J|A|C|S
ARTICLES
Published on Web 11/20/2003

On the Origin of Residual Dipolar Couplings from Denatured Proteins

Martti Louhivuori,[†] Kimmo Pääkkönen,[‡] Kai Fredriksson,[§] Perttu Permi,[||] Juhani Lounila,[⊥] and Arto Annila*,[||]

Contribution from the Department of Physical Sciences, University of Helsinki, Finland, VTT Biotechnology, Espoo, Finland, Department of Biosciences, University of Helsinki, Finland, Institute of Biotechnology, University of Helsinki, Finland, and Department of Physical Sciences, University of Oulu, Finland

Received April 2, 2003; E-mail: Arto.Annila@helsinki.fi

Abstract: Effects of steric obstruction on random flight chains are examined. Spatial probability distributions are elaborated to calculate residual dipolar couplings and residual chemical shift anisotropy, parameters that are acquired by NMR spectroscopy from solutes dissolved in dilute liquid crystals. Calculations yield chain length and residue position-dependent values in good agreement with simulations to provide understanding of recently acquired data from denatured proteins.

Introduction

Residual dipolar couplings (RDCs) acquired by NMR spectroscopy provide valuable structural information on biological macromolecules.[1] RDCs have been used in structure determination and refinement as well as in assembling protein models from peptide fragments[2] and building macromolecular complex from known subunits.[3] Residual dipolar couplings arise from the anisotropic molecular tumbling that can be enhanced by liquid crystal particles or axial matrixes.[4] RDCs depend on the average orientation, i.e., $\langle P_2(\cos\theta)\rangle = (3\langle\cos^2\theta\rangle - 1)/2$, of the internuclear vector with respect to the magnetic field. Likewise, chemical shifts depend on $\langle P_2\rangle$ of the shielding tensor.[5] The enhancement of alignment by obstruction is a well-understood phenomenon for rigid macromolecules, e.g., folded proteins.[6−8]

Recently, RDCs were measured from denatured proteins.[9] However, interpretations remained controversial, and conclusions about topology and spatial properties of the denatured ensemble were counterintuitive.[9] The data were even seen to support certain folding models.[10] It was implicitly assumed that a fully denatured protein would give zero RDCs per se. However, this assumption ignores classical results about random flight chains. Kuhn already had pointed out long ago that various segments of a random flight chain are not distributed spherically relative to the end-to-end displacement axis.[11] Nagai calculated the average orientation of a segment relative to the end-to-end direction,[12] and Solc and Stockmayer worked further to show that an instantaneous chain conformation departs from the spherical symmetry,[13] although the long time average is spherically symmetrical about the molecular center of mass. Fixman proceeded to compute radii of gyration for random flight chains of various lengths,[14] and Forsman and Hughes extended this to include chains in the vicinity of an interface.[15] In this paper, we calculate RDCs for a random flight chain from the first principles. With the proper understanding gained from these results, the unique power of RDCs as well as residual chemical shift anisotropy (RCSA) can be properly exploited to probe various loci along the chain.

Calculations

Our objective was to calculate residual dipolar couplings for flexible molecules, i.e., polymers without a definite shape (Figure 1). To show that the mere structure of a chain will cause nonvanishing RDCs, we

[†] Department of Physical Sciences, University of Helsinki.
[‡] VTT Biotechnology.
[§] Department of Biosciences, University of Helsinki.
[||] Institute of Biotechnology, University of Helsinki.
[⊥] Department of Physical Sciences, University of Oulu.

