J|A|C|S
ARTICLES
Published on Web 03/09/2006

Conformational Fluctuations Affect Protein Alignment in Dilute Liquid Crystal Media

Martti Louhivuori,[†] Renee Otten,[‡] Kresten Lindorff-Larsen,[§] and Arto Annila*[,†,||,⊥]

Contribution from the Department of Physical Sciences, Gustaf Hällströmin katu 2, University of Helsinki, Finland, The Groningen Biomolecular Sciences and Biotechnology Institute, Nijenborgh 4, University of Groningen, 9747 AG, Groningen, The Netherlands, Institute of Molecular Biology and Physiology, Universitetsparken 13, University of Copenhagen, Denmark, Institute of Biotechnology, and Department of Biosciences, Viikinkaari 1, University of Helsinki, Finland

Received November 9, 2005; E-mail: arto.annila@helsinki.fi

Abstract: The discovery of dilute liquid crystalline media to align biological macromolecules has opened many new possibilities to study protein and nucleic acid structures by NMR spectroscopy. We inspect the basic alignment phenomenon for an ensemble of protein conformations to deduce relative contributions of each member to the residual dipolar coupling signals. We find that molecular fluctuations can affect the alignment and discover a resulting emphasis of certain conformations. However, the internal fluctuations are largely uncorrelated with those of the alignment, implying that proteins have liquidlike molecular surfaces. Furthermore, we consider the implications of a dynamic bias to structure determination using data from the weak alignment method.

Introduction

Weak alignment of biological macromolecules in dilute aqueous liquid crystalline media[1] has had a tremendous impact on structural biology. Measurements of residual dipolar couplings (RDCs) by solution NMR spectroscopy deliver a wealth of information for protein structure determination, refinement, and cross-validation.[2−4] RDCs report from internuclear vector angles, effectively the directions of chemical bonds, within the biological macromolecule or its complexes. The distance-independent nature of RDC data is a remarkable asset especially for studies of nucleic acids.[5] These extended and proton-deficit molecules are not readily restrained by only short-range distances from nuclear Overhauser enhancements (NOEs) and dihedral angles from scalar couplings to build precise models. Furthermore, protein folds can be recognized using RDCs[6] and rapidly constructed by identifying corresponding structural fragments in the Protein Data Bank.[7] Conformational exchange,[8] domain orientations,[9,10] and flexibility[11,12] can be deduced from RDCs.

The weak alignment NMR spectroscopy also captures molecular dynamics[13−15] as RDCs explicitly probe bond vector fluctuations. Customarily thermal fluctuations of native proteins are thought to be small excursions from the lowest-energy state. These motions include bond librations and side-chain reorientations. Greater digressions involving for example secondary structure fluctuations are also expected to be contained in RDC data.[16−18] Moreover, conformational preferences in unfolded polypeptides have been inspected using the weak alignment method.[19−21] Here we examine how the various conformations contribute to the RDC signal—a signal that is an average over the entire ensemble of protein conformations present in a solution.

[†] Department of Physical Sciences, University of Helsinki.
[‡] The Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen.
[§] Institute of Molecular Biology and Physiology, University of Copenhagen.
[||] Institute of Biotechnology, University of Helsinki.
[⊥] Department of Biosciences, University of Helsinki.

