J Biomol NMR (2007) 39:1–16
DOI 10.1007/s10858-007-9169-3

ARTICLE

Theoretical framework for NMR residual dipolar couplings in unfolded proteins

O. I. Obolensky · Kai Schlepckow · Harald Schwalbe · A. V. Solov'yov

Received: 8 April 2007 / Accepted: 28 April 2007 / Published online: 7 July 2007

Abstract A theoretical framework for the prediction of nuclear magnetic resonance (NMR) residual dipolar couplings (RDCs) in unfolded proteins under weakly aligning conditions is presented. The unfolded polypeptide chain is modeled as a random flight chain while the alignment medium is represented by a set of regularly arranged obstacles. For the case of bicelles oriented perpendicular to the magnetic field, a closed-form analytical result is derived. With the obtained analytical expression the RDCs are readily accessible for any locus along the chain, for chains of differing length, and for varying bicelle concentrations. The two general features predicted by the model are (i) RDCs in the center segments of a polypeptide chain are larger than RDCs in the end segments, resulting in a bell-shaped sequential distribution of RDCs, and (ii) couplings are larger for shorter chains than for longer chains at a given bicelle concentration. Experimental data available from the literature confirm the first prediction of the model, providing a tool for recognizing fully unfolded polypeptide chains. With less certainty experimental data appear to support the second prediction as well. However, more systematic experimental studies are needed in order to validate or disprove the predictions of the model. The presented framework is an important step towards a solid theoretical foundation for the analysis of experimentally measured RDCs in unfolded proteins in the case of alignment media such as polyacrylamide gels and neutral bicelle systems which align biomacromolecules by a steric mechanism. Various improvements and generalizations are possible within the suggested approach.

Keywords Liquid crystal media · NMR residular dipolar couplings · Random flight model · Unfolded polypeptide chains

Introduction

It is now widely recognized that a thorough understanding of the structure and dynamics of unfolded proteins is of great importance for understanding protein folding, namely how a highly dynamic polypeptide chain devoid of any significant structure attains its biologically active three-dimensional conformation (Vendruscolo and Dobson 2005). Unfolded proteins are, however, not only important from the viewpoint of protein folding. It became evident only recently that intrinsically disordered proteins comprise a large part of the proteins being encoded in eukaryotic genomes (Fink 2005). Despite being (partially) unfolded they are able to fulfil biological functions, and as such they are involved in important regulatory processes in the cell such as cell cycle control and transcriptional and

O. I. Obolensky and Kai Schlepckow contributed equally to this work. O. I. Obolensky—On leave from: A.F. Ioffe Institute, St. Petersburg 194021, Russia. A. V. Solov'yov—On leave from: A.F. Ioffe Institute, St. Petersburg 194021, Russia.

O. I. Obolensky (✉) · A. V. Solov'yov
Frankfurt Institute for Advanced Studies, Max-von-Laue-Str. 1, 60438 Frankfurt am Main, Germany
e-mail: obolensky@fias.uni-frankfurt.de

K. Schlepckow · H. Schwalbe
Institute for Organic Chemistry and Chemical Biology, Center for Biomolecular Magnetic Resonance, Johann Wolfgang Goethe University, 60438 Frankfurt am Main, Germany

H. Schwalbe
e-mail: schwalbe@nmr.uni-frankfurt.de

K. Schlepckow
Frankfurt International Graduate School for Science, Johann Wolfgang Goethe University, 60438 Frankfurt am Main, Germany

translational regulation (Dyson and Wright 2005). Moreover, (partially) unfolded forms of proteins seem to play an important role in the onset of neurodegenerative diseases as it is the case in Alzheimer's (Mandelkow and Mandelkow 1998) or Parkinson's disease (Uversky et al. 2001).

High-resolution, liquid-state nuclear magnetic resonance (NMR) spectroscopy has proven to be invaluable in the investigation of the structure and dynamics of unfolded proteins. Several experimental observables can be used to gain detailed information about the heterogeneous ensemble of conformations at atomic level (Wirmer et al. 2005). Chemical shifts deviating from random coil values may indicate local conformational propensities (Bundi and Wuthrich 1979; Schwarzinger et al. 2001) whereas ^{15}N transverse magnetization relaxation measurements yield information about long-range interactions (Klein-Seetharaman et al. 2002) and conformational exchange within the ensemble (Tollinger et al. 2001). Elements of residual structure can be inferred from nuclear Overhauser effect (NOE, Neuhaus and Williamson 2000) data (Mok et al. 1999). Other approaches such as the use of spin labels and 2D photo-CIDNP (chemically induced dynamic nuclear polarization) enable the determination of residual long-range order (Gillespie and Shortle 1997; Gillespie and Shortle 1997; Bertoncini et al. 2005; Kristjansdottir et al. 2005) and of differential accessibility of aromatic side-chains involved in hydrophobic clustering (Schlorb et al. 2006), respectively.

Another especially valuable NMR observable is the direct dipole–dipole interaction between nuclear spins (the so-called residual dipolar coupling, RDC), e.g., between the spins of a ^{15}N and a ^{1}H nuclei. RDCs directly report on the average orientation of internuclear vectors with respect to the magnetic field. Net alignment is achieved by dissolving the protein in a liquid crystal medium which creates an anisotropic environment. As RDCs are correlated to each other via the alignment tensor, they provide unique long-range structural information. Therefore, they can be used in structure validation and refinement processes (Blackledge 2005).

Recently, RDCs have been measured on a variety of unfolded proteins and small peptides (Bertoncini et al. 2005; Shortle and Ackerman 2001; Ackerman and Shortle 2002a,b; Ohnishi and Shortle 2003; Alexandrescu and Kammerer 2003; Ding et al. 2004; Ohnishi et al. 2004; Fieber et al. 2004; Mohana-Borges et al. 2004; Fredriksson et al. 2004; Meier et al. 2004; Sallum et al. 2005; Bertoncini et al. 2005; Gebel et al. 2006; Binolfi et al. 2006; Sibille et al. 2006; Dames et al. 2006; Ohnishi et al. 2006; Meier et al. 2007). The interpretation of the RDC data is non-trivial due to the highly heterogeneous character of the unfolded ensemble. The earliest reports provided evidence for a native-like topology within the unfolded ensemble (Shortle and Ackerman 2001; Ackerman and Shortle 2002a,b). In a more recent study, however, it was argued in favor of simple local conformational propensities (Mohana-Borges et al. 2004). Due to this controversy in interpreting RDCs collected on unfolded proteins it is obvious that a theoretical treatment is needed in order to develop a solid foundation for data analysis.

A few theoretical approaches have been proposed for folded proteins. Reasonable predictions were obtained within the models including only steric effects (Zweckstetter and Bax 2000; Fernandes et al. 2001; Azurmendi and Bush 2002; Almond and Axelsen 2002). Upon inclusion of electrostatic contributions it was also possible to predict RDCs of folded proteins being dissolved in charged alignment media (Ferrarini 2003; Zweckstetter et al. 2004; Zweckstetter 2006).

However, these models fail in the case of unfolded proteins since not only one predominant conformation—as for a folded protein—but an ensemble of conformations, which might differ significantly from each other, has to be considered. Recently, two different groups reported similar approaches to provide a basis for the prediction of RDCs of unfolded proteins (Jha et al. 2005; Bernado et al. 2005). These studies employed the earlier introduced concept of a statistical coil (Smith et al. 1996; Fiebig et al. 1996; Schwalbe et al. 1997) in which ensembles of unfolded conformations are constructed from amino acid-specific distributions of Ramachandran angles ϕ/ψ taken from the loop regions of high-resolution X-ray structures in the protein data base. In (Bernado et al. 2005) steric overlap was avoided by residue-specific volume exclusion and reasonable agreement with experimental data was demonstrated. In (Jha et al. 2005) nearest neighbour effects were additionally taken into account, which led to a significant improvement in the agreement between calculated and experimental RDCs.

The method proposed in (Jha et al. 2005) and (Bernado et al. 2005) relies on the numerical sampling of the conformational space of the unfolded polypeptide. An attractive alternative to such a computational approach is the derivation of an analytic expression which can be used (i) to reveal and analyze the general trends of the behavior of the system, and (ii) to obtain quantitative predictions not having to perform the numerical sampling for each particular polypeptide. Unfortunately, this is not always possible.

In the present work, we show that a closed-form analytic expression for the RDCs in an unfolded polypeptide can be derived within the random flight chain model with the alignment medium represented by a set of planar obstacles.

A random flight chain is a rather crude approximation for an unfolded polypeptide. However, this idealized model captures many aspects of the behavior of unfolded

polypeptides. The random flight chain model allows an analytic description of many characteristics of unfolded polypeptides such as the radius of gyration or the mean end-to-end distance (Flory 1953).

The first attempt to obtain RDCs within the random flight chain model has been done by Annila and co-workers (Louhivuori et al. 2003) who were able to derive a representation of the RDCs as a threefold integral which had to be calculated numerically. This approach was later on extended to the valence chain model to take into account steric hindrance between chain residues (Louhivuori et al. 2004). Unfortunately, the formalism presented in the initial study (Louhivuori et al. 2003) contains three shortcomings. The most serious one is, essentially, adding the (weighed) probabilities of statistically independent events rather than multiplying them (see Eq. 1 in Louhivuori et al. 2003). Second, only one obstacle has been accounted for, while, according to the model, the polypeptide chain is confined between two obstacles. And finally, the one-dimensional random walk formalism has been employed for describing the probabilities of the possible chain conformations while, as we discuss below, the three-dimensional formalism must be used. The last shortcoming can be easily eliminated by a suitable re-scaling of the length of the chain segments. For the ranges of experimental parameters for which the polypeptide does not "feel" two obstacles simultaneously (this is a rather mild approximation, see below), the second shortcoming can also be eliminated by re-scaling. In this case the distance between the obstacles is actually twice as large as assumed in the original paper. The first shortcoming, however, cannot be remedied within the formalism presented in Louhivuori et al. (2003).