(1) (a) Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111−1114. (b) Mueller, G. A.; Choy, W. Y.; Yang, D.; Forman-Kay, J. D.; Venters, R. A.; Kay, L. E. *J. Mol. Biol.* **2000**, *300*, 197−212. (c) Bax, A.; Kontaxis, G.; Tjandra, N. *Methods Enzymol.* **2001**, *339*, 127−174. (d) Prestegard, J. H.; Al-Hashimi, H. M.; Tolman, J. R. *Q. Rev. Biophys.* **2000**, *33*, 371−424. (e) De Alba, E.; Tjandra, N. *Prog. NMR Spectrosc.* **2002**, *40*, 175−197.
(2) Delaglio, F.; Kontaxis, G.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 2142−2143.
(3) (a) Clore, M. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9021−9025. (b) Fischer, M. W.; Losonczi, J. A.; Weaver, J. L.; Prestegard, J. H. *Biochemistry* **2000**, *38*, 9013−9022. (c) Mattinen, M. L.; Pääkkönen, K.; Ikonen, T.; Craven, J.; Drakenberg, T.; Serimaa, R.; Waltho, J.; Annila, A. *Biophys. J.* **2002**, *83*, 1177−1183.
(4) (a) Bax, A.; Tjandra N. *J. Biomol. NMR* **1997**, *10*, 289−92. (b) Meier, S.; Haussinger, D.; Grzesiek, S. *J. Biomol. NMR* **2002,** *24*, 351−356. (c) Tycko, R.; Blanco, F. J.; Ishii, Y. *J. Am. Chem. Soc.* **2000**, *122*, 9340−9341.
(5) Wu, Z.; Tjandra, N.; Bax, A. *J. Am. Chem. Soc.* **2001**, *123*, 3617−3618.
(6) Bax, A.; Kontaxis, G.; Tjandra, N. *Methods Enzymol.* **2001**, *339*, 127−174.
(7) Zweckstetter M.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 3791−3792.
(8) Fernandes, M. X.; Bernado, P.; Pons, M.; Garcia de la Torre, J. *J. Am. Chem. Soc*. **2001**, *123*, 12037−12047.
(9) (a) Shortle, D.; Ackerman, M. S. *Science* **2001**, *293*, 487−489. (b) Ackerman, M. S.; Shortle, D. *Biochemistry* **2002**, *41*, 3089−3095. (c) Ackerman, M. S.; Shortle, D. *Biochemistry* **2002**, *41*, 13791−13797. (d) Shortle, D. *Adv. Protein Chem.* **2002**, *62*, 1−23. (e) Ohnishi, S.; Shortle, D. *Proteins: Struct., Funct., Genet.* **2003**, *50*, 546−551.
(10) (a) Plaxco K. W.; Gross, M. *Nat. Struct. Biol.* **2001**, *8*, 659−660. (b) Makarov, D. E.; Plaxco, K. W. *Protein Sci.* **2003**, *12*, 17−26.
(11) Kuhn, W. *Kolloid-Z.* **1934**, *68*, 2.
(12) (a) Nagai, K. *J. Chem. Phys.* **1964**, *40*, 2818. (b) Flory, P. J. *Statistical Mechanics of Chain Molecules*; Interscience Publishers: New York, 1969.
(13) Solc K.; Stockmayer, W. H. *J. Chem. Phys.* **1971**, *54*, 2756.
(14) Fixman, M. *J. Chem. Phys.* **1962**, *36*, *306*.
(15) Forsman, W. C.; Hughes, R. E. *J. Chem. Phys.* **1963**, *38*, 2118.
(16) Chandrasekhar, S. *Rev. Mod. Phys.* **1943**, *15*, 1−89.

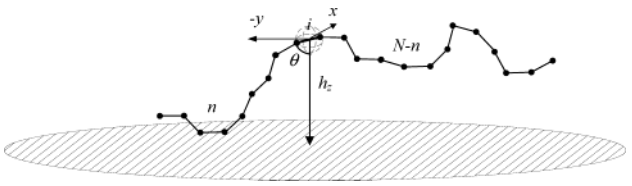

Figure 1. Random flight conformer with $N + 1$ segments hovering above an obstructing plane at height h_z. At locus i, placed at the origin, the chain is partitioned to two half chains containing n and $N - n$ segments each. Residual dipolar couplings and chemical shift anisotropy are functions of the average angle θ of i depicted here when the external magnetic field is chosen to be parallel with the z-direction.