(1) Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111−1114.
(2) Bax, A.; Grishaev, A. *Curr. Opin. Struct. Biol.* **2005**, *15*, 563−570.
(3) Clore, G. M.; Garrett, D. S. *J. Am. Chem. Soc.* **1999**, *121*, 9008−9012.
(4) Prestegard, J. H.; Bougault, C. M.; Kishore, A. I. *Chem. Rev.* **2004**, *104*, 3519−3540.
(5) Zhou, H.; Vermeulen, A.; Jucker, F. M.; Pardi, A. *Biopolymers* **1999**, *52*, 168−180.
(6) Annila, A.; Aitio, H.; Thulin, E.; Drakenberg, T. *J. Biomol. NMR* **1999**, *14*, 223−230.
(7) Kontaxis, G.; Delaglio, F.; Bax, A. *Methods Enzymol.* **2005**, *394*, 42−78.
(8) Pääkkönen, K.; Sorsa, T.; Drakenberg, T.; Pollesello, P.; Tilgmann, C.; Permi, P.; Heikkinen, S.; Kilpeläinen, I.; Annila, A. *Eur. J. Biochem.* **2000**, *267*, 6665−6672.
(9) Fischer, M. W.; Losonczi, J. A.; Weaver, J. L.; Prestegard, J. H. *Biochemistry* **1999**, *38*, 9013−9022.
(10) Bewley, C. A.; Clore, G. M. *J. Am. Chem. Soc.* **2000**, *122*, 6009−6016.
(11) Goto, N. K.; Skrynnikov, N. R.; Dahlquist, F. W.; Kay, L. E. *J. Mol. Biol.* **2001**, *308*, 745−764.
(12) Chou, J. J.; Li, S.; Klee, C. B.; Bax, A. *Nat. Struct. Biol.* **2001**, *8*, 990−997.
(13) Tolman, J. R.; Al-Hashimi, H. M.; Kay, L.; E.; Prestegard, J. H. *J. Am. Chem. Soc.* **2001**, *123*, 1416−1424.
(14) Meiler, J.; Prompers, J. J.; Peti, W.; Griesinger, C.; Brüschweiler, R. *J. Am. Chem. Soc.* **2001**, *123*, 6098−6107.
(15) Fredriksson, K.; Louhivuori, M.; Permi, P.; Annila, A. *J. Am. Chem. Soc.* **2004**, *126*, 12646−12650.
(16) Meiler. J.; Peti, W.; Griesinger, C. *J. Am. Chem. Soc.* **2003**, *125*, 8072−8073.
(17) Bernado, P.; Blackledge, M. *J. Am. Chem. Soc.* **2004**, *126*, 7760−7761.
(18) Bouvignies, G.; Bernado, P.; Meier, S.; Cho, K.; Grzesiek, S.; Bruschweiler, R.; Blackledge, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13885−13890.
(19) Shortle, D.; Ackerman, M. S. *Science* **2001**, *293*, 487−489.
(20) Mohana-Borges, R.; Goto, N. K.; Kroon, G. J.; Dyson, H. J.; Wright, P. E. *J. Mol. Biol.* **2004**, *340*, 1131−1142.
(21) Louhivuori, M.; Pääkkönen, K.; Fredriksson, K.; Permi, P.; Lounila, J.; Annila, A. *J. Am. Chem. Soc.* **2003**, *125*, 15647−15650.

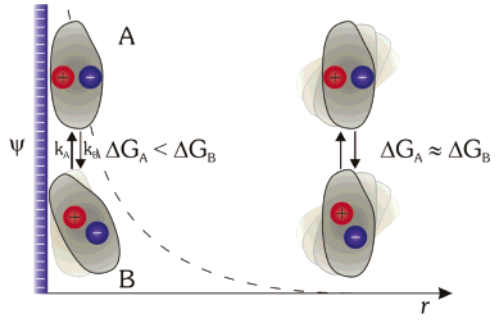

Figure 1. Schematic representation of a dilute liquid crystalline sample. A protein undergoes a two-state conformational isomerism accompanied by small charge displacements, summarized as a fluctuating net electric dipole moment. Far away from the liquid crystalline particle, depicted as a negatively charged wall, the two states *A* and *B* (on right) have similar free energies and are equally populated irrespective of their orientation (shown by shadows). At the surface of the charged obstruction (on left) there is a difference in the free energies because the two conformations cannot, due to steric clashes, reorient to experience the electric potential (ψ, dashed line) in the same way.

