Eur. Phys. J. D **51**, 41–49 (2009)
DOI: 10.1140/epjd/e2008-00195-x

THE EUROPEAN PHYSICAL JOURNAL D

Regular article

Semistiff polymer model of unfolded proteins and its application to NMR residual dipolar couplings

M. Čubrović[1,2], O.I. Obolensky[1,a], and A.V. Solov'yov[1,b]

[1] Frankfurt Institute for Advanced Studies, Ruth-Moufang-Str. 1, 60438 Frankfurt am Main, Germany
[2] Institute of Physics, P.O. Box 57, 11001 Belgrade, Serbia

Received 14 March 2008 / Received in final form 15 September 2008 / Published online 15 October 2008
© EDP Sciences, Società Italiana di Fisica, Springer-Verlag 2008

Abstract. We present a new statistical model of unfolded proteins in which the stiffness of polypeptide backbone is taken into account. We construct and solve a mean field equation which has the form of a diffusion equation and derive the distribution function for conformations of unfolded polypeptides. Accounting for the stiffness of the protein backbone results in a more accurate description of general properties of a polypeptide chain, such as its gyration radius. We then use the distribution function of a semistiff protein within a previously developed theoretical framework [J. Biomol. NMR **39**, 1 (2007)] to determine the nuclear magnetic resonance (NMR) residual dipolar couplings (RDCs) in unfolded proteins. The calculated RDC profiles (dependence of the RDC value on the residue number) exhibit a more prominent bell-like shape and a better agreement with experimental data as compared to the previous results obtained with the random flights chain model.

PACS. 87.10.-e General theory and mathematical aspects – 82.56.Pp NMR of biomolecules – 82.56.Dj High resolution NMR

1 Introduction and motivation

High-resolution, liquid-state nuclear magnetic resonance (NMR) spectroscopy has proven to be an invaluable tool in investigation of the structure and dynamics of biomacromolecules, including folded and, recently, unfolded proteins. One of the NMR observables from which one can infer structural and dynamical information is the so-called residual dipolar couplings (RDCs) [1]. These couplings are direct dipole-dipole interactions between the spins of two nuclei, e.g., ^{15}N and a ^{1}H nuclei, detected in NMR experiments by a shift in the resonant frequency of nuclear spin flip transitions. They can be measured independently for each amino acid residue in a polypeptide chain.

Analysis of RDC profiles (dependence of the RDC value on the residue number) has been shown to be very informative in structure validation and refinement of folded proteins [1]. However, for unfolded proteins reliable interpretation of RDC measurements remains elusive even though a significant amount of experimental data has been accumulated (see [2] for a survey). Theoretically, one can predict the RDC profiles by performing numerical sampling of the conformational space of the unfolded polypeptide. For example, in [3,4] ensembles of unfolded conformations were 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. This method allows one to predict the RDC profiles with a reasonable accuracy, but it lacks the ability to explain on a basic level the obtained results, serving, therefore, as a black box with a limited use for interpretation of experimental data.

In [2] it was shown that the basic trends in RDC profiles and the underlying physical and mathematical principles leading to these trends can be revealed by statistical analysis not based on numerical sampling of conformational space (see also similar, although less mathematically rigorous, work [5]). Two general features of RDC profiles were predicted for unfolded polypeptide chains. The first one is that shorter chains have larger (in absolute value) RDCs under same experimental conditions. The second feature predicted in [2,5] is that the RDCs are larger for the amino acid residues in the middle of the chain, leading to the bell-like shape of RDC profiles. Despite the simplicity of the model (random flights chain) used in [2,5] to mimic the unfolded polypeptides, both these trends seem to be present in the experiments carried out under conditions prohibiting formation of native-like structures [2].

[a] *Present address:* National Center for Biotechnology Information, NLM/NIH, 8600 Rockville Pike, Bethesda, MD 20894, USA
[b] e-mail: solovyov@fias.uni-frankfurt.de

Our goal in this paper is to improve the quality of the model which is used for simulating the unfolded polypeptide chain. We formulate here a new statistical model of unfolded proteins in which the stiffness of polypeptide backbone is taken into account. We demonstrate that accounting for the stiffness of the protein backbone results in a more accurate description of general properties of a polypeptide chain, such as its gyration radius.

Stiff polymer model is a well-known concept in polymer physics, much used since the pioneering paper of Sato et al. [6]. The idea is to introduce an energy cost for bending of the polymer, thus favoring the extended conformations. The most versatile formalism for doing so is the so-called wormlike chain model, in which the bending energy density is proportional to the square of the curvature of the polymer contour. It has been developed to its full strength only recently [7], with the introduction of new theoretical tools to account for various generalizations and boundary conditions. The starting point in the papers cited above is the mean-field description which has the form of a diffusion equation in tangent space. However, this approximation becomes less and less satisfactory for polymers with low stiffness [8]. We show in this paper (followed by a more technical discussion in [9]) that a better continuum limit for the case of low stiffness is obtained in the real space, which turns out to have, again, the form of the diffusion equation. Also, the question of how the wormlike chain model arises from discrete stiff chains has, to the best of our knowledge, not been addressed so far. In the following section we will start from a discrete model and pass to a continuum limit, which will turn out to be exactly the low-stiffness (the usual term in polymer physics is semiflexible or semistiff) limit of the wormlike chain model.

The developed semistiff polymer model is applied to the calculation of RDCs within the theoretical framework of [2]. The calculated RDC profiles exhibit a more prominent bell-like shape and a better agreement with experimental data as compared to the previous results obtained with the random flights chain model.

In the concluding section, we will also discuss possible further steps in interpreting the RDC measurements, as well as the importance of our results from a more general perspective.

2 Theory

2.1 Introduction to the problem

We first give a general and informal consideration of the problem, before describing in more detail the interaction potential in our system and the procedure to calculate the necessary quantities. The ultimate goal is finding the relation between the dipolar couplings and the physical parameters of the unfolded polypeptide, and extracting the information on the shape and other properties of the polypeptide from the RDC measurements. Unfolded polypeptides are in many aspects similar to simple linear polymers, having no well-defined secondary structure.

