höcherl, and T. Ilg. Large-scale production of the immunomodulator c-di-gmp from gmp and atp by an enzymatic cascade. *Appl Biochem Biotechnol*, 165(3-4):761–75, Oct. 2011.

[231] N. Sreerama and R. W. Woody. Computation and analysis of protein circular dichroism spectra. *Methods Enzymol*, 383:318–51, 2004.

[232] N. Sudarsan, E. R. Lee, Z. Weinberg, R. H. Moy, J. N. Kim, K. H. Link, and R. R. Breaker. Riboswitches in eubacteria sense the second messenger cyclic di-gmp. *Science*, 321(5887):411–3, July 2008.

[233] H. Surprenant, J. Sarneski, R. Key, J. Byrd, and C. Reilley. Carbon-13 nmr studies of amino acids: chemical shifts, protonation shifts, microscopic protonation behavior. *Journal of Magnetic Resonance (1969)*, 40(2):231–243, 1980.

[234] H. Takahashi and T. Shimizu. Phosphodiesterase activity of ec dos, a heme-regulated enzyme from escherichia coli, toward 3', 5'-cyclic diguanylic acid is obviously enhanced by o2 and co binding. *Chemistry Letters*, 35(8):970–971, 2006.

[235] R. Tal, H. C. Wong, R. Calhoon, D. Gelfand, A. L. Fear, G. Volman, R. Mayer, P. Ross, D. Amikam, H. Weinhouse, A. Cohen, S. Sapir, P. Ohana, and M. Benziman. Three cdg operons control cellular turnover of cyclic di-gmp in acetobacter xylinum: genetic organization and occurrence of conserved domains in isoenzymes. *J Bacteriol*, 180(17):4416–25, Sept. 1998.

[236] R. Tamayo, J. T. Pratt, and A. Camilli. Roles of cyclic diguanylate in the regulation of bacterial pathogenesis. *Annu Rev Microbiol*, 61:131–48, 2007.

[237] R. Tamayo, A. D. Tischler, and A. Camilli. The eal domain protein viea is a cyclic diguanylate phosphodiesterase. *J Biol Chem*, 280(39):33324–30, Sept. 2005.

[238] K. Tamiola, B. Acar, and F. A. A. Mulder. Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *J Am Chem Soc*, 132(51):18000–3, Dec. 2010.

[239] F. Tao, Y.-W. He, D.-H. Wu, S. Swarup, and L.-H. Zhang. The cyclic nucleotide monophosphate domain of xanthomonas campestris global regulator clp defines a new class of cyclic di-gmp effectors. *J Bacteriol*, 192(4):1020–9, Feb. 2010.

[240] V. Thanabal, D. O. Omecinsky, M. D. Reily, and W. L. Cody. The 13c chemical shifts of amino acids in aqueous solution containing organic solvents: application to the secondary structure characterization of peptides in aqueous trifluoroethanol solution. *J Biomol NMR*, 4(1):47–59, Jan. 1994.

[241] N. Tjandra, J. G. Omichinski, A. M. Gronenborn, G. M. Clore, and A. Bax. Use of dipolar 1h-15n and 1h-13c couplings in the structure determination of magnetically oriented macromolecules in solution. *Nat Struct Biol*, 4(9):732–8, Sept. 1997.

[242] M. Tollinger, N. R. Skrynnikov, F. A. Mulder, J. D. Forman-Kay, and L. E. Kay. Slow dynamics in folded and unfolded states of an sh3 domain. *J Am Chem Soc*, 123(46):11341–52, Nov. 2001.

[243] J. R. Tolman, J. M. Flanagan, M. A. Kennedy, and J. H. Prestegard. Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. *Proc Natl Acad Sci U S A*, 92(20):9279–83, Sept. 1995.

[244] A. Toumadje, S. W. Alcorn, and W. C. Johnson. Extending cd spectra of proteins to 168 nm improves the analysis for secondary structures. *Anal Biochem*, 200(2):321–31, Feb. 1992.