In this paper we present a general mathematical framework suitable for describing various characteristics of random flight chains, including RDCs in various alignment media. We use this framework to derive values of RDCs within the model suggested by Annila et al. We show that the RDCs can be represented in a form of an expansion over a small parameter $1/\sqrt{N}$ (where N is the number of segments in the chain). Only few first terms of this expansion are necessary for an accurate approximation of the exact formula. The series representation makes it possible to obtain analytical dependencies of the RDCs on obstacle concentration, chain length, and on position of a segment within the chain.

Two general features of RDCs in unfolded polypeptide chains are predicted by the random flight chain model. The first one is that RDCs are larger for segments in the middle of the chains as compared to the end segments. The second general feature predicted by the random flight chain model is that shorter chains exhibit larger RDCs at a given obstacle concentration.

We performed an extensive search of the available literature in order to find experimental data supporting or disproving the predictions of the model. We collected enough evidence to argue that the first feature is an important characteristic of RDCs in fully unfolded polypeptides which can serve for recognizing the fully unfolded states in further experiments. The evidence supporting the second predicted feature is less solid, as there are only two experimental studies which provide possibilities for comparing the theoretical and experimental data. The data from both studies appear to agree with the theoretically predicted behavior, but more systematic investigations are required.

Theory of residual dipolar couplings in random flight chains

General approach

Let us start with formulating the random flight chain model. A random flight chain is constructed as a set of line segments connecting subsequent steps of a three-dimensional random walk. The unfolded polypeptide is, therefore, represented in the model by a sequence of infinitely thin rods of equal, fixed length attached one to another at the tips. Each rod/segment represents a structural subunit of the polypeptide. The segments are randomly oriented, there is no interaction, including steric hindrance, with the other segments.

The dipolar coupling between nuclei P and Q depends on the angle β between the internuclear vector and the magnetic field (Ernst et al. 1987):

$$D_{\mathrm{PQ}} = \frac{\mu_0 \hbar \gamma_{\mathrm{P}} \gamma_{\mathrm{Q}}}{4\pi^2 R_{\mathrm{PQ}}^3} \left\langle \frac{3\cos^2\beta - 1}{2} \right\rangle. \quad (1)$$

Here, γ_{P} and γ_{Q} are the gyromagnetic ratios of the nuclei, R_{PQ} is the internuclear distance.

In the case of axially symmetric segments the dipolar couplings can be expressed via the angle θ between the axis of the segment to which the nuclei belong and the magnetic field (see Fig. 1):

$$D_{\mathrm{PQ}} = \frac{\mu_0 \hbar \gamma_{\mathrm{P}} \gamma_{\mathrm{Q}}}{4\pi^2 R_{\mathrm{PQ}}^3} \frac{3\cos^2\alpha_{\mathrm{PQ}} - 1}{2} \left\langle \frac{3\cos^2\theta - 1}{2} \right\rangle, \quad (2)$$

where α_{PQ} is the angle between the internuclear vector and the axis of the segment.

Since the angle $\theta = \theta(t)$ changes with time due to fluctuations of the chain, one has to average the coupling over the time of the measurements; that is denoted by the angular brackets:

$$\langle \cos^2\theta \rangle = \frac{1}{\tau} \int_0^\tau \cos^2\theta \, \mathrm{d}t. \quad (3)$$

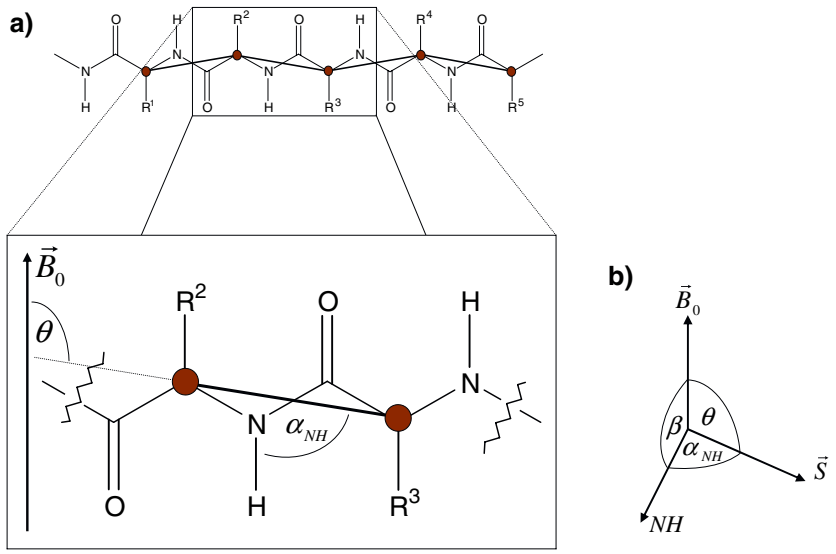

Fig. 1 (**a**) Representation of an unfolded polypeptide chain by a random flight chain. The segments of the random flight chain are being built by connecting subsequent C^{α} atoms. Additional information about the molecular structure of the polypeptide chain is being neglected. H^{α} atoms are not shown. Zoom in: Given segment is at angle θ with respect to the magnetic field. The vector connecting nuclei N and H belonging to the segment is at angle α_{NH} (corresponding to α_{PQ} in Eq. 2) with respect to the unit vector along the segment axis. The R^n indicate residue-specific side-chains. (**b**) The angles between the vector of the magnetic field B_0, the unit vector S along the segment of interest, and the internuclear vector connecting nuclei N and H, respectively

For residual dipolar couplings, the experimental parameter is averaged over a much longer time than the characteristic times of thermal motion of segments of the polypeptide chain. If no special aligning conditions are imposed (alignment due to the presence of the magnetic field is negligible in the case of diamagnetic proteins), the chain will freely fluctuate in solution and the segment S_i will spend equal times at each possible value of θ and averaging over a long time τ will result in zero coupling, $\langle\cos^2\theta\rangle = 1/3$, $D_{PQ} = 0$. If, however, segments of the chain have a preferred orientation in space, the dipolar couplings become an observable quantity and can serve as an important source of information about the structural properties of the polypeptide. The optimal experimental conditions correspond to a very small degree of alignment, $\langle\cos^2\theta\rangle - 1/3 \sim 10^{-3}$ to 10^{-4}, hence the name *residual* dipolar couplings. The random flight chain model is suitable for the NMR experiments in which alignment is caused by a spatial obstruction of chain motions due to various kinds of oriented media (bicelles in an external magnetic field, strained gels, etc.).

The statistical mechanics postulates (Landau and Lifschitz 1958) that averaging over infinite time is equivalent to averaging over the phase space of the system with a probability distribution function describing how frequently the system visits each point of its phase space. Since τ is much longer than the characteristic times of fluctuations and, additionally, since large ensembles of the polypeptide chains contribute to the measured signal, one can replace the averaging over time in Eq. 2 by averaging over the configurational space of the polypeptide.

In this section we show how the results of the theory of random walks allow one to construct a probability distribution function which depends on the angle θ only. To obtain such a function $P(\theta)$ one has to integrate over all other variables characterizing the configurational space of a polypeptide. The distributions $P(\theta)$ are different for each chain segment i, and they also depend on the length of the chain N, $P(\theta) = P_i^{(N)}(\theta)$; we will omit these indices as their values are always clear from the context. In the next section we will apply this general recipe of finding $P(\theta)$ in calculating the RDCs for a particular experimental setup.

The desired distribution $P(\theta)$ gives the probability $d\mathcal{P}$ of finding a given segment S_i of the chain oriented at an angle in an infinitesimally small vicinity of θ, $d\mathcal{P} = P(\theta)d\theta$. Then

$$\langle\cos^2\theta\rangle = \int_0^{\pi} P(\theta)\cos^2\theta\sin\theta\,d\theta = \int_{-1}^{1} P(\theta)\cos^2\theta\,d(\cos\theta). \tag{4}$$

The sine of θ in Eq. 4 appears as a remnant of the integration over the complete phase space of the chain in which the integration over the polar angle $d\Omega = \sin\theta\,d\theta\,d\phi$ is done. The probability distribution must be properly normalized,

$$\int_{-1}^{1} P(\theta)\,d(\cos\theta) = 1. \tag{5}$$

If no alignment is imposed, all values of θ are equally probable, $P(\theta) = 1/2, \langle \cos^2 \theta \rangle = 1/3$, and the RDC is zero.

The function $P(\theta)$ can generally be thought of as the ratio of the number of polypeptide conformations in which the given segment S_i is at angle θ with respect to the magnetic field to the total number of polypeptide conformations (cf. generic definition of probability, the number of successful events divided by the total number of events). In continuous space it is impossible to explicitly enumerate all the conformations of the polypeptide chain. Therefore, one has to operate with continuous distributions characterizing the density of conformations in a space of continuous variables.

In order to find the density of conformations $P(\theta)$ we use the results of the theory of random walks which connect the starting point of a walk and the number of random steps with a probability to arrive at some ending point. The segment S_i, for which $P(\theta)$ is sought, divides the chain into two parts (before and after S_i) of lengths $N_1 = i - 1$ and $N_2 = N - i$. Each subchain corresponds to a random walk characterized by the starting point (defined by the position and orientation of the segment S_i), number of steps (N_1 and N_2, respectively), and by the ending point (ending points of the chain). Therefore, the function $P(\theta)$ can be expressed via the random walk distribution functions integrated over the positions of the end points of the chain and over the position and azimuthal orientation of the given segment S_i. In this probability distribution all other variables of the configurational space of the polypeptide chain (i.e., the coordinates of the rest of the segments) have already been integrated over.