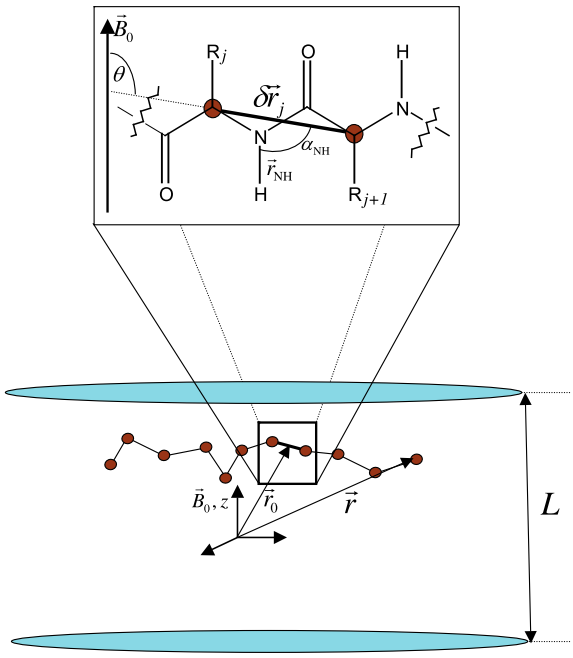


Fig. 1. Schematic picture of a polymer chain in a restricting medium modelled with a set of parallel barriers. The barrier-to-barrier distance is L. External magnetic field vector is $B_0\mathbf{e}_z$. The vectors $\mathbf{r}$ and $\mathbf{r}_0$ are the position vectors of points in the chain, defined in the fourth section. Inset: structure of a single monomer unit. The angles α_{NH} and Θ are defined for each segment. The residues (side chains) are denoted by R_j and R_{j+1}, while the radius vector of the jth segment is $\delta\mathbf{r}_j$.

For clarity we will pose the problem for a discrete chain first, although the calculations will be performed in the continuum limit. The quantity to be calculated is the dipole-dipole coupling between two nuclei. We will deal with the ^{15}N–^{1}H couplings in this paper. The energy (actually, the frequency) of the coupling is given by [2]:

$$D_{\mathrm{NH}} = \frac{\mu_0 \hbar \gamma_{\mathrm{N}} \gamma_{\mathrm{H}}}{4\pi^2 r_{\mathrm{NH}}^3} P_2(\alpha_{\mathrm{NH}}) \langle P_2(\cos\Theta)\rangle, \qquad (1)$$

where μ_0 is the magnetic permeability constant, γ_{N} and γ_{H} are gyromagnetic ratios for the nitrogen and the hydrogen nucleus and r_{NH} is the internuclear distance (assumed to be fixed). The function P_2 stands for the Legendre polynomial of the second order. The angles α_{NH} and Θ characterize the orientation of a chain segment with respect to the external field. These are in turn determined by the conformation of the chain, therefore connecting the measured couplings to the structure of the polypeptide. The average (denoted by angular brackets) is to be carried out over all possible conformations of the chain, i.e. over the state space of the chain.

The meaning of the angles α_{NH} and Θ is best seen from Figure 1. The former is the angle between the vector $\mathbf{r}_{\mathrm{NH}}$, i.e. the internuclear vector, and the axis of the jth segment, denoted in Figure 1 by $\delta\mathbf{r}_j$. On the other hand, Θ stands for the angle between $\delta\mathbf{r}_j$ and the z-axis, which is the direction of the external magnetic field.

The origin of non-zero RDCs lies in the restricting medium (bicelles or polyacrylamide gels, usually) in which the denaturated protein is solvated in experiments. Actual shape and geometry of the confining barriers may vary but the overall effect will be similar and will result in an effective confinement of the polymer. In the simplest approximation, one may regard the restricting barriers to be parallel to each other, at some distance L, as given in the Figure 1. The influence of the confinement is crucial even if the distance L is large compared to the length of the polypeptide (as is the case for typical experimental conditions). Confinement only induces a mild "deformation" of the average shape of the polypeptide coil. It is this deformation, however, which gives rise to non-zero expectation value of the term $P_2(\cos\Theta)$. The calculation of this value will be in focus of the rest of the paper.

2.2 Stiffness of polypeptide chain

The discrete chain in represented as an array of N segments, each denoted by index $j = 1 \ldots N$. The information on structure of the chain is contained in the distribution function $P(N, \mathbf{r}, \mathbf{r}_0)$ which gives probability to find the chain in a conformation with the end points at $\mathbf{r}_0$ and $\mathbf{r}$. The usual approach of polymer physics would be to write the action (or, equivalently, path integral) for the chain based on some empirical potential [10]. As we will, for the most part, work in the mean-field approximation, we will refrain from this approach and write directly the equation for the distribution function. The statistical weight of each conformation is determined by its Boltzmann weight, supposing that the system is sufficiently close to equilibrium. To specify these weights completely, one needs to employ an empirical potential for segment-to-segment interactions.

Empirical potentials for polypeptides, generally, may include the following terms: bond extensibility, bond angle stiffness, rotation about the so-called Ramachandran dihedral angles (see, e.g., [11]) and non-bonded interactions, including, possibly, Coulombic interactions, hydrogen bonds, excluded volume interactions, etc. We will deal with the non-bonded interactions in a separate publication [12]; it can be shown that these do not contribute significantly to the problem of interpreting the NMR spectroscopic data that we are primarily concerned with in this paper. Bond extensibility is, in general, negligible in polypeptides and better results are usually obtained in the framework of fixed bond lengths [13]. So, all of our segments are assumed to have the same length a.

The stiffness is an all-present effect in polymer physics and is usually characterized either by the persistence length L_p [10] or in terms of the bond angle θ and its discrepancies from some optimal value θ_0 [14]. The connection between the two descriptions is given by $L_p = ak_\theta\beta$; as usual, $\beta = (kT)^{-1}$. In our model, for typical values of k_θ and β, L_p is a few segments long. However, L_p is not very practical for nonzero θ_0. Geometrically, it is equal to the segment length of an effective freely jointed chain with the same macroscopic properties as the stiff chain. This geometric analogy is lost for $\theta_0 > 0$.