At the Surface of a Liquid Crystalline Particle

A biological macromolecule is constantly fluctuating. Thus, at any moment the sample contains an ensemble of conformations that differ from each other, albeit in the case of a folded polymer, on average, only slightly. The consequences of dynamics are fundamental to the interpretation of all NMR data including those obtained from weakly aligned samples. We begin by illustrating a two-state conformational isomerism involving only a charged protein side chain teetering between two rotameric states (Figure 1).

Far away from a charged liquid crystalline (LC) particle, where the protein's electric potential gradient vanishes, the two states have, irrespective of molecular orientation, comparable Gibbs free energies, ΔG_0. Thus, the Boltzmann equilibrium populations P_A and P_B are virtually equal and thus denoted by P_0. We refer to these conditions, present also in the isotropic aqueous solution, as the isotropic equilibrium.

At the surface of the charged liquid crystalline particle conformation A is, due to the favorable orientation of its net electric dipole moment, preferred over conformation B. A reorientation of B would turn the electric dipole in a more attractive orientation but due to steric clashes would also displace B into a weaker potential further away from the nematogen. Thus, the energy of B is inevitably higher than that of A and near the nematogen $P_A > P_B$. Also, even in the absence of Coulombic interactions, the modulation in the van der Waals terms caused by molecular shape fluctuations would cause a minute overall preference for elongated conformations.

In general, for an arbitrary ensemble of conformations the population of a conformation k in the whole sample is governed by $\Delta G_k(\Omega)$

$$P_k = \int \exp[-\Delta G_k(\Omega)]\mathrm{d}\Omega/Q = \int w_k(\Omega)\mathrm{d}\Omega. \qquad (1)$$

The concise notation using w_k, i.e., the probability density, includes the normalization by the partition function Q, i.e., the sum over all conformations in all orientations and positions Ω. Far away from the nematogen, ΔG_k does not depend on Ω. There the isotropic populations that we denote by P_0 exist. In the vicinity of the liquid crystalline particle the various free energy contributions depend on translational and rotational coordinates of the solute−nematogen coupling. The dependence of w_k on Ω varies from one conformation to another; thus, in general, $P_k \neq P_0$. Based on this reasoning there is in the LC-medium a small shift in the total population from the equilibrium population present in the isotropic solution.

The statistical ensemble description given above is equivalent to the portrayal of a conformation over a long trajectory in time. The ergodic correspondence is particularly illuminating when the exchange between conformations is fast, giving only one signal in which the populations are not observable as distinct. Consider a protein diffusing to the obstructing nematogen. The fast flickering motion of a side chain will begin to cease in the vicinity of the charged LC-particle as the energy difference between the states A and B will develop and may eventually become too large for the rotamer to exist in the state B. The translational diffusion and rotational tumbling of the molecule are too slow to allow relaxation by molecular reorientation. In general, interactions between the nematogen and the protein may hamper motions that occur normally when the molecules are not in contact. The differential alignment effects as a whole are not large but worth considering when examined by residual dipolar couplings.

Contributions to the RDC Signal

The presence of a population shift raises the question whether the dynamic bias described above contributes sufficiently to the RDC signal to affect its structural interpretation. At first sight it seems that the tiny variation in the total populations is utterly insignificant. Indeed residual dipolar couplings have been successfully employed in protein structure determination and other analyses, assuming implicitly equal weights for all conformations. However, it is eye-opening to examine the dynamic effects analytically and via simulations.

We follow the nomenclature where the total RDC signal for an internuclear vector is an average over all corresponding directions found in all conformations, denoted by N, in all molecular orientations and positions present in the sample[22,23]

$$\mathrm{RDC} = \frac{1}{N}\sum_{k=1}^{N} \int w_k(\Omega)P_2(c_z^k)\mathrm{d}\Omega$$
$$= \frac{1}{N}\sum_{k=1}^{N}\Big(\sum_{ij=\{x,y,z\}} S_{ij}^k c_i^k c_j^k\Big). \qquad (2)$$