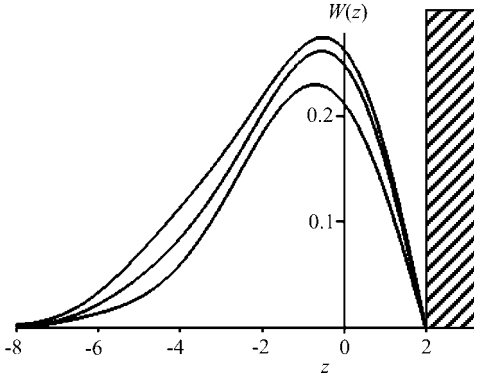

Figure 2. Random flight binomial distribution depicted for a nine-residue chain ($N + 1 = 9$) when $n = 7$ and $N - n = 1$ in the vicinity of the barrier ($h_z = 2$). When the segment i at the origin has the orientation $\theta = 0$ (lowest curve), the long half chain is at the closest to the barrier and is obstructed most. When $\theta = \pi$ (uppermost), the obstruction is at its minimum. The curve for $\theta = \pi/2$ is also shown (middle).

employ a simple model, the random flight chain. The results serve to show how RDCs originate from the structure of a chain. Undoubtedly, natural polypeptides display details beyond this simplification. To begin with, we formulate spatial distributions for an ensemble of conformers under the assumption that no conformational changes are induced by the obstruction itself, and then we proceed to calculate the average orientation of each segment in the chain.

For a given locus i in a chain that consists of $N + 1$ segments there are n segments before i and $N - n$ segments after i. Therefore, when viewed from the locus i the spatial probability distribution function W also contains two parts: W_n and W_{N-n}. The local viewpoint is justified because RDC is a local probe, not a whole-chain average. We chose the magnetic field $\mathbf{B}_o$ as well as the normal of the obstructing plane along the z-direction. This geometry corresponds to the case where the director $\mathbf{n}$ of the medium is parallel to $\mathbf{B}_o$. For the other stable liquid crystal condition, i.e., $\mathbf{n} \perp \mathbf{B}_o$, the outcome of the following calculation scales by a factor of $-1/2$. Axial matrixes can be oriented to any direction, e.g., along the magic angle to suppress dipolar contribution.

Proteins are often modeled as valence chains to account for the steric hindrance between residues. By introducing the concept of statistical element, any valence chain that is sufficiently long can be described by a Gaussian chain, i.e., a random flight chain. For proteins, one statistical element will encompass at least several residues. The precise number depends on the amino acid sequence. Therefore, our interest is on rather short Gaussian chains. The appropriate starting point for the calculation is the binominal distribution given by Chandrasekhar[16] and denoted here as W^{free}. To ensure that the chain will not penetrate into the barrier, the probability must vanish at the surface. To this end, the effect of the barrier itself is modeled as a distribution W^{bar}. The distribution for the obstructed chain is then obtained as $W^{obs} = W^{free} - W^{bar}$.[16] Furthermore, the position of each half-chain distribution depends on the direction of segment i placed in the origin. The direction of segment i is conveniently given by the spherical coordinates of θ and ϕ. According to the geometry in Figure 1, the two distributions, one

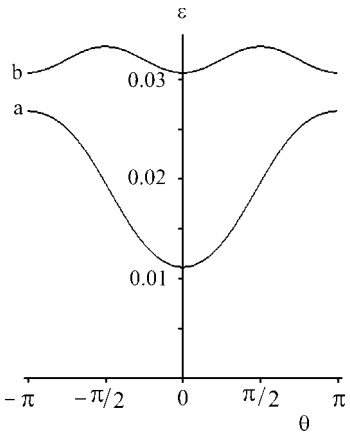

Figure 3. The error $\epsilon = (I^{exp} - I^{bin})/I^{bin}$ as a function of θ that is introduced when replacing the obstructed binomial distribution by the obstructed exponential distribution for a nine-residue chain ($N + 1 = 9$) placed at the distance $h_z = 2$ from the barrier (a) near the end of chain ($n = 7$, $N - n = 1$) and (b) in the middle of the chain ($n = 4$, $N - n = 4$).