For most polymers, the optimal angle is $\theta_0 = 0$; in our case, the structure of polypeptide backbone –N–C–C–N– results in a non zero value of θ_0 [11], which is actually the tetrahedral angle characterizing the bond geometry of the carbon atom. The radius-vector of the jth segment relative to the end of the previous segment will be denoted by $\delta\mathbf{r}_j$. The bond stiffness is obviously a nearest-neighbor interaction, involving pairs of subsequent segments. Dihedral angles are for *unfolded* polypeptides usually considered in a purely local approximation, thus leading to no site-to-site interaction. Therefore, the potential of our system is of the form:

$$V = \sum_{j=1}^{N-1} V_\theta(\theta_j) + \sum_{j=1}^{N} V_{\phi\psi}(\phi_j, \psi_j). \tag{2}$$

Still, even in this approximation, distribution of dihedral angles shows nontrivial behavior if the polypeptide is not a homopolymer, i.e. if various segments have different energy minima. One final remark is that we assume various degrees of freedom to be decoupled; it is also a common approximation, and a necessary one for the problem to be tractable.

The Ramachandran part of the potential, $V_{\phi\psi}$, cannot be treated in the mean-field approximation for the reason mentioned in the previous paragraph: the energy landscape is strongly site-specific and therefore evades a description in the framework of mean-field theory. On the other hand, the decoupling of the degrees of freedom suggests that the effects of stiffness can be considered independently. In this paper, we will explore exactly the influence of stiffness, leaving the theory for Ramachandran angles for further work.

Hence, we are only interested for the potential V_θ. An often-employed potential in both analytical and numerical work, with slight differences from author to author, described in [14], is the following one:

$$V_\theta(\theta_j) = -k_\theta \cos(\theta - \theta_0) + O\left((\theta - \theta_0)^3\right). \tag{3}$$

The correction to the cosine term in (3) can be any function which is "small" compared to the leading term in the cosine, i.e. containing only third and higher order terms in the angular displacement $\theta_j - \theta_0$. It will turn out later that these corrections are, in our method, of secondary importance anyway, so the exact form of this correction is not relevant. In other words, the exact form of the anharmonic terms is not relevant; a different form would produce different higher-order terms for the diffusion coefficient but these are (by definition) beyond the scope of any model based on diffusion equation.

In the following subsection, we will give our mean-field model for a semistiff (semiflexible) chain. For some purposes the mean-field treatment can provide sufficiently accurate estimates and it is also of interest for other problems, not connected to protein physics.

2.3 Diffusion equation formalism for semistiff polymers

For the rest of this paper, we will take the continuum limit. The index of a segment in the chain (chemical coordinate s) is now a variable taking values from the interval $(0, N)$, where N is the total segment number. This framework is, of course, only suitable for the chains which are not too short.

The formalism we employ here is best suited for small stiffness; typical values of k_θ in (3) are of the order of 10ϵ, with $\epsilon = 10^{-23}$ J (0.624×10^{-4} eV), which is small compared to systems like double-stranded DNA.

One can start from the statistical weight of the jth segment having a bond angle θ_j expressed in terms of its radius-vector $\delta\mathbf{r}_j$:

$$p_0(\delta\mathbf{r}_j) = \frac{\beta k_\theta}{4\pi a^2 \sinh \beta k_\theta}\delta(\delta r_j - a)\exp(\beta k_\theta \cos\theta_j). \quad (4)$$

The above equation is nothing but the Boltzmann weight with appropriate normalization. Since the experiments with unfolded proteins are usually performed at room temperature, we take $T = 300$ K for all calculations throughout the paper. In other words, the chain is modelled as a random walk with one-step memory (which is implicitly included in (4) via the bond angle depending on the previous segment). It is a variation on the persistent random walk problem, well-known and often encountered in theory of stochastic processes [15]. The usual formalism of master equations leads to the conclusion that the continuum limit of this process is a diffusion equation; we show that in more detail in a separate publication [9].

For a three-dimensional model, diffusion coefficient becomes the diffusion tensor $\hat{\mathcal{D}}$ represented with a three-by-three matrix, the component $\mathcal{D}_{ij}$ being, by definition:

$$\mathcal{D}_{ij} = \frac{1}{2}\int d\delta\mathbf{r} p_0(\delta\mathbf{r})\delta r_i \delta r_j. \quad (5)$$

A straightforward calculation then shows that the off-diagonal components vanish; furthermore, the two "transversal" components (perpendicular to the tangent vector at the given point) are equal and will be denoted by $\mathcal{D}_\perp$; the "longitudinal" one is denoted by $\mathcal{D}_\parallel$. They are obtained to be:

$$\mathcal{D}_\perp = \frac{a^2}{\sinh\beta k_\theta}\left(\frac{\cosh\beta k_\theta}{\beta k_\theta} - \frac{\sinh\beta k_\theta}{\beta^2 k_\theta^2}\right) \quad (6)$$

$$\mathcal{D}_\parallel = \frac{a^2\cos^2\theta_0}{\sinh\beta k_\theta}\left(\frac{\sinh\beta k_\theta}{\beta^2 k_\theta^2} + \frac{\sinh\beta k_\theta}{2} - \frac{\cosh\beta k_\theta}{\beta k_\theta}\right). \quad (7)$$

The above result was derived by rotating the diffusion tensor in the local tangent coordinate system. The higher order terms can be included to modify the cosine potential, by means of perturbative corrections (so-called higher-order correlation terms) to the diffusion coefficient. The full formalism for computation of these corrections can be found in [16]. For example, the harmonic potential for the bond angle, also often employed [14] in various models, can be modelled in this way. Let us right away define also the coefficient $\mu \equiv 2\mathcal{D}_\perp/\mathcal{D}_\parallel$, as we will use it throughout the paper.

The non-isotropic diffusion tensor gives rise to the following diffusion equation:

$$\frac{\partial P}{\partial N} = \mathcal{D}_\parallel\frac{\partial^2 P}{\partial r^2} + \frac{2\mathcal{D}_\parallel}{r}\frac{\partial P}{\partial r} + \frac{\mathcal{D}_\perp}{r^2}\triangle_{S^2}P, \quad (8)$$

where $\triangle_{S^2}$ is the angular part of the Laplacian in spherical coordinates, i.e., the two-dimensional Laplace-Beltrami operator.