We introduce a function $W(\mathbf{r}_0, \mathbf{r}_1, \mathbf{r}_2, \mathbf{s}_i)$ proportional to the density of polypeptide conformations such that the given segment S_i is oriented along the unit vector $\mathbf{s}_i = \{\cos\phi \sin\theta, \sin\phi \sin\theta, \cos\theta\}$ (here and throughout the paper the magnetic field is taken to be along the z-axis, θ and ϕ are the polar angles), while the ends of the polypeptide chain are located at the points $\mathbf{r}_1$ and $\mathbf{r}_2$. Vector $\mathbf{r}_0$ points to the middle of S_i (see Fig. 2). Then the expression for $P(\theta)$ can be written as

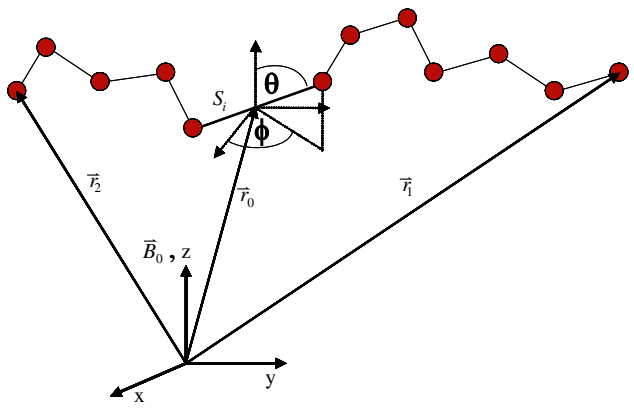

Fig. 2 Definition of the angle ϕ and vectors $\mathbf{r}_0, \mathbf{r}_1, \mathbf{r}_2$. Segment S_i is enlarged for illustrative purposes

$$P(\theta) = \frac{\mathcal{N}(\theta)}{\int_{-1}^{1} \mathcal{N}(\theta)\, \mathrm{d}(\cos\theta)}, \tag{6}$$

where

$$\mathcal{N}(\theta) = \int \mathrm{d}\mathbf{r}_0 \int_0^{2\pi} \mathrm{d}\phi \int \mathrm{d}\mathbf{r}_1 \int \mathrm{d}\mathbf{r}_2 W(\mathbf{r}_0, \mathbf{r}_1, \mathbf{r}_2, \mathbf{s}_i). \tag{7}$$

Roughly speaking, in Eq. 7 we find the number of conformations of the chain for a given position $\mathbf{r}_0$ and orientation $\mathbf{s}_i$ of the segment S_i by integrating over the number of conformations with positions of the end points $\mathbf{r}_1$ and $\mathbf{r}_2$. Then, we integrate over the position ($\mathbf{r}_0$) and orientation (ϕ) of S_i. Integration over $\mathbf{r}_0$, $\mathbf{r}_1$, and $\mathbf{r}_2$ is carried out over the whole accessible space. This results in the total number of conformations in which S_i is at angle θ with respect to the z-axis. The total number of conformations is obtained by integrating $\mathcal{N}(\theta)$ over the interval $(0,\pi)$.

The densities of conformations for the subchains before and after the segment S_i are given by a probability distribution of a random walk with the corresponding length, starting and ending points. Let us denote this distribution by $G(\mathbf{a}, \mathbf{r})$. For the first subchain the ending point of the random walk is $\mathbf{r}_1$, the starting point is one of the tips of segment $S_i, \mathbf{a}_1 = \mathbf{r}_0 + \mathbf{s}_i/2$. The segment length is taken to be 1, defining therefore the scale for all length-related quantities. For the second subchain the random walk ends in $\mathbf{r}_2$ and it starts from the other tip of the segment $S_i, \mathbf{a}_2 = \mathbf{r}_0 - \mathbf{s}_i/2$. The subchains are independent, hence the probability distributions are multiplied, that is, the number of conformations of the whole chain is a product of the numbers of conformations of the subchains:

$$W(\mathbf{r}_0, \mathbf{r}_1, \mathbf{r}_2, \mathbf{s}_i) = G(\mathbf{r}_0 + \mathbf{s}_i/2, \mathbf{r}_1) G(\mathbf{r}_0 - \mathbf{s}_i/2, \mathbf{r}_2). \tag{8}$$

The random walk probability distribution $G(\mathbf{a}, \mathbf{r})$ which is to be substituted into Eq. 8 should be determined based on the particular geometry of obstructing media, but it can also accommodate further improvements in the model (self-avoiding random walk, non-Markovian processes, etc.).

In the simplest case—no alignment, no interactions between the segments of the chain—$G(\mathbf{a}, \mathbf{r})$ was first obtained by Lord Rayleigh (Lord Rayleigh 1964). It reads as (Chandrasekhar 1943):

$$G(\mathbf{a}, \mathbf{r}) = \frac{1}{(2\pi n/3)^{3/2}} \exp\left(-\frac{3(\mathbf{r} - \mathbf{a})^2}{2n} \right). \tag{9}$$

Here n is the number of the steps in the random walk. Substitution of Eq. 9 into Eqs. 8, 7 and 6 gives a constant (1/2), which does not depend on θ, as it is expected.

We note that, in principle, the binominal distributions should be used in Eq. 9 instead of the Gaussian functions. In the limit of the infinitely long chain, $n \rightarrow \infty$, the binominal distributions coincide with the Gaussians. Therefore, strictly speaking, the analysis based on the use of the Gaussian functions is applicable for long enough subchains, i.e. for segments which are far from the ends of a long chain. However, already for $n > 2$ the Gaussians are a rather good approximation.

Alignment media in the random flight chain model are represented by a set of barriers obstructing the random walk, and the function $G(\mathbf{a}, \mathbf{r})$ has to be modified correspondingly. The role of the obstructing media is to exclude from the configurational space all the conformations of the chain which cross any of the barriers. This is achieved by modeling the obstructing media as a set of absorbing barriers. Indeed, the probability distribution of a random walk in the presence of absorbing barriers disregards the pathways connecting the starting and the ending points, which cross any of the barriers. Representing the obstructing media as a set of reflecting barriers would contradict the physical picture behind the random flight chain model, since in this case some random trajectories (=conformations) are counted twice (see, e.g. Chandrasekhar 1943).

It is instructive to make the following general remark. The integral over the probability distribution $G(\mathbf{a}, \mathbf{r})$ in the presence of obstructing media is less than 1,

$$\mathcal{I}(\mathbf{a}) = \int G(\mathbf{a}, \mathbf{r}) \mathrm{d}\mathbf{r} < 1. \qquad (10)$$

In the language of the random walk, the difference $1 - \mathcal{I}$ reflects the probability of absorption. In the chain conformations language, the integral $\mathcal{I}$ corresponds to the fraction of possible polypeptide conformations in the obstructed space as compared to the number of conformations of a free chain.

After integrating over $\mathbf{r}_1$ and $\mathbf{r}_2$ in Eq. 7 one arrives at

$$\mathcal{N}(\theta) = \int \mathrm{d}\mathbf{r}_0 \int_0^{2\pi} \mathrm{d}\phi \mathcal{I}(\mathbf{r}_0 + \mathbf{s}_i/2)\mathcal{I}(\mathbf{r}_0 - \mathbf{s}_i/2). \qquad (11)$$

Residual dipolar couplings in the presence of bicelles

Let us consider a particular type of experiment for the measurement of RDCs in which alignment of the polypeptide chains is achieved by placing bicelles into the solution. Bicelles are large, disk-shaped bodies which weakly interact with polypeptides and orient in the presence of an externally applied magnetic field (Sanders and Schwonek 1992). We assume that the bicelles are aligned perpendicular to the magnetic field. Neglecting the finite size of the bicelles we approximate them by a set of parallel infinite planes.

Typical interplanar distances encoutered in experiments are 400–600 Å in the case of phospholipid bicelles (Bax and Tjandra 1997) and a few nanometers in the case of *n*-alkyl-poly(ethylene glycol)/*n*-alkyl alcohol bicelles (Ruckert and Otting 2000), respectively. We have taken the length of one amino acid residue (C^{α}–C^{α} distance $\approx$ 3.8 Å) to be the unit of length in our model, hence, an L value of 150 corresponding to an interplanar distance of about 570 Å seems to be a reasonable choice for the calculations.

We note that the model—random walk between two infinite planes oriented perpendicular to the z-axis while there are no obstructions in the x- and y-directions—implies the one-dimensional character of the random walk problem. However, the random walk probability distribution has to be taken from the 3D random walk formalism rather than from the one-dimensional one. The reason is that in the 3D case the characteristic width of the random walk probability distribution (which translates into the radius of gyration, end-to-end distance, etc.) is a factor of $\sqrt{3}$ smaller than in the 1D case. When applying the 1D formalism in this model, one effectively replaces the steps of *equal length* with the steps of *equal projection* onto the z-axis.

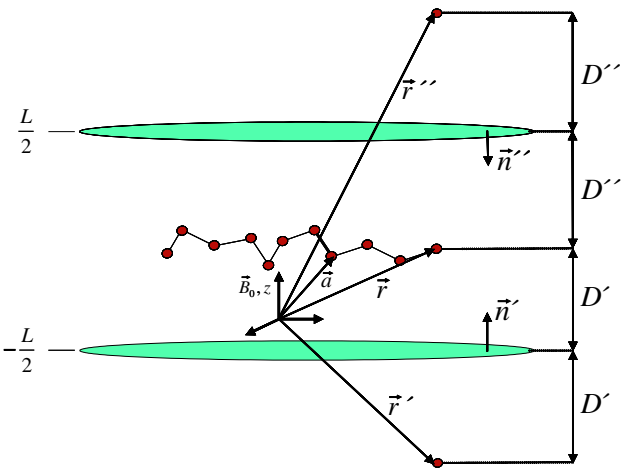

Fig. 3 Definition of the vectors $\mathbf{r}'$ and $\mathbf{r}''$. The two ellipsoids represent two bicelle particles oriented perpendicular with respect to B_0. The bicelles are approximated by infinite planes in the calculations. The segment which divides the chain in two sub-chains is thickened for illustrative purposes

The probability distribution of the 3D random walk between two parallel infinite absorbing barriers is easily obtained by generalization (based on symmetry considerations) of a one-dimensional walk in the presence of one barrier (Chandrasekhar 1943),

$$G(\mathbf{a},\mathbf{r}) = \frac{1}{(2\pi n/3)^{3/2}}\left[\exp\left(-\frac{3(\mathbf{r}-\mathbf{a})^2}{2n}\right) - \exp\left(-\frac{3(\mathbf{r}'-\mathbf{a})^2}{2n}\right) - \exp\left(-\frac{3(\mathbf{r}''-\mathbf{a})^2}{2n}\right)\right] \quad (12)$$

Here $\mathbf{a}$ is the starting point of an n-steps walk, $\mathbf{r}$ is the ending point, $\mathbf{r}'$ is the point symmetric to $\mathbf{r}$ with respect to mirror reflection in the plane of the first barrier, $\mathbf{r}' = \mathbf{r} - 2D'\mathbf{n}'$ (where D' is the distance between $\mathbf{r}$ and the first barrier, $\mathbf{n}'$ is the unit normal to the plane of the first barrier). Analogously, $\mathbf{r}'' = \mathbf{r} - 2D''\mathbf{n}''$ is the image of the point $\mathbf{r}$ in the plane of the second barrier (Fig. 3).