The vector direction, in the form of the second Legendre polynomial P_2, is given by the projection c_z^k along the magnetic field B_0. For each conformation k the integration over orientations and positions using the weight $w_k(\Omega)$ expresses interactions with the nematogen. These differ slightly from one conformation to another. Thus, the individual conformations of each ensemble do not align exactly in the same way. The molecular alignment in the Cartesian frame $\{x,y,z\}$ is expressed by tensor components S_{ij}^k and the internuclear vector by projections $c_i^k c_j^k$. The common index k symbolizes the link between the conformation and its alignment. Solutes that are far away from the nematic wall do not align, that is, w_k is invariant for them. It is only the molecules that sense the Coulombic, van der Waals, or other interactions with the nematogen that will align; that is, w_k depends on Ω.

(22) Fernandes, M. X.; Bernado, P.; Pons, M.; Garcia de la Torre, J. *J. Am. Chem. Soc.* **2001**, *123*, 12037−12047.
(23) Bax, A.; Kontaxis, G.; Tjandra, N. *Methods Enzymol.* **2001**, *339*, 127−174.

Thus, the overall average, i.e. the integration in eq 2 over the sample volume, results in the weak alignment, i.e., $\mathbf{S}^k \neq 0$, and nonzero RDCs.

Although it is generally accepted that RDCs do arise from a conformational ensemble, they are customarily used to refine a single structural model, here denoted by $c_i c_j$, using a common alignment tensor S_{ij} in a set of equations

$$\mathrm{RDC} = \sum_{ij=\{x,y,z\}} S_{ij} c_i c_j. \tag{3}$$

This approach is motivated because often the information content in measured RDCs, just as that in NOEs or scalar couplings, is insufficient to describe an ensemble, i.e., dynamics. This way of looking at eq 3 with no correlation between the alignment and the conformation unlike that in eq 2 causes one to think that some sort of an effective alignment for the whole ensemble would exist. However, it is not obvious why the alignment and coordinates could be considered to be independent from each other, and indeed the simplification from eq 2 to eq 3 warrants further analysis. Even if a lack of correlation is proven by practice, it does not necessary imply that all conformations would contribute equally to the RDC signal.

The correspondence between the population P_k of a conformation and its magnitude of alignment, i.e. the general degree of order[13] S_k, can be derived analytically when a molecule is approximated by an ellipsoid and steric interactions dominate alignment.[22] The relationship is linear when the molecular shape perturbations are small (see Supporting Information). For globular proteins we find

$$S_k \approx {P_k}^r/5P_0 \tag{4}$$

where ${P_k}^r$ is the fraction of all conformations subject to interactions with the nematogen. These molecules are referred to be in the restricted region.[22] In a dilute medium the vast excess of nonaligned conformations P_0 will scale down the alignment to weak and the approximation $P_k \approx P_0$ holds. Thus, the directional information content in RDCs stems from ${P_k}^r$, and the scale is dictated by P_0—de facto by the concentration of the medium. To remove the scaling dependency that may vary from one sample to another, we use a parameter relative to the average to express the contribution of a conformation to the overall RDC signal.

$$W_k = S_k/\langle S_k \rangle \approx {P_k}^r/\langle {P_k}^r \rangle \tag{5}$$

The W_k weight parameter contains no information of the molecules in the isotropic volume fraction V_0. This makes sense as V_0 contains no means to distinguish molecules from one and other, because all the molecules in V_0 taken together will give zero RDCs.

Intuitively, it is obvious that the alignment of a nearly spherical object will be affected by small surface protrusions. Analytical inspection provided in the Supporting Information confirms that small variations in the molecular shape of a small globular protein translate to significant variation in ${P_k}^r$ and thus also in the W_k values. Also the charge distribution within the protein, often summarized as a net charge and an electric dipole moment, will perturb the ${P_k}^r$ values along with the variation in the molecular shape.