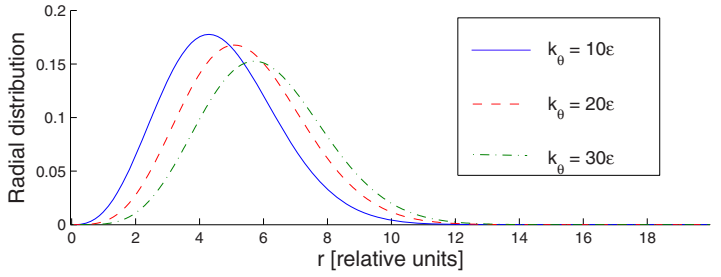
We first look for a fundamental solution (in terminology of the theory of partial differential equations), i.e. for a solution in the whole space, vanishing at the infinity and starting at $\mathbf{r}_0$, leading to the initial-boundary condition $P(0, \mathbf{r}, \mathbf{r}_0) = \delta(\mathbf{r} - \mathbf{r}_0)/4\pi$. Then one can use well-developed tools for solving diffusion equations. The easiest way is to rewrite (8) as the Schrödinger equation in imaginary time for a particle in a spherical potential given by $U(r) = \ell(\ell+1)(\mu/2-1)/r^2$. It is easy to see that $U(r)$ defined in this way is a well only for $\mu < 2$, i.e. for $\mathcal{D}_\perp < \mathcal{D}_\parallel$, otherwise, it is repulsive. The physical interpretation of this fact is that the diffusion with large $\mathcal{D}_\perp$ corresponds to the states with high angular momenta (notice the position of $\mathcal{D}_\perp$ in (8)). But since arbitrarily high angular momenta are only possible for unbounded states, this means the the imaginary time description of the diffusion corresponds to a particle in a repulsive potential. Conversely, when $\mathcal{D}_\parallel$ dominates over $\mathcal{D}_\perp$, the primary contribution to the energy comes from the radial part of the Laplace operator; hence, angular momentum cannot be arbitrarily high, which is consistent with a bounded state in a potential well.

However, one can use the same eigenbasis for both of the above cases; only the coefficients of the expansion will be different. The eigenfunctions of the radial part of the equation read as:

$$u_1(\ell, \mathcal{E}, r) = \frac{C_1(\ell, \mathcal{E})}{\sqrt{r}}J_\kappa\left(-\sqrt{\frac{\mathcal{E}}{\mathcal{D}_\parallel}}r\right) \quad (9)$$

$$u_2(\ell, \mathcal{E}, r) = \frac{C_2(\ell, \mathcal{E})}{\sqrt{r}}Y_\kappa\left(-\sqrt{\frac{\mathcal{E}}{\mathcal{D}_\parallel}}r\right), \quad (10)$$

with $\kappa = [1/4 + \mu\ell(\ell+1)/2]^{1/2}$, and correspond to the states of definite energy $\mathcal{E}$ and angular momentum ℓ of the particle. The Bessel functions of the first (second) kind are denoted by J_α and Y_α. Right away we see that $C_2 = 0$ for all $\mathcal{E}$ and ℓ, as the Bessel functions of the second kind diverge at short distances. Hence, only the (9) states contribute to the solution.



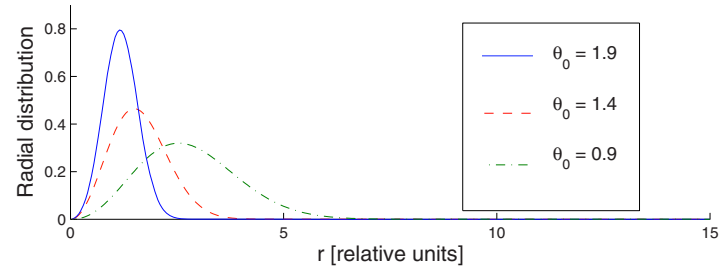


Fig. 2. (Color online) Top – radial distribution functions $P(N,r)$ for various values of stiffness, with $\theta_0 = 1.9$. Bottom – radial distribution functions $P(N,r)$ for various values of θ_0, with $k_\theta = 50\epsilon$. The length of the chain $N = 50$.

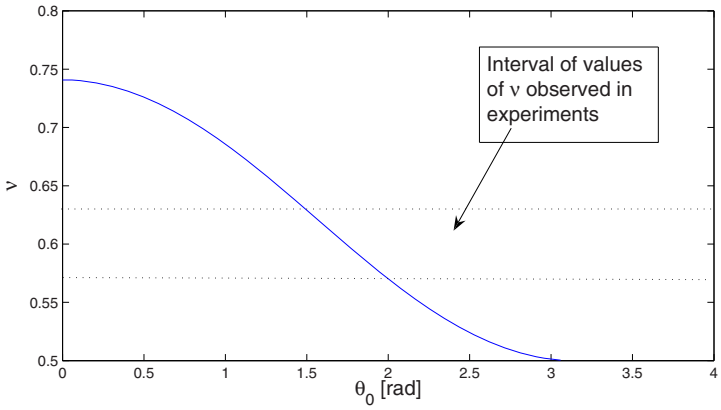


Fig. 3. Dependence of the scaling exponent ν for the gyration radius on θ_0. We have set $k_\theta = 20\epsilon$. The dashed lines denote the interval of the gyration radii measured experimentally in unfolded polypeptides (about 0.6). The exponent ν is defined through the scaling relation $\langle R_g^2 \rangle \propto N^{2\nu}$.