Equation 12 is valid for $n < L$, when the chain cannot "feel" two barriers simultaneously. However, completely or almost completely extended conformations for which the size of the chain approximately equals N are extremely improbable. Rather, the typical size of the chain is of the order of $\sqrt{N}$. Therefore, in fact Eq. 12 works well until $N \sim L^2$, i.e. it is satisfied for most of the experimental conditions.

The second and the third terms in the square brackets in Eq. 12 correspond to exclusion of the trajectories/conformations which cross the first and the second barriers, respectively. These terms are responsible for the integrals (10) being less than 1.

Let us now calculate this integral. For convenience, we will assume that the barriers are perpendicular to the z-axis, crossing it at the points $-L/2$ and $L/2$ (see Fig. 3). Then, integrations over x and y result in the factor $(2\pi n)/3$. The integration over z from $-L/2$ to $L/2$ results in a set of error functions Erf(x):

$$\mathcal{I}(\mathbf{a}) = \int G(\mathbf{a},\mathbf{r})\mathrm{d}\,\mathbf{r} = \mathrm{Erf}\left[\sqrt{\frac{3}{2n}}\left(\frac{L}{2}-a_z\right)\right] + \mathrm{Erf}\left[\sqrt{\frac{3}{2n}}\left(\frac{L}{2}+a_z\right)\right] - 1. \quad (13)$$

Substituting Eq. 13 into Eq. 11 one obtains

$$\mathcal{N}(\theta) = \int \mathrm{d}\mathbf{r}_0 \int_0^{2\pi} \mathrm{d}\phi \left\{\mathrm{Erf}\left[\sqrt{\frac{3}{2N_1}}\left(\frac{L}{2}-z_0-c/2\right)\right] + \mathrm{Erf}\left[\sqrt{\frac{3}{2N_1}}\left(\frac{L}{2}+z_0+c/2\right)\right] - 1\right\} \times \left\{\mathrm{Erf}\left[\sqrt{\frac{3}{2N_2}}\left(\frac{L}{2}-z_0+c/2\right)\right] + \mathrm{Erf}\left[\sqrt{\frac{3}{2N_2}}\left(\frac{L}{2}+z_0-c/2\right)\right] - 1\right\}, \quad (14)$$

where $N_1 = i - 1$ and $N_2 = N - i$ are the lengths of the parts of the chain before and after the given segment S_i, $(N_1 + N_2 + 1 = N)$, z_0 is the z-component of the vector $\mathbf{r}_0$, and

$$c \equiv \cos\theta \quad (15)$$

is the z-component of the vector $\mathbf{s}_i$.

In Eq. 14 due to the symmetry of the problem there are no dependences on x_0, y_0, and the azimuthal angle ϕ. The integrals over these variables produce the infinitely large factor $2\pi\sigma$, where σ is the (infinite) surface of the barriers. However, exactly the same factor is produced in the denominator of Eq. 17, so the value of $P(\theta)$ remains finite. Therefore, the probability distribution is expressed via the one-dimensional integral over the position of the middle of the given segment S_i, z_0:

$$P(c) = \frac{\mathcal{E}(c)}{\int_{-1}^{1}\mathcal{E}(c)\mathrm{d}(c)}, \quad (16)$$

where

$$\mathcal{E}(c) = \int_{-L/2+|c|/2}^{L/2-|c|/2}\left\{\mathrm{Erf}\left[\sqrt{\frac{3}{2N_1}}\left(\frac{L}{2}-z_0-c/2\right)\right] + \mathrm{Erf}\left[\sqrt{\frac{3}{2N_1}}\left(\frac{L}{2}+z_0+c/2\right)\right] - 1\right\} \times \left\{\mathrm{Erf}\left[\sqrt{\frac{3}{2N_2}}\left(\frac{L}{2}-z_0+c/2\right)\right] + \mathrm{Erf}\left[\sqrt{\frac{3}{2N_2}}\left(\frac{L}{2}+z_0-c/2\right)\right] - 1\right\}\mathrm{d}z_0. \quad (17)$$

At this stage of derivation, the reason for the non-zero RDCs can be clearly seen. Indeed, $\mathcal{E}(c)$ is just a renormalized number of chain conformations for which the segment of interest at a particular angle θ ($c \equiv \cos\theta$). In

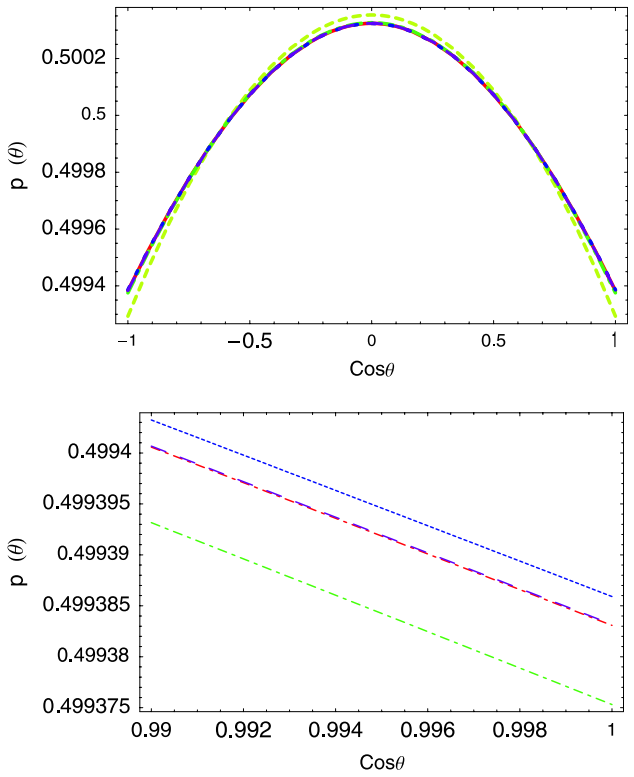

Fig. 4 Comparison of $P(\theta)$ for the center segment of a 21mer chain calculated with the series representation (19) (*dashed-dotted green* (including terms up to $|c^3|$), *dotted blue* (including terms up to c^4), and *dashed purple* (inlcuding terms up to $|c^5|$) curves) and with the exact expression (17) (*red*). The bottom graph is a zoom in of the bottom right part of the top graph

Eq. 17 there are terms $|c|/2$ in the limits of the integration which are due to the fact that the center of the given segment can not approach the barriers closer than $|c|/2$. The conformations in which the segment of interest is parallel to the barriers ($c = 0$) have larger space available to them and, therefore, they are more populated.

Expression (17) can be further simplified by expanding the error functions in a series over $|c|/2$. Using the standard Taylor series

$$\begin{aligned}\mathrm{Erf}(x+\delta) = \mathrm{Erf}(x) &+ \frac{2}{\sqrt{\pi}}\exp(-x^2)\delta - \frac{2}{\sqrt{\pi}}x\exp(-x^2)\delta^2 \\ &+ \frac{2}{\sqrt{\pi}}\frac{2x^2-1}{3}\exp(-x^2)\delta^3 \\ &- \frac{2}{\sqrt{\pi}}\frac{2x^2-3}{6}x\exp(-x^2)\delta^4 + \cdots \end{aligned} \tag{18}$$

one obtains after some algebra

$$\begin{aligned}\mathcal{E}(c) = L &- \frac{2}{\sqrt{\pi}}\sqrt{\frac{2(N_1+N_2)}{3}} - \frac{2}{\sqrt{\pi}}\sqrt{\frac{3}{2(N_1+N_2)}}c^2 \\ &+ \frac{1}{\pi}\sqrt{\frac{1}{N_1N_2}}|c|^3 + \frac{1}{2\sqrt{\pi}}\sqrt{\frac{3}{2(N_1+N_2)^3}}c^4 \\ &- \frac{3}{20\pi}(N_1+N_2)\sqrt{\frac{1}{N_1^3N_2^3}}|c|^5 + \cdots \end{aligned} \tag{19}$$

As shown in Fig. 4, one has to include terms up to $|c|^5$ in order to achieve accuracy (10^{-5}) needed for the calculations of the RDCs.

The denominator in Eq. 16 is calculated, then, as follows:

$$\begin{aligned}\int_{-1}^{1}\mathcal{E}(c)\mathrm{d}\,c = 2L &- \frac{4}{\sqrt{\pi}}\sqrt{\frac{2(N_1+N_2)}{3}} \\ &- \frac{4}{3\sqrt{\pi}}\sqrt{\frac{3}{2(N_1+N_2)}} + \frac{1}{2\pi}\sqrt{\frac{1}{N_1N_2}} \\ &+ \frac{1}{5\sqrt{\pi}}\sqrt{\frac{3}{2(N_1+N_2)^3}} \\ &- \frac{1}{20\pi}(N_1+N_2)\sqrt{\frac{1}{N_1^3N_2^3}} + \cdots \end{aligned} \tag{20}$$

Equations 16, 19, 20 define the probability distribution for a given segment to be found at a particular angle θ. With the given probability distribution, we can now find the average angular distribution in Eq. 2.