for each half chain, can be divided into the Cartesian base. The distributions along the x- and y-direction are the familiar free binomial distributions. The modulation on the position of W_n and W_{N-n} in the x,y-plane due to ϕ is inconsequential for the calculation of $\langle P_2 \rangle$ from the axial symmetric distributions. However, the modulation due to θ, a deceptively tiny effect, is the sole source of the nonvanishing $\langle P_2 \rangle$. Hence, we proceed to consider only the z-part of the distribution, with the aim of calculating $\langle P_2 \rangle$ for every locus i in the chain (Figure 2).

As exemplified by Figure 2, the angle θ will affect the concentration of various chain conformations. To facilitate the computation, we replace the binomial distribution by an exponential distribution (eq 1). For simplicity, we define the length of the segment l to be unity.

$$W(z,h_z,c,n,N) = \frac{1}{N}\{nW_n^{obs}(z,h_z,c,n) + (N-n)W_{N-n}^{obs}(z,h_z,c,N-n)\}$$

$$= \frac{1}{N}\{n[W_n^{free}(z,c,n) - W_n^{bar}(z,h_z,c,n)] + (N-n)[W_{N-n}^{free}(z,c,N-n) - W_{N-n}^{bar}(z,h_z,c,N-n)]\}$$

$$= \frac{\sqrt{2n/\pi}}{N}\left\{\exp\left[-\frac{(z+1/2c)^2}{2n}\right] - \exp\left[-\frac{(2h_z - z + 1/2c)^2}{2n}\right]\right\} + \frac{\sqrt{2(N-n)/\pi}}{N}\left\{\exp\left[-\frac{(z-1/2c)^2}{2(N-n)}\right] - \exp\left[-\frac{(2h_z - z - 1/2c)^2}{2(N-n)}\right]\right\}$$

where $c = \cos\theta$ (1)

We estimate the error done by substituting the binomial with the exponential distribution by integrating the probabilities over z to give I^{bin} and I^{exp}. The relative error ϵ depends on the position i in the chain and is modulated by θ (Figure 3), but it remains small for every locus even in the case of short chains of few elements close to the barrier. It is instructive to realize that the error and the integrated values themselves, is modulated by θ. At the chain terminus only the long half chain will sense the obstruction, whereas for any other locus the overall modulation is a weighted sum of the half-chain modulations. At the central locus, equally long half chains experience the obstruction to the same degree to give modulation by 2θ.

Residual Dipolar Coupling and Chemical Shift Anisotropy. Consider two nuclei A and B in the segment i separated by the internuclear distance r_{AB}. The residual dipolar coupling D_{AB} for this pair of nuclei is also proportional to the product of gyromagnetic ratios $\gamma_A\gamma_B$ and magnetic permeability μ_o, and Planck's constant that are included in a constant denoted as the maximal dipolar coupling D_{AB}^{max}. The vector r_{AB} is at an angle α_{AB} with respect to the internal coordinate system of the segment. As the chain is tumbling, the segment has an

instantaneous angle θ with respect to the magnetic field $\mathbf{B}_o$ (Figure 1). The distribution W governing the averaging of $\cos^2\theta$ is a function of the position of the segment i in the chain, i.e., each segment has its own $\langle P_2\rangle$. Consequently, each segment has its own alignment tensor $\mathbf{A}$ analogous to the molecular alignment tensor of a rigid molecule. In general, $\mathbf{A}$ is not axially symmetric, but in our case rhombicity can be disregarded because the distributions for the random flight chain are axially symmetric. Therefore, D_{AB} is proportional to the axial component $A_{||}$.

$$D_{AB} = \frac{\mu_o \gamma_A \gamma_B \hbar}{4\pi^2 r_{AB}^3} \frac{3\cos^2\alpha_{AB} - 1}{2} \left\langle \frac{3\cos^2\theta - 1}{2} \right\rangle = D_{AB}^{max} \frac{3\cos^2\alpha_{AB} - 1}{2} A_{||} \quad (2)$$