The solution that satisfies our boundary conditions is then obtained by standard methods and reads as:

$$P_\ell(N,\mathbf{r},\mathbf{r}_0) = C_n(\kappa,N)\left(\frac{r}{r_0}\right)^{2\kappa+1/2}\frac{1}{\mathcal{D}_\parallel N} \times \exp\left(-\frac{r^2+r_0^2}{4\mathcal{D}_\parallel N}\right)\mathcal{I}_\kappa\left(\frac{|\mathbf{r}\cdot\mathbf{r}_0|}{2\mathcal{D}_\parallel N}\right), \quad (11)$$

where $\mathcal{I}$ stands for the modified Bessel function of the first kind. The normalization constant $C_\mathrm{n}(N)$ is equal to:

$$C_n(\kappa,N) = 3\pi 2^{2+\kappa}\kappa(\mathcal{D}_\parallel N)^{1+\kappa/2}\Gamma(1+\kappa)\Gamma(3\kappa/2) \times_1 F_1\left(1+3\kappa/2, 1+\kappa, \frac{r_0^2}{4\mathcal{D}_\parallel N}\right), \quad (12)$$

with ${}_1F_1(a,b,x)$ denoting the confluent hypergeometric function of its arguments (see [17] for a definition). We obtain the above result by expanding (11) into power series, integrating and resuming. Dependence of the normalization constant on N and κ is explicitly written, as this dependence will play a role in later sections. The large-N asymptotic form of $C_\mathrm{n}(\kappa,N)$ reads as:

$$C_\mathrm{n}(\kappa,N) \approx 3\pi 2^{2+\kappa}\kappa(\mathcal{D}_\parallel N)^{1+\kappa/2} \times \Gamma\left(\frac{3\kappa}{2}\right)\left(1+\frac{3\kappa+2}{2\kappa+2}\frac{r_0^2}{4\mathcal{D}_\parallel N}\right). \quad (13)$$

We will need this asymptotic form later on. Notice that the normalization constant is dependent on N, as one would expect. The solution explicitly depends on ℓ. It is actually the sum of all partial waves (characterized by the value of ℓ) that provide a solution of finite norm (i.e. no scattering to infinity, which would correspond to the "blowup" of the chain) and finite localization radius (i.e. no "falling to the center", which would correspond to the collapse of the chain).

To understand better the general properties of the model, it is helpful to analyze the behavior of the solution (11) depending on the parameters k_θ and θ_0. We will first discuss the radial distribution function $P(N,r)$, defined as $\int\int d\phi\, d\theta \sin\theta r^2 P(N,\mathbf{r},\mathbf{r}_0)$. The results for selected values are given in Figure 2. It is seen that in the whole physically sensible range of parameters, the stiffness

k_θ, once it moves away from zero, only mildly flattens the distribution function. On the other hand, the bending angle does influence it significantly. An informal explanation is that letting k_θ grow, provided it is neither too close to zero nor too large, results in less bending of the chain but (as k_θ is not very large) the chain still does bend relatively often and still has the shape of a slightly flattened sphere; therefore, various parts of the chain still propagate in almost uncorrelated directions and it is not very important how long (on average) they are. On the other hand, large preferential angle θ_0 brings a systematic effect, which accumulates and substantially changes the shape of the chain.

It is also instructive to look at the behavior of the gyration radius (expectation value of the squared distance from the center of mass of the polymer), given in Figure 3. As one could expect, it grows significantly with θ approaching zero, as in that case, the most extended configurations are preferred. This case agrees with the equations for the gyration radius cited in [10]. On the other hand, for values of θ_0 close to π, the chain, on a macroscopic scale, behaves almost as a Gaussian freely jointed chain, hence ν approaches its Gaussian value 0.5. The gyration radius for the continuum limit in the case of non-zero θ_0 has, to the best of our knowledge, not been addressed so far. The plot in Figure 3 shows the range of the preferential bond angle values that correspond to the experimental result, $\nu \approx 0.6$ [3]. This range roughly corresponds to the value of θ_0 suggested by the geometry of bonds in polypeptides: $\theta_0 \approx 1.8$ radians [14]; in our calculations, we use $\theta_0 = 1.83$ radians. Hence, our model is able to reproduce the observed scaling exponent of the gyration radii and allows us to conclude that the proximity of its value to the scaling exponent ν_{saw} of the self-avoiding random walk ($\nu_{saw} \approx 0.59$) is pure coincidence. The distribution of bond angles (which are dominant degrees of freedom in terms of typical energies and time scales) alone accounts for the gyration radius scaling, whereas the self-avoidance (together with other non-bonded interactions) only produces small corrections (for more details see [12]).

3 The calculation of RDC values

3.1 Basic considerations

Having described our model of unfolded polypeptides, we now turn to the calculation of RDC values. The general theory is given in [2] and the basic idea is also mentioned in the introduction section of this paper. Here we discuss the more formal aspects of the procedure and state the results.

We will consider the simplest model, in which the restricting medium is modelled as a set of parallel planar absorbing barriers at the distance L from each other, as shown in Figure 1 [2]. This effectively means that all the paths which intersect the barrier are removed from consideration. The exact solution with these boundary conditions is difficult to find; therefore, we resort to the method of images, common in problems such as diffusion and electrostatics [19]. Staying at the first order approximation, the solution reads as [2]:

$$f(N,\mathbf{r},\mathbf{r}_0) = P(N,\mathbf{r},\mathbf{r}_0) - P(N,\mathbf{r}',\mathbf{r}_0) - P(N,\mathbf{r}'',\mathbf{r}_0), \tag{14}$$

with $\mathbf{r}'$ and $\mathbf{r}''$ being the points symmetric to $\mathbf{r}$ with respect to the barriers, and $f(N,\mathbf{r},\mathbf{r}_0)$ denoting the probability density function for the appropriate boundary conditions (whereas P stands for the fundamental solution in the whole space).