Equation 21 is the final result of our derivation.

$$D_{PQ} \approx \frac{\mu_0\hbar\gamma_P\gamma_Q}{4\pi^2R_{PQ}^3}\frac{3\cos^2\alpha_{PQ}-1}{2} \times \frac{-\frac{8}{15\sqrt{\pi}}\sqrt{\frac{3}{2(N_1+N_2)}} + \frac{1}{4\pi}\sqrt{\frac{1}{N_1N_2}} + \frac{4}{35\sqrt{\pi}}\sqrt{\frac{3}{2(N_1+N_2)^3}} - \frac{N_1+N_2}{32\pi}\sqrt{\frac{1}{N_1^3N_2^3}}}{2L - \frac{4}{\sqrt{\pi}}\sqrt{\frac{2(N_1+N_2)}{3}} - \frac{4}{3\sqrt{\pi}}\sqrt{\frac{3}{2(N_1+N_2)}} + \frac{1}{2\pi}\sqrt{\frac{1}{N_1N_2}} + \frac{1}{5\sqrt{\pi}}\sqrt{\frac{3}{2(N_1+N_2)^3}} - \frac{N_1+N_2}{20\pi}\sqrt{\frac{1}{N_1^3N_2^3}}}. \tag{21}$$

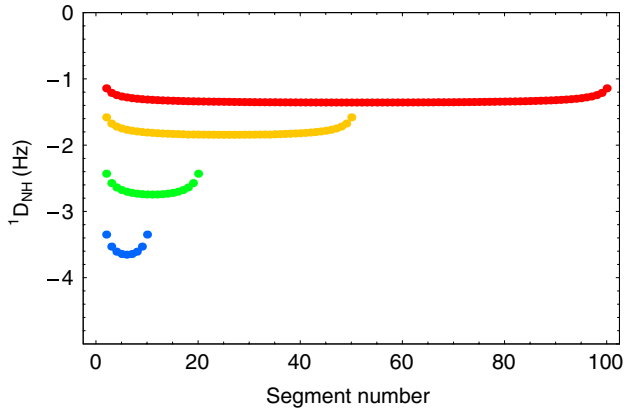

Fig. 5 NH residual dipolar couplings as a function of the segment number for random flight chains of different lengths: 11 segments (*blue*), 21 segments (*green*), 51 segments (*yellow*), and 101 segments (*red*)

Results and discussion

Predictions of the model

In order to reveal the two general patterns in the behavior of the RDCs which are predicted by the random flight chain model, let us further simplify expression (21) by expanding it over $1/L$ and $1/\sqrt{N}$ provided that $L \gg \sqrt{N} \gg 1$. This condition is fulfilled for typical ranges of experimental parameters.

We keep only the first terms in the expansion:

$$D_{PQ} \approx -\frac{\mu_0 \hbar \gamma_P \gamma_Q}{4\pi^2 R_{PQ}^3} \frac{3\cos^2 \alpha_{PQ} - 1}{2} \times \frac{1}{L}\left[\frac{4}{15\sqrt{\pi}}\sqrt{\frac{3}{2(N_1+N_2)}} - \frac{1}{8\pi}\sqrt{\frac{1}{N_1 N_2}}\right], \quad N = N_1 + N_2 + 1. \tag{22}$$

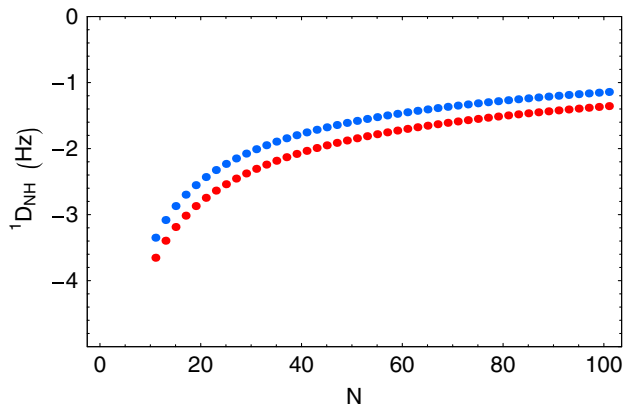

Fig. 6 NH residual dipolar couplings as a function of the chain length N for the segment next to the end segment (upper curve) and the center segment (lower curve)

From Eq. 22 it is evident that shorter chains have larger (in absolute value) RDCs at fixed distance L. From the first glance, this is a counter-intuitive result. Indeed, one would expect that the lengthier chains, which are "bigger", should be more affected by the presence of the obstacles as compared to "smaller" shorter chains. However, within the same theoretical framework, one can show that the chain clews almost never approach the obstacles closer than their characteristic radius $\sim \sqrt{N}$; the conformations in which the chain is thinly spread along the obstacle are statistically insignificant. In the vicinity of the obstacle the average "shape" of the chain resembles a sphere flattened on one side, in contrast to the unperturbed spherical shape characteristic of non-obstructed chains. The deformation is maximum at the surface of the chain clew while the inner region is less affected. From this point of view it is evident that the influence of the obstacle is greater for clews with greater surface-to-volume ratios. That is, the short chains are more affected by the obstacles and therefore have larger RDCs.

The second general feature predicted by the random flight chain model is that the RDCs are larger for the segments in the middle of the chain as compared to the end segments,[1] leading to the bell-like shape of the RDC profile, as illustrated in Fig. 5. This feature can also be explained in terms of average shapes of the chain clews. The reason why the middle segments are more affected by the obstacles is the following. As we have just discussed the chain clews resemble flattened spheres with maximum deformation on one side of the clew at the contact with the

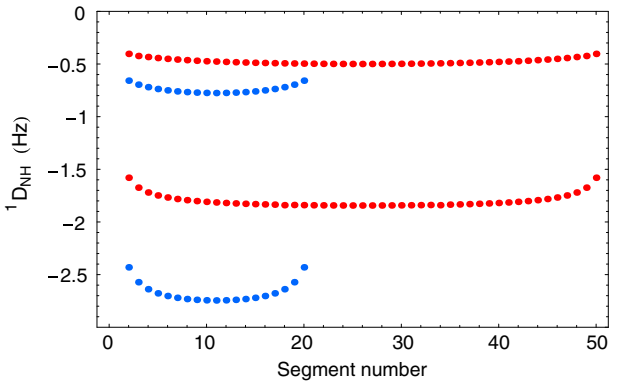

Fig. 7 Comparison of NH residual dipolar couplings as a function of the segment number for random flight chains consisting of 21 and 51 segments (*blue* and *red*, correspondingly). The upper profiles are obtained with the expression from Louhivuori et al. (2003), the lower profiles are the present results reproduced from Fig. 5

[1] It is assumed that the internuclear vectors throughout the chain are on average oriented at the same angle α_{PQ} with respect to the symmetry axis of the corresponding segment.

Table 1 Summary of experimental conditions and residual structure exhibited by unfolded polypeptides on which RDCs have been measured

Polypeptide (chain length)	Unfolding conditions	Alignment media	Residual structure	$^1D_{PQ}$	Model agreement
SNase[a] Δ131Δ (131) (Shortle and Ackerman 2001)	Truncation; truncation + urea	12% PAG (compr.)[b]; 5% C*m*E*n*[c]/alcohol	Long-range; native-like topology	$^1D_{NH}$	–[d]
SNase[a] (149) and SNase[a] Δ131Δ (131) (Ackerman and Shortle 2002)	Urea; acid; truncation; truncation + urea; truncation + detergent	6–14% PAG (compr.)[b]; 8–12% PAG (str.)[b]; 3–7% C8E5[c]/1-octanol; 6.5% CPBr[e]/1-hexanol	Long-range; native-like topology	$^1D_{NH}$	–[d]
Mutants (131) and fragments (101) of SNase[a] Δ131Δ (Ackerman and Shortle 2002)	Truncation; truncation + urea	12% PAG (compr.)[b]	Long-range; native-like topology; local propensities	$^1D_{NH}$	–[d]
Peptides (2, 4, 10, and 15) (Ohnishi and Shortle 2003)	Intrinsically disordered	18% PAG (compr.)[b]; 8–10% C8E5[c]/1-octanol	Local propensities	$^1D_{NH}$	–[f]
S-peptide of RNaseA[g] (22) (Alexandrescu and Kammerer 2003)	Decreased salt concentration; increased temperature	5.2% C8E5[c]/1-octanol	α-Helical	$^1D_{NH}$	No bell shape
GB1[h] (56) (Ding et al. 2004)	Acid + increased temperature	6% PAG (str.)[b]; fd[i] (~30 mg/mL)	Not native-like	$^1D_{NH}$	–[f]
Eglin C (70) (Ohnishi et al. 2004)	Acid + urea	12% PAG (compr.)[b]	Native-like global structure	$^1D_{NH}$, $^1D_{C^\alpha H^\alpha}$	No bell shape
ACBP[j] (86) (Fieber et al. 2004)	GdnCl; acid	7% PAG (str.)[b]	α-Helical propensity; native-like long-range interactions in acid solution; suggested loss of long-range interaction in acid solution upon I27A mutation	$^1D_{NH}$	Approximate bell shape for GdnCl; no bell shape for acid
apoMb[k] (153) (Mohana-Borges et al. 2004)	Acid + urea; acid	10% PAG (compr. + str.)[b]	α-Helical propensity for acid; neither secondary nor tertiary interactions for acid + urea	$^1D_{NH}$	Bell shape for acid + urea; no bell shape for acid
Polyglutamic acid (21) (Fredriksson et al. 2004)	Intrinsically disordered	6.5% DMPC/DHPC[l]	No res. structure reported	RDC values not given explicitly	RDCs only measured for E11 and E21
CTD[m] of T4 fibritin (29) (Meier et al. 2004)	Increased temperature	7% PAG (compr. + str.)[b]	Local propensities	$^1D_{NH}$, $^1D_{C^\alpha H^\alpha}$	No bell shape
αS[n] (140) and fragment thereof (32) (Bertoncini et al. 2005)	Intrinsically disordered; urea; increased temperature; polyamine binding	Pf1[o](10 mg/mL); 5% C8E5[c]/octanol	Long-range interactions, weakened to different extents by polyamine ligands and disrupted by elevated temperature; disruption of hydrophobic clusters by urea	$^1D_{NH}$	Approximate bell shape for full-length αS in C8E5[c]/octanol in presence of 8 M urea and upon spermine binding (in the latter case PRE[p] data still show existent long-range interaction)
SNase[a](149) and fragment thereof (102) (Sallum et al. 2005)	Acid + urea	7.5% PAG (str.)[b]	No long-range interactions	$^1D_{NH}$	Approximate bell shape in both cases
Mutants of αS[n] (140) (Bertoncini et al. 2005)	Intrinsically disordered; urea	Pf1(10 mg/mL)[o]; 5% C8E5[c]/octanol	For A18C basically as for WT[q]; A30P and A53T induce local effects and affect long-range interactions as compared to WT∝S	$^1D_{NH}$	Approximate bell shape for A30P and A53T in C8E5[c]/octanol and for A30P in C8E5[c]/octanol/8 M urea, respectively