The expression for the frequency shift $\Delta\omega_A$ of the nucleus A with respect to the isotropic Larmor frequency ω_o due to the molecular alignment is analogous to that of D_{AB}:

$$\Delta\omega_A = \Delta\sigma_A \gamma_A B_o \frac{3\cos^2\alpha_A - 1}{2} \left\langle \frac{3\cos^2\theta - 1}{2} \right\rangle = \Delta\omega_A^{max} \frac{3\cos^2\alpha_A - 1}{2} A_{||} \quad (3)$$

where $\Delta\sigma_A$ is the difference between the axial and transverse components of the shielding tensor and α_A is the angle between the shielding tensor's long-axis and $A_{||}$. In practice, the chemical shift anisotropies for carbonyl and aromatic carbons as well as amide nitrogens and phosphorus are large enough to give detectable effects.

The average of $\cos^2\theta$ contained in the expression of D_{AB} and $\Delta\omega_A$ will be obtained by an integration over all angles and space weighted by the distribution and normalized. It is convenient to replace $\cos^2\theta$ by c^2 and the term $\sin\theta\, d\theta$ required for uniform sampling by dc. The chain is confined at all positions along z between two barriers separated by the distance L. Furthermore, the space is confined by the fact that the segment i in any chain cannot approach the barrier closer than $1/2|c|$ without collision.

$$\langle c^2\rangle = \int_{-1}^{1} dc \int_{1/2|c|}^{L} dh \int_{-L+h_z}^{h_z} dz\, W(z,h_z,c,n,N)c^2 \Big/ \int_{-1}^{1} dc \int_{1/2|c|}^{L} dh \int_{-L+h_z}^{h_z} dz\, W(z,h_z,c,n,N) \quad (4)$$

In principle, the space available to a finite chain should be partitioned to a restricted volume where the fully extended chain can reach the barrier to give $\langle c^2\rangle \neq 1/3$ and to a free region where the chain cannot reach the barrier to result in $\langle c^2\rangle = 1/3$.[8] However, when using the decaying exponential functions that in all cases will extend to the barriers with a finite value, it is consistent to extend the integration over the entire available space.

Results

The calculations through eqs 1 and 4 give $\langle P_2\rangle$, which can then be used to calculate D_{AB} and $\Delta\omega_A$ via eqs 2 and 3 for any particular choice of nuclei as a function of the segment position n and chain length N (Figure 4).

The results are easy to rationalize. For longer chains, $\langle P_2\rangle$ will become smaller because the distributions of larger number of uncorrelated chain segments will become progressively more spherical. For an infinitely long chain, $\langle P_2\rangle$ will approach zero. It is important for the studies of proteins that polypeptides of hundreds of residues, which are equivalent to chains containing tens of Gaussian segments, will give an observable fraction of D_{AB}^{max}. For example, for the amide group $D_{NH}^{max} = 21.7$ kHz, resulting in few hertz couplings that are readily observable. Likewise, $\Delta\omega_N$'s for amides with $\Delta\sigma_N = -170$ ppm[17,18] are expected to be obtained quantitatively. The dependence on the segment position is understood similarly. The distribution becomes increasingly spherical as the segment locus approachs the chain termini, i.e., the long half chain begins to dominate the distribution.

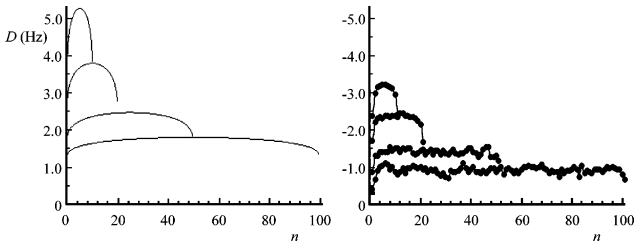

Figure 4. Calculated (left) and simulated (right) RDCs for an amide nuclear pair parallel to the segment as a function of the segment position for random flight chains with $N = 10$, 20, 50, and 100. Bicelle concentration was 5% (w/v).