As can be seen from the defining expression, the RDC of the jth segment is determined by the value of the angle Θ of the $C_j^\alpha - C_{j+1}^\alpha$ segment with respect to the direction of the magnetic field. Therefore, we wish to find the distribution function for this angle, denoted $f(\Theta)$. As elucidated in more detail in [2], $f(\Theta)$ equals the joint cumulative distribution for a chain of length j starting at $\mathbf{r}_j - \delta\mathbf{r}_j/2$ and reaching some position $\mathbf{r}_{01}$, and a chain of length $N-j$ starting at $\mathbf{r}_j + \delta\mathbf{r}_j/2$ and ending at some position $\mathbf{r}_{02}$. Cumulative distribution, by definition, enumerates the states with prescribed position of one end of the chain ($\mathbf{r} \pm \delta\mathbf{r}_j/2$), independently of the coordinate of the other end ($\mathbf{r}_{01}$ or $\mathbf{r}_{02}$). Therefore, for $C(N,\mathbf{r})$ we have, in general:

$$C(N,\mathbf{r}) = \int d\mathbf{r}_0 f(N,\mathbf{r},\mathbf{r}_0). \tag{15}$$

At this point, one should notice that the dependence of the cumulative distributions on x and y vanishes due to symmetry. We may therefore denote them by $C(N,z)$. Integrating over the initial positions of the chains, $\mathbf{r}_{01}$ and $\mathbf{r}_{02}$, we arrive at the following equation for $f(\Theta)$:

$$f(\Theta) \propto \int_{-L/2}^{L/2} dz\, C\left(j, z - \frac{a\cos\Theta}{2}\right) \times C\left(N-j, z + \frac{a\cos\Theta}{2}\right), \tag{16}$$

where the proportionality constant is easy to determine from the previous equations, keeping track of all constant factors from the beginning. We have exploited the fact that the segment lengths are all equal ($|\delta\mathbf{r}_j| = a$), as well as the definition of the angle Θ.

3.2 Angular averaging: elementary method

The final step is performing the necessary integrations, i.e., calculating the average over $f(\Theta)$. Conceptually the simplest way of doing this is expanding (16) in a power series and integrating it term by term. This is the most feasible way for obtaining quick, low-accuracy estimates. We first sketch this method.

One starts by expanding the cumulative distribution functions in powers of $a\cos\Theta$. The odd terms obviously vanish. The even terms are then integrated by parts bearing in mind the fact that the distribution $f(N,\mathbf{r},\mathbf{r}_0)$ vanishes at the boundaries due to confinement. The averaging

in equation (1) is then readily performed. The result up to the fourth order in a has the form:

$$\begin{aligned} D_{\mathrm{NH}} \;=\; & K P_2(\alpha_{\mathrm{NH}})\times \\ & \Bigg[B \;-\frac{3}{8}\frac{C_n(\kappa,N_1)C_n(\kappa,N-N_1-1)}{C_n(\kappa,N)}a^2 \\ & -\frac{1}{384}\frac{C_n(\kappa-1,N_1)C_n(\kappa-1,N)}{C_n(\kappa,N)}a^2 \\ & +\frac{1}{64}\frac{C_n(\kappa-1,N_1)C_n(\kappa-1,N-N_1-1)}{C_n(\kappa-1,N)}a^4\Bigg]. \end{aligned} \tag{17}$$

The constant K is defined in (1) and reads as:

$$K=\frac{\mu_0\hbar\gamma_{\mathrm{N}}\gamma_{\mathrm{H}}}{4\pi^2 r_{\mathrm{NH}}^3}. \tag{18}$$

The constant term B in (17) is small (about two orders of magnitude smaller than the a-dependent terms) and we ignore it in our calculations. When calculating D_{NH}, we have used the asymptotic form for the normalization constant, as given in (13). Bearing in mind the limited accuracy of our formalism (simple toy-model potential, mean-field approach, etc.), one may safely ignore also the quartic term, as well as the second quadratic term (due to its large denominator). We have found for the examples below that this approximation leads to insignificant changes of the computed RDC curves.

3.3 Angular averaging: advanced method

A more elaborate but substantially more general scheme, allowing in principle calculations of arbitrarily high accuracy, is based upon the formalism of quantum theory of angular momentum. We again start from (16), which is the exact result (not approximate, like the series expansion subsequently performed in the previous subsection). The idea is to refrain from using the closed-form solution (11) and use the series expansion of $P(N,\mathbf{r},\mathbf{r}_0)$ over the radial eigenfunctions (9) and spherical harmonics. The former is more convenient and more informative for most purposes but the latter allows us to use numerous identities of the angular momentum theory to obtain simpler expressions for the average of $P_2(\cos\Theta)$.

The starting point is the solution in the whole space:

$$P(N,\mathbf{r},\mathbf{r}_0)=\sum_{\ell=0}^{\infty}A_\ell R_\ell(N,|\mathbf{r}-\mathbf{r}_0|)P_\ell(\cos\theta), \tag{19}$$

where A_ℓ are the appropriate coefficients determined by the eigenfunctions (9), R_ℓ are radial functions, obtained by integrating the eigenfunctions over the "energy" variable $\mathcal{E}$, and θ is the azimuthal angle. Reflections in the planes $z=\pm L/2$, which give the image solutions, are then readily obtained in the form $\hat{\mathcal{R}}_{\pi,\mathbf{e}_z}\hat{P}\hat{T}_{\pm\frac{L}{2}\mathbf{e}_z}P(N,\mathbf{r},\mathbf{r}_0)$, which is easy to prove from elementary considerations. The operators denote the rotation for a given angle about the given axis, spatial inversion and spatial translation for a given vector, in that order. The rotation for π about the z-axis and the spatial inversion act upon the angular part simply by multiplying it with $(-1)^l$. Only the translation has a nontrivial action. A lengthy calculation, making use of the Wigner functions and identities with Clebsch-Gordan coefficients as given in [20], results in:

$$\begin{aligned} P_\ell(\theta',\phi) =& K_\ell\sum_{\lambda=0}^{\infty}(-1)^{\lambda+\ell}f_r(\lambda,\ell,r) \\ & \times\sum_{\Lambda=|\ell-\lambda|}^{\ell+\lambda}\frac{(2\ell+1)(2\lambda+1)}{2\Lambda+1}|\langle\Lambda 0|\ell 0\lambda 0\rangle|^2. \end{aligned} \tag{20}$$

For the left image, where we have introduced the notation:

$$\begin{aligned} K_\ell=\sum_{j=0}^{\ell} & \left(-\frac{L}{2}\right)^{\ell-j}\sqrt{2(\ell-j)-1}\frac{(\ell+j)!}{(\ell-j)!} \\ & \times\left[\frac{(2\ell)!(2j)!}{(2\ell+2j)!}\right]^{1/2}, \end{aligned} \tag{21}$$

$$\begin{aligned} f_r(\lambda,\ell,r) =& \frac{(2rL)^\ell}{(4r^2+L^2)^{\lambda+\ell/2}}\;F \\ & \times\left(\frac{2\lambda+\ell}{4},\;\frac{2\lambda+\ell+2}{4};\;\lambda+\frac{3}{2},\frac{16r^2L^2}{(4r^2+L^2)^2}\right), \end{aligned} \tag{22}$$

and the angular brackets stand for the Clebsch-Gordan coefficients, whereas F is the confluent hypergeometric function, and the angle in the new coordinate system is denoted by θ'. The other image looks the same, except that the functions (22) now contain $-L$ in place of L. Finding the cumulative distribution is straightforward, inserting the expressions for $P(N,\mathbf{r},\mathbf{r}_0)$ and its two images in (15) and integrating. Notice that the initial position is contained only in the radial functions R_ℓ, which can be integrated analytically as their integrals reduce to the familiar Bessel integrals.