Table 1 continued

Polypeptide (chain length)	Unfolding conditions	Alignment media	Residual structure	$^1D_{PQ}$	Model agreement
SNase[a] Δ131Δ (131) (Gebel et al. 2006)	Truncation + urea	6% PAG (str.)[b]; composite of 6% PAG (str.)[b] and Pf1[o] (3 and 10 mg/mL) at different relative angles	Tertiary interactions	$^1D_{NH}$	Approximate bell shape in 6% PAG (str.)[b]
αS (140) (Binolfi et al. 2006)	Intrinsically disordered	5% C8E5[c]/octanol	Long-range interaction	$^1D_{NH}$	No bell shape
Tau (441) (Sibille et al. 2006)	Intrinsically disordered	5% C12E5[c]/1-hexanol	Local propensities	$^1D_{NH}$	No bell shape
Peptides (9 and 12) (Dames et al. 2006)	Intrinsically disordered	10% PAG (str.)[b]; Pf1[o] (20 mg/mL)	Local propensities; pronounced effects induced by G, P, W, and Y	$^1D_{NH}$, $^1D_{C^\alpha H^\alpha}$	Approximate bell shape in PAG[b] for longer peptide; slight deviation from bell shape in PAG[b] for shorter peptides
Peptides of varying length and amino acid composition/sequence (6, 7, 8, 12, 15) (Ohnishi et al. 2006)	Intrinsically disordered	18% PAG (compr.)	Preference for extended conformation	$^1D_{NH}$, $^1D_{C^\alpha H^\alpha}$	Approximate bell shape for $^1D_{NH}$ of Ac-YGEGSGAGTGDG; other sequences either too short (6–8 res.) or RDCs only available for part of the sequence (12/15 res.)
Ubiquitin (76) (Meier et al. 2007)	Acid + urea	10% PAG	Native-like long-range structure	$^1D_{NH}$, D_{H^N,H^N}	Approximate bell shape

The most right column reports on whether the sequential distribution of RDCs agrees with the first prediction of the model

[a] SNase: Staphylococcal nuclease

[b] PAG: Polyacrylamide gel (compressed/stretched)

[c] C*m*E*n*/C8e5/C12E5/C12E5: *n*-Alkyl-poly (ethylene glycol)/*n*-Octyl-penta(ethylene glycol)/*n*-Dodecyl-penta(ethylene glycol)

[d] NH RDCs only available for part of the sequence

[e] CPBr: Cetylpyridinium bromide

[f] Residue assignment not given

[g] RNaseA: Ribonuclease A

[h] GB1:B1 domain of protein G

[i] fd: fd bacteriophage

[j] ACBP: Acyl-coenzyme A binding protein

[k] apoMb: apo-Myoglobin

[l] DMPC/DHPC: Dimyristoyl-/Dihexanoyl-phosphatidycholine

[m] CTD: C-terminal domain

[n] αS: α-Synuclein

[o] Pf1: Pf1 bacteriophage

[p] PRE: Paramagnetic relaxation enhancement

[q] Wild type

obstacle. In this situation the end segments (located mostly on the surface of the clew) are affected most, while the middle segments are affected less. However, when the clew rotates (keeping the distance to the obstacle constant) other end segments become affected by the obstacle while the middle segments are still constrained. This rotational degree of freedom leads to the fact that on average the middle segments are more strongly aligned and the RDCs in such segments are larger.

These two features are illustrated in Figs. 5 and 6. The presented data correspond to $^1D_{NH}$ couplings using a value of –21,700 Hz for the constant term in Eq. 2. The angle α_{NH} (α_{PQ}, see Eq. 2) was taken to be 90°, a value close to ≈ 85° dictated by the rigid part of the geometry of the chain. One can easily check that deviations from 90° value do not exceed 10–15° (0.26 rad) in all sensible conformations (α-helix, β-sheet, polyproline II) being populated in unfolded polypeptides. Such deviations are insignificant for the RDC values, since $\cos^2(90° \pm 15°) \approx 0.07 \ll 1$ (see Eqs. 21 and 22).

Before we proceed with comparing these two general features with experimental data which have been collected on unfolded proteins and model peptides up to date, let us compare our results with those obtained in (Louhivuori et al. 2003). In Fig. 7 we present the RDC profiles calculated with the expressions from (Louhivuori et al. 2003) (two upper curves) and with the use of the present formalism (two lower curves). It is seen from the figure, that the shortcomings mentioned in the introduction lead to the three times smaller RDCs (in absolute values), and also to a less pronounced bell-shape.

Comparison with experiment

When comparing the theoretical predictions and experimental data one has to bear in mind the following two important aspects. First, the random flight chain model provides us with a model for an unfolded protein in which long-range interactions are not taken into account. Second, the derived expressions only hold true in the case of a steric alignment mechanism since we do not treat electrostatic contributions as they occur, e.g., when using the filamentous bacteriophages Pf1 (Hansen et al. 1998) and fd (Clore et al. 1998). We therefore limit our comparison to RDCs which have been experimentally measured in bicelle and polyacrylamide alignment media (Tycko et al. 2000; Sass et al. 2000). Moreover, the derivation in the section "Residual dipolar couplings in the presence of bicelles"is carried out for obstruction in one dimension, while for polyacrylamide motion of the polypeptide chain is obstructed in all three dimensions. Therefore, certain deviations from the predicted behavior are possible for the polyacrylamide media. For the complete set of experimental conditions and details about residual structure concerning all investigated proteins and peptides, please, refer to Table 1.

$^1D_{NH}$ dependence on segment number

The $^1D_{NH}$ plotted as a function of the segment number follow a symmetric bell shape (see Fig. 5). Inspection of experimentally determined $^1D_{NH}$ reveals bell-like shapes in a few cases (Fieber et al. 2004; Mohana-Borges et al. 2004; Sallum et al. 2005; Dames et al. 2006; Ohnishi et al. 2006; Meier et al. 2007). Saying bell-like we talk of profiles in which NH RDCs are (i) on average larger in the middle of the chain as compared to the chain ends, and (ii) exhibit a uniform sign. The first report of such a profile was the paper by Poulsen and co-workers (Fieber et al. 2004). They measured $^1D_{NH}$ of the GdnCl-denatured state of acyl-coenzyme A binding protein (ACBP). Except for two amino acid residues (change in sign) an overall bell shape is observed. However, this is not the case for the

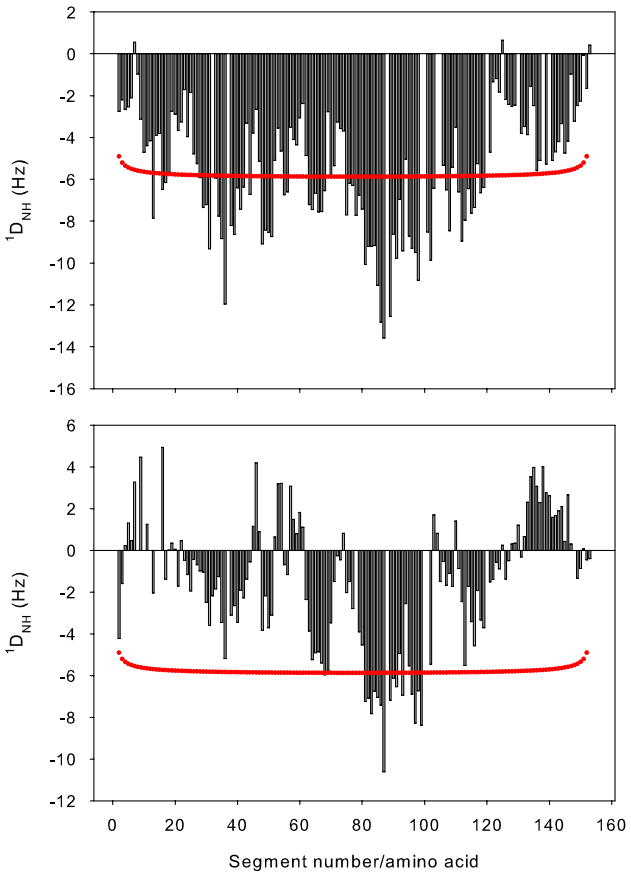

Fig. 8 Experimental $^1D_{NH}$ for the urea- and acid-denatured states of apoMb (*black bars*; top and bottom graphs, respectively) and predicted $^1D_{NH}$ for a chain of equal length (*red*). RDC profiles of apoMb were taken from Mohana-Borges et al. (Mohana-Borges et al. 2004), digitized and replotted. For the prediction an interplanar distance of $L = 38$ was taken

acid-denatured state of ACBP as the sign of the dipolar coupling changes along the sequence multiple times. Since it is known that denaturants such as GdnCl lead to more extended chain conformations whereas acid denaturation promotes more compact states of unfolded proteins, the GdnCl-denatured state of ACBP might be a good model case for a fully unfolded protein following the prediction made by the random flight chain model. A higher degree of compactness in acid-denatured ACBP, however, favors elements of residual structure and therefore leads to deviations from the bell-shaped distribution of RDCs exhibited by the random flight chain. Indeed, the occurrence of hydrophobic long-range interactions has been reported (Fieber et al. 2004).