To validate our calculations, we constructed models of random flight chains from segments of pseudo-atoms to simulate RDCs. Gaussian chains were made simply by a random selection of intersegmental angles. RDCs from large families, tens of thousands conformers, were simulated using the program PALES.[7] The magnitude of the simulated values depends on the nuclear pair in question ($D_{NH}^{max} = 21.7$ kHz, $r = 1.04$ Å), concentration of the bicelles (0.05 w/v), and their degree of alignment (1.0). In the simulations, bicelles had their normals perpendicular to $\mathbf{B}_o$. The simulated and calculated data were comparable and display comparable values and show similar overall dependence of RDC on the chain length and locus position. Small differences at the chain termini may stem from the fact that we used the exponential distribution rather than the binomial distribution. It is also conceivable that the differences could result from finite size effects in the simulations compared with the infinitely thin segments used in the analytical calculation.

Discussion

At first sight it may appear puzzling that a seemingly structureless random flight chain would give nonvanishing RDCs. The paradox is resolved by the realization that the random flight chain has a structure: the structure of *a chain*. The segments are coupled and not free, e.g., as atoms of the ideal gas. Furthermore, it may appear perplexing that the random flight ensemble customarily described by a spherical distribution will give nonvanishing averages of P_2. The dilemma is solved by understanding that RDC or RCSA originates from individual loci along the chain, not from the whole chain averages. For any given locus the distribution is not spherical. In other words, the parameters proportional to $\langle P_2\rangle$ are ensemble and time averages for a particular locus but not averages over all loci, unlike many other observables obtained by other means, e.g., radius of gyration. Finally, it may seem unexpected that a Gaussian chain gives a dispersion of values superficially similar

(17) Tjandra, N.; Szabo, A.; Bax, A. *J. Am. Chem. Soc.* **1996**, *118*, 6986−6991.
(18) Lipsitz, R. S.; Tjandra, N. *J. Magn. Reson.* **2003**, *164*, 171−176.
(19) (a) Clore, G. M.; Gronenborn, A. M.; Tjandra, N. *J. Magn. Reson.* **1998**, *131*, 159−162. (b) Clore, G. M.; Gronenborn, A. M.; Bax, A. *J. Magn. Reson.* **1998**, *133*, 216−221.

to powder patterns measured from rigid macromolecules.[19] For random flight chains the origin of dispersion is the locus-dependent θ averaging, while any particular internuclear direction or shielding anisotropy is collinear with those in other loci, i.e., α is the same. In a well-defined three-dimensional structure, the θ averaging is the same for every locus, that is to say, the internuclear vectors share the same molecular alignment frame and the source of the dispersion is the α sampling of directions by internuclear vectors at the various loci.

RDC data published up to this date are in agreement with our model. The magnitudes of measured couplings are comparable with our predictions and simulations. RDCs should be measured in similar conditions for chains of various lengths, preferably homopolymers, to verify the chain length dependence proposed by our model. The residue position dependence is a small effect close to experimental precision. Even without sequence-specific assignments of measured values to the specific loci, the distribution of D_{AB} can reveal the characteristics of a coil, i.e., the distribution is offset from zero. The position and the width of distributions report of the spatial extent of the chain.

With this understanding, it seems most likely that the nonvanishing RDCs of denatured proteins in recent studies in fact originate from chainlike molecules without a defined three-dimensional spatial organization or nativelike topology, contrary to previous interpretations.[9] Nevertheless, RDCs are informative probes of residual structure above and beyond that of the intrinsic structure of a chain. Once the characteristic random coil values are subtracted from data, the genuine signatures of residual structure become apparent. In this way, RDCs offer unprecedented means to characterize ensembles of weakly structured polypeptides and possibly even folding intermediates.

Acknowledgment. We are grateful to Dr. Gerhard Hummer for informative discussions. This work has been supported by the Academy of Finland, Structural Biology Program.

JA035427V