The last step is multiplying the two cumulative distribution functions as in (16) and integrating over z. At this step the symmetry of the problem nullifies all the terms containing the product of an even and an odd functions of z, and the triangle rule for addition of angular momenta further reduces the number of non-zero terms. We are thus left with a *finite* sum which, to the second order, gives the result for $P_2(\cos\Theta)$ that coincides with (17). The fourth order term differs from the corresponding term in (17), however, as in this approach, due to the orthogonality of the Legendre polynomials, we capture the *exact* value of the coefficient in front of the fourth (or any other desired) order term in the expansion. In the elementary method, the expansion is in powers of a *and* in number of images. The latter expansion is an uncontrollable approximation, since the $2n$th image can contribute a term of the order $2n-2$ in a. The advanced method captures the whole contribution of given order in a.

At the present level of accuracy of our model, this increase in accuracy is not crucial. However, the generality of the formalism applied in this subsection might prove to be necessary if the subtler effects as the Ramachandran rotations or long-range interactions are included. Also, the described method allows an equally straightforward calculation of the expectation value of $P_\ell(\cos\Theta)$ for any ℓ. This case will appear if other observables in addition to direct dipole-dipole couplings are measured. We therefore propose this approach for any further work on this problem.

4 Examples

The purpose of this section is to test our predictions on experimental data and judge the accuracy and usefulness of our theory. Therefore, we do not analyze in detail any of the systems and contend ourselves just to compare our curves with the experimental ones.

In all the cases that we consider, we take the stiffness k_θ equal to 20ϵ and preferential bond angle $\theta_0 = 1.83$ (in radians). These values have been recommended in [14] and also according to other authors the peptide bond is expected to be well described by these values. The interplanar distance is taken fixed to $L = 100a$, where a is the length of a single segment. In both experiments that we analyze [21,22], this length is cited to be about 40 nm while the segment length is 0.38 nm. So, our value for the interplanar distance roughly corresponds to the experimental one; exact equality is not essential since the experimental setup is difficult to control concerning the interplanar distance [22] and the actual distribution of interplanar values is probably rather fat-tailed. In our theory, the segment length does not enter the final expressions and therefore can take any arbitrary value.

The first case we consider is the urea-denaturated apomyoglobin, an experiment reported in [21], and analyzed also in the previous study by two of the authors [2]. The result is seen in Figure 4. The same paper also reports on measurements of acid denaturated apomyoglobin, which cannot be well described with our model, probably because the native-like topology is still present in this case, as the authors themselves state [21].

Another example is ubiquitin, one of the proteins which are intrinsically disordered also in their native state. The measurements are taken from [22].

The second case, in Figure 5, shows somewhat better agreement with experiment than the first one (χ^2 about 30 percent better). In part, this is probably due to the difference between the two proteins. Ubiquitin is known to be a strongly disordered protein [22] and behaves essentially as a perfect statistical coil, so various local deviations from the mean value of RDCs tend to average out. On the other hand, apomyoglobin probably retains some native-like structure even in the unfolded state; this is particularly probable for the regions formed by several subsequent segments which are completely above or below the average RDC value, that are visible in the measurements given in Figure 4. These are probably regions with strong close-neighbor interactions, that behave like partially folded secondary structures and therefore choose one of the conformations, some of them with significantly higher probability than the others. Finally, we point out that both examples demonstrate that the current model provides a more realistic description of the polypeptide than the non-interacting random flight model.

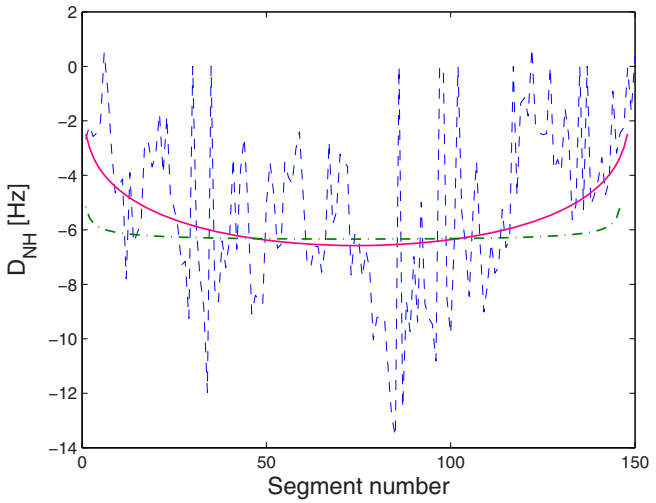


Fig. 4. (Color online) Comparison of experimental (blue dashed line) and theoretical (red full line) RDCs for unfolded apomyoglobin. The prediction of the random flight theory [2] is also shown (green dash-dotted line). General bell shape is observed but it is obvious that local conformational properties induce large deviations from the mean field curve, predicted by our model.