The conclusion drawn from the RDC data collected on the GdnCl-denatured and acid-denatured states of ACBP is supported by the report by Mohana-Borges et al. (Mohana-Borges et al. 2004). Corresponding NH RDCs were measured on urea-denatured and acid-denatured apoMb. The urea-denatured state of apoMb exhibits a bell-shaped RDC profile whereas in the case of acid-denatured apoMb the dipolar coupling sign changes along the amino acid sequence several times. This is illustrated in Fig. 8 where NH RDCs measured for the two different unfolded states of apoMb and predicted NH RDCs for a chain of equal length are shown. We note that the simple model which we use in this work, of course, does not capture any amino acid specific features that are clearly observed in experiments.

Bell-shaped RDC distributions were observed in two further studies. In the first study, NH RDCs for the urea-denatured state of full-length staphylococcal nuclease (SN) reported by Sallum et al. follow the prediction made by the random flight chain model (Sallum et al. 2005). This is as well the case for the second study where NH RDCs were measured on model peptides of 9 and 12 residues in length (Dames et al. 2006). Although no denaturant was present in the experiments (Dames et al. 2006), the lengths of the peptides chosen and their amino acid composition preclude the occurrence of long-range interactions and (in most cases) local order which makes them an ideal model for a fully unfolded polypeptide chain. Therefore, not surprisingly, for almost all peptides studied in 10% PAG (stretched) an approximately bell-shaped RDC profile is observed (Dames et al. 2006). Likewise, the same reasoning applies to the 12-residue peptide Ac-YGEGSGAGTGDG for which $^1D_{NH}$ were measured in 18% PAG (compressed) in the absence of denaturant (Ohnishi et al. 2006). In a further study, RDCs were measured for the central and C-terminal residues of a 21-residue polyglutamate chain (Fredriksson et al. 2004). The axial alignments of the respective C^{α} fragments, which were derived from $C^{\alpha}H^{\alpha}$, $C^{\alpha}CO$, and $C^{\alpha}C^{\beta}$ RDCs, show a sixfold larger value for the central fragment as compared to the C-terminal fragment. While a sequential RDC profile is not available, this observation is in qualitative agreement with the prediction that the chain center shows larger alignment than the chain ends.

As mentioned above, the presence of residual intramolecular interactions within ensembles of unfolded proteins might lead to deviations from the characteristic distribution of $^1D_{NH}$ observed along a random flight chain which is likely the case for the acid-denatured states of ACBP (Fieber et al. 2004) and apoMb (Mohana-Borges et al. 2004). We therefore also analyzed $^1D_{NH}$ profiles of proteins which are reported to exhibit long-range interactions in their unfolded state ensembles. Since long-range interactions are not explicitly taken into account in the random flight chain model one would expect deviations from the predicted bell-shaped profile.

The pioneering work of Shortle and co-workers on the Δ131Δ construct of SN is an especially attractive case in this context as the persistence of a native-like topology under even highly denaturing conditions (8 M urea) was reported in three subsequent studies (Shortle and Ackerman 2001; Ackerman and Shortle 2002a,b). However, NH RDCs are available only for part of the sequence as the complete backbone resonance assignment could not be undertaken in an earlier study which was attributed to exchange broadening (Alexandrescu et al. 1994). Hence, an analysis of the profile's shape was not conducted.

The most recent RDC study on Δ131Δ is consistent with the earlier reports (Shortle and Ackerman 2001; Ackerman and Shortle 2002a,b) stating a 'specific tertiary structural arrangement, not inconsistent with a nativelike topology' (Gebel et al. 2006). Interestingly, the $^1D_{NH}$ profile in 6% PAG (stretched) deviates only slightly from the bell shape predicted by the random flight chain model suggesting that the assumed tertiary structural arrangement is not very much reflected in the RDCs measured in PAG alone.

For another protein, eglin C, a native-like global structure was also reported (Ohnishi et al. 2004). A bell-shaped profile cannot be inferred from sequential NH RDCs of the urea-denatured state supporting the author's conclusion that residual structure is being populated. The finding that native-like hydrophobic long-range interactions persist in acid-denatured ACBP is in line with our above reasoning (Fieber et al. 2004). Recent work on α-synuclein (αS) also showed that long-range contacts greatly affect the sequential $^1D_{NH}$ profile which is reflected in large couplings in specific parts of the amino acid sequence (Bertoncini et al. 2005a,b). Perturbation of these long-range contacts by the addition of urea (Bertoncini et al. 2005a) and by single point mutations (Bertoncini et al. 2005b) resulted in significant decreases in respective RDCs.

Overall, we conclude that long-range interactions in unfolded states of proteins are likely reflected in $^1D_{NH}$ profiles which should significantly deviate from the bell shape exhibited by random flight chains. Of course, it cannot be ruled out that different intramolecular interactions present within unfolded states of proteins may compensate each other such that NH RDCs still follow a symmetric bell shape. An example may be the case of the acid- and urea-denatured state of ubiquitin. In the corresponding report, native-like long-range interactions were shown to exist for this state by the measurement of long-range D_{H^N,H^N} while the $^1D_{NH}$ profile is still approximately bell-shaped (Meier et al. 2007).

We also note that there are further reports on NH residual dipolar couplings in short peptides and unfolded proteins. In some cases residue assignments were not given which did not allow us to include these data into our comparison (Ohnishi and Shortle 2003; Ding et al. 2004). In other cases local structural propensities might also lead to deviations from the predicted bell-shaped profile (Alexandrescu and Kammerer 2003; Meier et al. 2004; Sibille et al. 2006).

$^1D_{NH}$ dependence on chain length

We shall now give attention to the chain length dependence of NH RDCs. The random flight chain model predicts stronger alignment for shorter chains as compared to longer chains (Figs. 5 and 6).

As the random flight chain represents a model for a fully unfolded polypeptide chain this prediction is solely compared to experimental data from peptides and unfolded states of proteins which are reported to lack long-range contacts in their conformational ensembles (Fieber et al. 2004; Mohana-Borges et al. 2004; Sallum et al. 2005; Dames et al. 2006). Based on our model, corresponding RDCs should in principle be a direct measure of the extent of alignment. Chain lengths of the polypeptides which were considered are 9 and 12 (model peptides, Dames et al. 2006), 86 (ACBP, Fieber et al. 2004), 102 (truncation fragment of SN, Sallum et al. 2005), 149 (full-length SN, Sallum et al. 2005) and 153 residues (apoMb, Mohana-Borges et al. 2004). $^1D_{NH}$ were all measured in PAG with acrylamide concentrations ranging from 7% (Fieber et al. 2004) to 10% (Mohana-Borges et al. 2004; Dames et al. 2006). It is evident that these differences in the acrylamide concentration and further differences in experimental parameters such as temperature impair a direct comparison of most of the data sets.

In two of the above mentioned studies, however, NH RDCs were measured for chains of different length under identical experimental conditions. This is the case for the model peptides investigated by Grzesiek and co-workers (Dames et al. 2006) and full-length SN and the truncation fragment thereof studied by Sallum et al. (Sallum et al. 2005). From pure inspection of the reported dipolar couplings it is not possible to conclude whether the shorter chain exhibits stronger alignment in comparison to the longer chain. We therefore digitized reported dipolar couplings for all relevant polypeptides and determined their sequence average. We believe that this is a reasonable approach as the amino acid sequences of the polypeptides studied in each report are very similar. The sequence-averaged dipolar coupling should thus be a good measure of the overall chain alignment.

The sequenced-averaged NH dipolar coupling of the 9-residue peptide (Dames et al. 2006) is –4.79 Hz whereas in the case of the longer 12-residue peptide it equals –4.10 Hz. Similarly, for the 102-residue truncation fragment of SN (Sallum et al. 2005) we found an average NH dipolar coupling of –4.96 Hz while for the 149-residue full-length protein an average NH dipolar coupling of –4.36 Hz was computed. Therefore, in both studies the longer chain shows a (on absolute values) smaller sequenced-averaged NH dipolar coupling. This result is in agreement with the prediction made by the random flight chain model. Even the $\sim 1/\sqrt{N}$ dependence inferred from Eq. 22 is remarkably precisely represented in the calculated average dipolar couplings as is illustrated in Table 2.

We note that $^1D_{NH}$ were measured for peptides of varying length under identical experimental conditions in another study (Ohnishi et al. 2006). However, the fact that the studied sequences are either too short (6–8 residues) or RDCs are only available for part of the sequence (12 and

Table 2 Scaling of the sequence-averaged $^1D_{NH}$ with the chain length for two available independent studies (Sallum et al. 2005; Dames et al. 2006)

Ref	Length $N^{(1)}$	$\langle {}^1D_{NH}\rangle$ $(N^{(1)})$	Length $N^{(2)}$	$\langle {}^1D_{NH}\rangle$ $(N^{(2)})$	$\frac{\langle {}^1D_{NH}\rangle (N^{(2)})}{\langle {}^1D_{NH}\rangle (N^{(1)})}$	$\sqrt{\frac{N^{(1)}}{N^{(2)}}}$
Sallum et al. (2005)	102	–4.96	149	–4.36	0.88	0.83
Dames et al. (2006)	9	–4.79	12	–4.10	0.86	0.85

In each report, polypeptides exhibit similar amino acid sequences and experimental conditions did not vary. The average $^1D_{NH}$ for the 9-residue peptide is the average of the $^1D_{NH}$ of 9 different 9-residue peptides differing only by the amino acid at position 5 (peptides having I, N, Q, T, D, E, K, V, and L at position 5, respectively, were analyzed)

15 residues) did not allow us to conduct an analysis as it was undertaken for the two other studies.