[245] N. T. Tran, C. D. Den Hengst, J. P. Gomez-Escribano, and M. J. Buttner. Identification and characterization of cdgb, a diguanylate cyclase involved in developmental processes in streptomyces coelicolor. *J Bacteriol*, 193(12):3100–8, June 2011.

[246] M. J. Trimble and L. L. McCarter. Bis-(3'-5')-cyclic dimeric gmp-linked quorum sensing controls swarming in vibrio parahaemolyticus. *Proc Natl Acad Sci U S A*, 108(44):18079–84, Nov. 2011.

[247] R. Tycko, F. J. Blanco, and Y. Ishii. Alignment of biopolymers in strained gels: A new way to create detectable dipole-dipole couplings in high-resolution biomolecular nmr. *J Am Chem Soc*, 122(8):9340–41, Aug. 2000.

[248] V. Uversky. Intrinsically disordered proteins from a to z. *The international journal of biochemistry & cell biology*, 2011.

[249] V. N. Uversky, C. J. Oldfield, and A. K. Dunker. Intrinsically disordered proteins in human diseases: introducing the d2 concept. *Annu Rev Biophys*, 37:215–46, 2008.

[250] N. Vajpai, M. Gentner, J.-R. Huang, M. Blackledge, and S. Grzesiek. Side-chain chi(1) conformations in urea-denatured ubiquitin and protein g from (3)j coupling constants and residual dipolar couplings. *J Am Chem Soc*, 132(9):3196–203, Mar. 2010.

[251] P. Vallurupalli, D. F. Hansen, and L. E. Kay. Structures of invisible, excited protein states by relaxation dispersion nmr spectroscopy. *Proc Natl Acad Sci U S A*, 105(33):11766–71, Aug. 2008.

[252] P. Vallurupalli, D. F. Hansen, E. Stollar, E. Meirovitch, and L. E. Kay. Measurement of bond vector orientations in invisible excited states of proteins. *Proc Natl Acad Sci U S A*, 104(47):18473–7, Nov. 2007.

[253] E. Veliath. *Synthesis and characterization of C8 analogs of c-di-GMP; New synthetic method for 5'-capped oligoribonucleotides*. PhD thesis, Rutgers, The State University of New Jersey, 2011.

[254] E. Veliath, S. Kim, B. L. Gaffney, and R. A. Jones. Synthesis and characterization of c8 analogs of c-di-gmp. *Nucleosides Nucleotides Nucleic Acids*, 30(11):961–78, Nov. 2011.

[255] A. Verma, K. Halder, R. Halder, V. K. Yadav, P. Rawal, R. K. Thakur, F. Mohd, A. Sharma, and S. Chowdhury. Genome-wide computational and expression analyses reveal g-quadruplex dna motifs as conserved cis-regulatory elements in human and related species. *J Med Chem*, 51(18):5641–9, Sept. 2008.

[256] D. Vuzman, A. Azia, and Y. Levy. Searching dna via a "monkey bar" mechanism: the significance of disordered tails. *J Mol Biol*, 396(3):674–84, Feb. 2010.

[257] D. Vuzman and Y. Levy. Dna search efficiency is modulated by charge composition and distribution in the intrinsically disordered tail. *Proc Natl Acad Sci U S A*, 107(49):21004–9, Dec. 2010.

[258] H. Wang, J.-H. Wu, J. C. Ayala, J. A. Benitez, and A. J. Silva. Interplay among cyclic diguanylate, hapr, and the general stress response regulator (rpos) in the regulation of vibrio cholerae hemagglutinin/protease. *J Bacteriol*, 193(23):6529–38, Dec. 2011.

[259] J. Wang, J. Zhou, G. P. Donaldson, S. Nakayama, L. Yan, Y. fai Lam, V. T. Lee, and H. O. Sintim. Conservative change to the phosphate moiety of cyclic diguanylic monophosphate remarkably affects its polymorphism and ability to bind dgc, pde, and pilz proteins. *J Am Chem Soc*, 133(24):9320–30, June 2011.