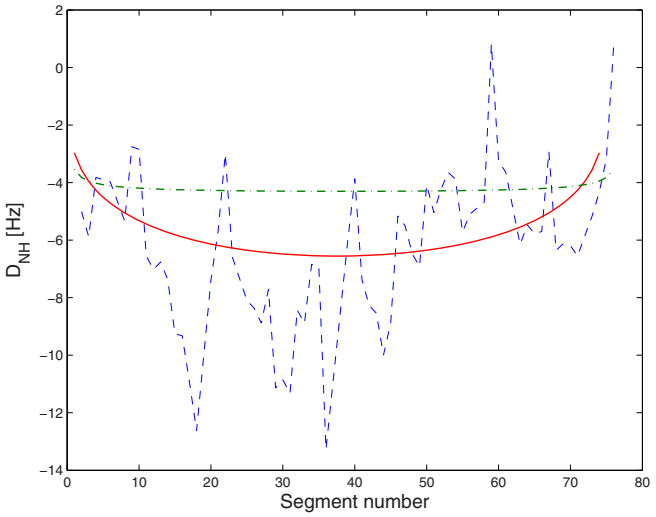


Fig. 5. (Color online) Comparison of experimental (blue dashed line) and theoretical (red full line) RDCs for unfolded ubiquitin. The random flight theory prediction [2] is also shown (green dash-dotted line). One again sees the local variations superimposed on the global bell-shaped curve.

5 Discussion and conclusion

We have presented a theoretical method for reproducing the spatial structure of unfolded polypeptides, in particular the NMR spectroscopic measurements of NH dipolar couplings. Our approach requires the use of the empirical potentials and model parameters; therefore, it is not an ab initio approach. Nevertheless, all the parameters entering the calculation are either measured (or controlled) in the experiment itself (temperature, interplanar distance) or more or less standard and well-known values (optimal bond angle, bond stiffness). Bond stiffness is the "most empirical" of them but it also seems to vary very little in various numerical models [14,18]. The results seem encouraging and reveal general properties of disordered proteins.

First, it seems that the assumption of the effective decoupling of the degrees of freedom is justified by the RDC curves. The global shape of the chain, which is determined primarily by the statistical nature of polypeptide chain conformations in unfolded state and is well described within the semistiff polymer model, gives rise to the bell shape of the curves, also detected in experiments. On the other hand, the specificities of the segments lead to the local deviations of the RDC values from the smooth bell-shaped distribution. We plan to extend our model in further work, applying the linear response theory in order to reproduce these local structures.

The method will be subject to a number of improvements in the future. Besides applying the linear response formalism to improve the results, we also plan to asses in more detail the influence of long range interactions and intrachain contacts. Also, it is possible to use the results of the numerical work to identify the optimal Ramachandran angles for unfolded polypeptides. This will allow us to account for even richer secondary structure than that produced by a restricted database search, as the problem of weighting would be eliminated, with the energy values of different (ϕ, ψ) points being read off numerically obtained potential energy surfaces.

We are grateful to NoE EXCELL for the support of this research, as well as to Alexander Yakubovich for helpful discussions. M. Č. also acknowledges his gratitude to FIAS and Meso-Bio-Nano group for financial support and warm hospitality during the work on this project. Work at the Institute of Physics is supported by the Ministry of Science Project OI141031.

References

1. M. Blackledge, Prog. Nucl. Magn. Reson. Spectrosc. **46**, 23 (2005)
2. O.I. Obolensky, K. Schlepckow, H. Schwalbe, A.V. Solov'yov, J. Biomol. NMR **39**, 1 (2007)
3. A.K. Jha, A. Colubri, K.F. Freed, T.R. Sosnick, Proc. Natl. Acad. Sci. **102**, 13099 (2005)
4. P. Bernado et al., Proc. Natl. Acad. Sci. **102**, 17002 (2005)
5. M. Louhivuori et al., J. Am. Chem. Soc. **125**, 15647 (2003)
6. N. Sato, K. Takahashi, Y. Yunoki, J. Phys. Soc. Jpn **22**, 219 (1967)
7. B. Hamprecht, H. Kleinert, Phys. Rev. E **71**, 031803 (2005); H. Kleinet, A. Chervyakov, J. Phys. A: Math. Gen. **39**, 8231 (2006)
8. H. Kleinert, *Path integrals in quantum mechanics, statistics, polymer physics and financial markets* (Berlin, 2003)
9. M. Čubrović, O.I. Obolensky, A.V. Solov'yov, submitted
10. K. Kroy, E. Frey, Phys. Rev. Lett. **77**, 2581 (1996); H. Changbong, D. Thirumalai, J. Chem. Phys. **124**, 104905 (2006)
11. A.B. Rubin, *Biophysics* (Nauka, Moscow, 2004) (in Russian)
12. M. Čubrović, O. Obolensky, A.V. Solov'yov, in preparation
13. R.K. Murarka, A. Liwo, H.A. Scheraga, J. Chem. Phys. **127**, 155103 (2007)
14. D.K. Klimov, D. Thirumalai, Phys. Rev. Lett. **79**, 317 (1997); T. Veitshans, D.K. Klimov, D. Thirumalai, Fold. Des. **7**, 11 (1997)
15. B.D. Hughes, *Random Walks and Random Environments* (Clarendon Press, Oxford, 1995)
16. G.M. Zaslavsky, Phys. Rep. **371**, 461 (2002); W. Ebeling, I. Sokolov, *Statistical Thermodynamics and Stochastic Theory of Nonequilibrium systems* (World Scientific, 2002)
17. A.D. Polianin, *Handbook of linear partial differential equations* (CRC Press, 2002)
18. I.A. Solov'yov, A.V. Yakubovich, A.V. Solov'yov, W. Greiner, Eur. Phys. J. D **46**, 227 (2007) ; A.V. Yakubovich, I.A. Solov'yov, A.V. Solov'yov, W. Greiner, Eur. Phys. J. D **46**, 215 (2007)
19. E.M. Lifshitz, L.D. Landau, L.P. Pitaevskii, *Electrodynammics of continuous media* (World Scientific, 1989)
20. D.A. Varshalovich, A.N. Moskalev, V.K. Khersonskii, *Quantum theory of angular momentum* (World Scientific, 1988)
21. W. Fieber, S. Kristjansdottir, F.M. Poulsen, J. Mol. Biol. **339**, 1191 (2004)
22. S. Meier, S. Grzesiek, M. Blackledge, J. Am. Chem. Soc. **129**, 9799 (2007)
23. H.M. Berman et al., Nucl. Acids Res. **28**, 235 (2003)