Furthermore, it is important to state that in the latter studies weak alignment was accomplished by using PAG whereas our formalism was derived for the situation in which weak alignment is achieved by the presence of regularly arranged infinite planes resembling the bicelle case. The good agreement of the analyzed experimental data with the prediction made by the model might therefore indicate that the situation in polyacrylamide gels is well described by the herein presented theoretical framework. Systematic studies of the effect of chain length on the extent of alignment under identical experimental conditions are nevertheless needed in order to evaluate the prediction.

Conclusions

In this report, we present a theoretical framework for the prediction of residual dipolar couplings in unfolded proteins. The framework is rather general and can serve as a basis for determining RDCs in unfolded polypeptide chains under a wide spectrum of experimental conditions. The framework allows one to employ various models for the polypeptide chains and for aligning media in order to find RDCs with the desired degree of accuracy.

Using the framework we showed that within a simple model which approximates bicelles as infinite planes and in which unfolded polypeptides are simulated by random flight chains it is possible to obtain a closed-form analytical result for the RDCs. In this case RDCs are readily accessible for chains of differing length, for different loci along the chain and for varying bicelle concentrations.

The two general features predicted by the model are (i), a sequentially symmetric bell-shaped distribution of RDCs with center segments showing larger alignment than segments at the ends of the chain and (ii), larger alignment for shorter chains than for longer chains at a given bicelle concentration (interplanar distance L).

Experimental data available from the literature confirm the first prediction of the model, providing, therefore, a tool for recognizing fully unfolded polypeptide chains. With less certainty experimental data appear to support the second prediction as well. However, more systematic experimental studies are needed in order to validate or disprove the predictions of the model.

We note that the employed model represents the ideal case of fully unfolded protein and it additionally contains significant simplifications. Advancing the model will account for the excluded volume, the actual geometry of the alignment medium, long-range interactions, etc. Unfortunately, accounting for these factors will probably prevent one from deriving a closed-form analytical solution for the RDCs but will allow to obtain better agreement with experimental data. Such developments can be implemented within the general framework of random walk dynamics formulated in the present work.

Acknowledgements This work is partially supported by the European Commission within the Network of Excellence project EXCELL and by the EU project UPMAN. The Center for Biomoleuclar Magnetic Resonance is supported by the State of Hess. We thank Julia Wirmer for critical reading of the manuscript.

References

Ackerman MS, Shortle D (2002a) Biochemistry 41:3089–3095
Ackerman MS, Shortle D (2002b) Biochemistry 41:13791–13797
Alexandrescu AT, Kammerer RA (2003) Prot Sci 12:2132–2140
Alexandrescu AT, Abeygunawardana C, Shortle D (1994) Biochemistry 33:1063–1072
Almond A, Axelsen JB (2002) J Am Chem Soc 124:9986–9987
Azurmendi HF, Bush CA (2002) J Am Chem Soc 124:2426–2427
Bax A, Tjandra N (1997) J Biomol NMR 10:289–292
Bernado P, Blanchard L, Timmins P, Marion D, Ruigrok RW, Blackledge M (2005) Proc Natl Acad Sci USA 102:17002–17007
Bertoncini CW, Jung YS, Fernandez CO, Hoyer W, Griesinger C, Jovin TM, Zweckstetter M (2005a) Proc Natl Acad Sci USA 102:1430–1435
Bertoncini CW, Fernandez CO, Griesinger C, Jovin TM, Zweckstetter M (2005b) J Biol Chem 280:30649–30652
Binolfi A, Rasia RM, Bertoncini CW, Ceolin M, Zweckstetter M, Griesinger C, Jovin TM, Fernandez CO (2006) J Am Chem Soc 128:9893–9901
Blackledge M (2005) Prog Nucl Magn Reson Spectrosc 46:23–61
Bundi A, Wuthrich K (1979) Biopolymers 18:285–297
Chandrasekhar S (1943) Rev Mod Phys 15:1–89
Clore GM, Starich MR, Gronenborn AM (1998) J Am Chem Soc 120:10571–10572
Dames SA, Aregger R, Vajpai N, Bernado P, Blackledge M, Grzesiek S (2006) J Am Chem Soc 128:13508–13514
Ding K, Louis JM, Gronenborn AM (2004) J Mol Biol 335:1299–1307
Dyson HJ, Wright PE (2005) Nat Rev Mol Cell Biol 6:197–208
Ernst RR, Bodenhausen G, Wokaun A (1987) Principles of nuclear magnetic resonance in one and two dimensions. Oxford University Press, Oxford
Fernandes MX, Bernado P, Pons M, Garcia de la Torre J (2001) J Am Chem Soc 123:12037–12047
Ferrarini A (2003) J Phys Chem B 107:7923–7931
Fieber W, Kristjansdottir S, Poulsen FM (2004) J Mol Biol 339:1191–1199
Fiebig KM, Schwalbe H, Buck M, Smith LJ, Dobson CM (1996) J Phys Chem 100:2661–2666
Fink AL (2005) Curr Opin Struct Biol 15:35–41
Flory PJ (1953) Principles of polymer chemistry. Cornell University Press, Ithaca
Fredriksson K, Louhivuori M, Permi P, Annila A (2004) J Am Chem Soc 126:12646–12650
Gebel EB, Ruan K, Tolman JR, Shortle D (2006) J Am Chem Soc 128:9310–9311
Gillespie JR, Shortle D (1997) J Mol Biol 268:158–169

Gillespie JR, Shortle D (1997) J Mol Biol 268:170–184
Hansen MR, Mueller L, Pardi A (1998) Nat Struct Biol 5:1065–1074
Jha AK, Colubri A, Freed KF, Sosnick TR (2005) Proc Natl Acad Sci USA 102:13099–13104
Klein-Seetharaman J, Oikawa M, Grimshaw SB, Wirmer J, Duchardt E, Ueda T, Imoto T, Smith LJ, Dobson CM, Schwalbe H (2002) Science 295:1719–17222
Kristjansdottir S, Lindorff-Larsen K, Fieber W, Dobson CM, Vendruscolo M, Poulsen FM (2005) J Mol Biol 347:1053–1062
Landau LD, Lifschitz EM (1958) Course of theoretical physics, vol V. Pergamon Press, Oxford
Lord Rayleigh (1964) Lord Rayleigh collected papers, vol 6. Dover Publications, New York
Louhivuori M, Paakkonen K, Fredriksson K, Permi P, Lounila J, Annila A (2003) J Am Chem Soc 125:15647–15650
Louhivuori M, Fredriksson K, Paakkonen K, Permi P, Annila A (2004) J Biomol NMR 29:517–524
Mandelkow E-M, Mandelkow E (1998) Trends Cell Biol 8:425–427
Meier S, Guthe S, Kiefhaber T, Grzesiek S (2004) J Mol Biol 344:1051–1069
Meier S, Strohmeier M, Blackledge M, Grzesiek S (2007) J Am Chem Soc 129:754–755
Mohana-Borges R, Goto NK, Kroon GJ, Dyson HJ, Wright PE (2004) J Mol Biol 340:1131–1142
Mok YK, Kay CM, Kay LE, Forman-Kay J (1999) J Mol Biol 289:619–638
Neuhaus D, Williamson MP (2000) The nuclear overhauser effect in structural and conformational analysis, 2nd edn. Wiley-VCH
Ohnishi S, Shortle D (2003) Proteins 50:546–551
Ohnishi S, Lee AL, Edgell MH, Shortle D (2004) Biochemistry 43:4064–4070
Ohnishi S, Kamikubo H, Onitsuka M, Kataoka M, Shortle D (2006) J Am Chem Soc 128:16338–16344
Ruckert M, Otting G (2000) J Am Chem Soc 122:7793–7797
Sallum CO, Martel DM, Fournier RS, Matousek WM, Alexandrescu AT (2005) Biochemistry 44:6392–6403
Sanders CR II, Schwonek JP (1992) Biochemistry 31:8898–8905
Sass HJ, Musco G, Stahl SJ, Wingfield PT, Grzesiek S (2000) J Biomol NMR 18:303–309
Schlorb C, Mensch S, Richter C, Schwalbe H (2006) J Am Chem Soc 128:1802–1803
Schwalbe H, Fiebig KM, Buck M, Jones JA, Grimshaw SB, Spencer A, Glaser SJ, Smith LJ, Dobson CM (1997) Biochemistry 36:8977–8991
Schwarzinger S, Kroon GJ, Foss TR, Chung J, Wright PE, Dyson HJ (2001) J Am Chem Soc 123:2970–2978
Shortle D, Ackerman MS (2001) Science 293:487–489
Sibille N, Sillen A, Leroy A, Wieruszeski JM, Mulloy B, Landrieu I, Lippens G (2006) Biochemistry 45:12560–12572
Smith LJ, Bolin KA, Schwalbe H, MacArthur MW, Thornton JM, Dobson CM (1996) J Mol Biol 255:494–506
Tollinger M, Skrynnikov NR, Mulder FA, Forman-Kay JD, Kay LE (2001) J Am Chem Soc 123:11341–11352
Tycko R, Blanco FJ, Ishii Y (2000) J Am Chem Soc 122:9340–9341
Uversky VN, Li J, Fink AL (2001) J Biol Chem 276:10737–10744
Vendruscolo M, Dobson CM (2005) Philos Trans A Math Phys Eng Sci 363:433–450
Wirmer J, Schlorb C, Schwalbe H (2005) Protein folding handbook, Part I, 1st edn. Wiley-VCH, Weinheim, pp 737–794
Zweckstetter M (2006) Eur Biophys J 35:170–180
Zweckstetter M, Bax A (2000) J Am Chem Soc 122:3791–3792
Zweckstetter M, Hummer G, Bax A (2004) Biophys J 86:3444–3460