[260] Y. Wang and D. J. Patel. Solution structure of a parallel-stranded g-quadruplex dna. *J Mol Biol*, 234(4):1171–83, Dec. 1993.

[261] Y. Wang and D. J. Patel. Solution structure of the tetrahymena telomeric repeat d(t2g4)4 g-tetraplex. *Structure*, 2(12):1141–56, Dec. 1994.

[262] Y. Wang and D. J. Patel. Solution structure of the oxytricha telomeric repeat d[g4(t4g4)3] g-tetraplex. *J Mol Biol*, 251(1):76–94, Aug. 1995.

[263] Y. X. Wang, J. Jacob, F. Cordier, P. Wingfield, S. J. Stahl, S. Lee-Huang, D. Torchia, S. Grzesiek, and A. Bax. Measurement of 3hjnc' connectivities across hydrogen bonds in a 30 kda protein. *J Biomol NMR*, 14(2):181–4, June 1999.

[264] P. Wassmann, C. Chan, R. Paul, A. Beck, H. Heerklotz, U. Jenal, and T. Schirmer. Structure of bef3- -modified response regulator pled: implications for diguanylate cyclase activation, catalysis, and feedback inhibition. *Structure*, 15(8):915–27, Aug. 2007.

[265] N. J. West and L. J. Smith. Side-chains in native and random coil protein conformations. analysis of nmr coupling constants and [chi] 1 torsion angle preferences1. *J Mol Biol*, 280(5):867–877, 1998.

[266] D. S. Wishart, D. Arndt, M. Berjanskii, P. Tang, J. Zhou, and G. Lin. Cs23d: a web server for rapid protein structure generation using nmr chemical shifts and sequence data. *Nucleic Acids Res*, 36(Web Server issue):W496–502, July 2008.

[267] D. S. Wishart, C. G. Bigam, A. Holm, R. S. Hodges, and B. D. Sykes. 1h, 13c and 15n random coil nmr chemical shifts of the common amino acids. i. investigations of nearest-neighbor effects. *J Biomol NMR*, 5(1):67–81, Jan. 1995.

[268] P. G. Wolynes, J. N. Onuchic, and D. Thirumalai. Navigating the folding routes. *Science*, 267(5204):1619–20, Mar. 1995.

[269] R. Woody. Circular dichroism. *Methods in enzymology*, 246:34–71, 1995.

[270] R. W. Woody. Circular dichroism of protein-folding intermediates. *Methods Enzymol*, 380:242–85, 2004.

[271] P. E. Wright and H. J. Dyson. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol*, 293(2):321–31, Oct. 1999.

[272] P. E. Wright and H. J. Dyson. Linking folding and binding. *Curr Opin Struct Biol*, 19(1):31–8, Feb. 2009.

[273] J. R. Wyatt, P. W. Davis, and S. M. Freier. Kinetics of g-quartet-mediated tetramer formation. *Biochemistry*, 35(24):8002–8, June 1996.

[274] M. Xu, O. Beresneva, R. Rosario, and H. Roder. Microsecond folding dynamics of apomyoglobin at acidic ph. *J Phys Chem B*, (in press), Apr. 2012.

[275] H. Yan and A. Aguilar. Synthesis of 3', 5'-cyclic diguanylic acid (cdigmp) using 1-(4-chlorophenyl)-4-ethoxypiperidin-4-yl as a protecting group for 2'-hydroxy functions of ribonucleosides. *Nucleosides, Nucleotides, and Nucleic Acids*, 26(2):189–204, 2007.

[276] C.-Y. Yang, K.-H. Chin, M. L.-C. Chuah, Z.-X. Liang, A. H.-J. Wang, and S.-H. Chou. The structure and inhibition of a ggdef diguanylate cyclase complexed with (c-di-gmp)(2) at the active site. *Acta Crystallogr D Biol Crystallogr*, 67(Pt 12):997–1008, Dec. 2011.

[277] F. Zähringer, C. Massa, and T. Schirmer. Efficient enzymatic production of the bacterial second messenger c-di-gmp by the diguanylate cyclase ydeh from e. coli. *Appl Biochem Biotechnol*, 163(1):71–9, Jan. 2011.

[278] H. Zhang, M. Levine, and H. L. Ashe. Brinker is a sequence-specific transcriptional repressor in the drosophila embryo. *Genes Dev*, 15(3):261–6, Feb. 2001.

[279] Z. Zhang, B. L. Gaffney, and R. A. Jones. c-di-gmp displays a monovalent metal ion-dependent polymorphism. *J Am Chem Soc*, 126(51):16700–1, Dec. 2004.

[280] Z. Zhang, S. Kim, B. L. Gaffney, and R. A. Jones. Polymorphism of the signaling molecule c-di-gmp. *J Am Chem Soc*, 128(21):7015–24, May 2006.

[281] J. Zhao, E. Veliath, S. Kim, B. L. Gaffney, and R. A. Jones. Thiophosphate analogs of c-di-gmp: impact on polymorphism. *Nucleosides Nucleotides Nucleic Acids*, 28(5):352–78, May 2009.

[282] J. Zurdo, J. I. Guijarro, J. L. Jimenez, H. R. Saibil, and C. M. Dobson. Dependence on solution conditions of aggregation and amyloid formation by an sh3 domain. *J Mol Biol*, 311(2):325–40, Aug. 2001.

Acknowledgments

I would sincerely like to thank Prof. Dr. Stephan Grzesiek for giving me the opportunity to conduct this PhD thesis under his supervision. His guidance, support and encouragement strongly contributed to this thesis. I would also like to express my gratitude for his trust by giving me the opportunity to work on several diverse projects.

I wish to thank PD Dr. Daniel Häussinger for agreeing to be the co-referee of this thesis as well as his support, advice and help during the past four years. I would also also like to thank Prof. Dr. Dagmar Klostermeier and Prof. Dr. Torsten Schwede as my PhD committee members for their guidance, support and valuable discussions.

I sincerely thank Prof. Dr. Tilman Schirmer for giving me the opportunity to work on the very interesting c-di-GMP project as well as his support and help at all stages of the project.

A special thanks goes to Dr. Navratna Vajpai who helped me particularly in the first year as a PhD student. I am also thankful for the opportunity to work with Navratna and Dr. Jie-rong Huang on the side-chain rotamer project.

I want to thank Dr. Martin G. Allan and Franziska Zähringer for the collaboration on the c-di-GMP project. A special thanks goes to Franziska and Dietrich Samoray for the preparation and purification of c-di-GMP.

Many thanks go to Tobias Schmidt for his valuable contributions (MD simulations) to the Brinker project. I would like to thank Dr. Hans-Jürgen Sass for help with the CS-Rosetta calculations. I also want to thank Marco Rogowski and Klara Rathgeb-Szabo for help with protein purification. A special thanks goes to Dr. Florence Cordier for the nice and always fast correspondence. I wish to thank my Blockkurs students (Michaela, Melissa, Matthias and Christof) for their effort and motivation during their labweek on the Brinker project. I really enjoyed to work with this team of highly motivated students.

Many thanks go to Chaim Howald for the collaboration on the gel-device project. I am also very grateful for the help from Raymond Strittmatter, Karl Vogt and Röbi Wyss, who produced and modified various gel-devices.

I would like to thank Simon Ittig for the short, but successful collaboration on bacterial metabolites.

I am very grateful to Dr. Björn Burmann, Tobias Schmidt and Dr. Hans-Jürgen Sass for proof-reading parts of this thesis. I want to thank Dr. Lydia Nisius, Dr. Björn Burmann, PD Dr. Brian Cutting and Dr. Hans-Jürgen Sass for the numerous helpful discussions and their support.

I thank all present and past group members with whom I had the pleasure to work with in the last four years for their company. This includes also the present and past members of the Hiller lab as well as the Pervushin group.

Last, but not least I would like to thank all my friends, my family, particularly my parents, my sister and Susanne as well as her family for their continuous support, motivation and encouragement.