Equation 5 provides a relationship between the relative contribution of an individual conformation to the total RDC signal and the general degree of order for that conformation. As the expression was derived by assuming that a conformation can be modeled as an ellipsoid, it is important to demonstrate with more realistic models that RDC contribution of each conformation indeed scales with the amount of alignment, i.e. the normalized general degree of order. To this end we shall show the contributions made to RDCs by various conformations of human ubiquitin and the B domains of protein G—two molecules that have been extensively studied using liquid crystal NMR spectroscopy.[24,25]

To this end we generated conformational ensembles of human ubiquitin, the first and third B domains of protein G, and cardiac troponin C regulatory domain by molecular dynamics (MD) using CHARMM.[26] The simulated structures were: ubiquitin (1D3Z),[27] GB1 (3GB1),[28] GB3 (1P7E),[29] cardiac troponin C in its closed state (1SPY)[30] and open state (1MXL)[31] without the troponin I peptide. Simulations of ubiquitin and GB1 were performed with the CHARMM22 parameter set[32] with a modified backbone potential[33] and generalized Born model.[34] Simulations of troponin C were performed using the EEF1 potential.[35] In the case of troponin C, harmonic restraints (force constant 1 kcal mol^{-1} Å^{-2}) were applied to the heavy atoms in the polypeptide backbone of residues in secondary structural elements to prevent large-scale structural changes that might occur in particular upon removal of the peptide in the open state. An additional harmonic restraint (force constant 10 kcal mol^{-1} Å^{-2}) was added to the Ca^{2+} ion in the calcium-bound open state. The proteins were first heated to 300 K during 200 ps and then equilibrated for 2 ns at this temperature before sampling for 48, 34, and 10 ns for ubiquitin, GB1, and troponin C, respectively.

Subsequently, the PALES program[36,37] was used to predict encounters with neutral bicelles[38] and with negatively charged filamentous phages (Pf1)[39] for each conformation from the molecular dynamics (MD) ensembles. This method for modeling alignment has been very successful in predicting RDCs in a number of cases when van der Waals interactions dictate the alignment. When electrostatics are also considered, the extended PALES method based on Debye−Hückel theory improved the prediction over the plain steric simulation to give satisfactory results, for example for ubiquitin and for the B domains of

(24) Cornilescu, G.; Marquardt, J. L.; Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 6836−6837.
(25) Clore, G. M.; Starich, M. R.; Gronenborn, A. M. *J. Am. Chem. Soc.* **1998**, *120*, 10571−10572.
(26) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(27) Cornilescu, G.; Marquardt, J. L.; Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 6836−6837.
(28) Kuszewski, J.; Gronenborn, A. M.; Clore, G. M. *J. Am. Chem. Soc.* **1999**, *121*, 2337−2338.
(29) Ulmer, T. S.; Ramirez, B. E.; Delaglio, F.; Bax, A. *J. Am. Chem. Soc.* **2003**, *125*, 9179−9191.
(30) Spyracopoulos, L.; Li, M. X.; Sia, S. K.; Gagne, S. M.; Chandra, M.; Solaro, R. J.; Sykes, B. D. *Biochemistry* **1997**, *36*, 12138−12146.
(31) Li, M. X.; Spyracopoulos, L.; Sykes, B. D. *Biochemistry* **1999**, *38*, 8289−8298.
(32) MacKerell, A. D., Jr.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(33) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III. *J. Am. Chem. Soc.* **2004**, *126*, 698−699.
(34) Im, W.; Lee, M. S.; Brooks, C. L., III. *J. Comput. Chem.* **2003**, *24*, 1691−1702.
(35) Lazaridis, T.; Karplus, M. *Proteins* **1999**, *35*, 133−152.
(36) Zweckstetter, M.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 3791−3792.
(37) Zweckstetter, M.; Hummer, G.; Bax, A. *Biophys. J.* **2004**, *86*, 3444−3460.
(38) Bax, A.; Tjandra, N. *J. Biomol. NMR.* **1997**, *10*, 289−292.
(39) Hansen, M. R.; Mueller, L.; Pardi, A. *Nat. Struct. Biol.* **1998**, 1065−1074.

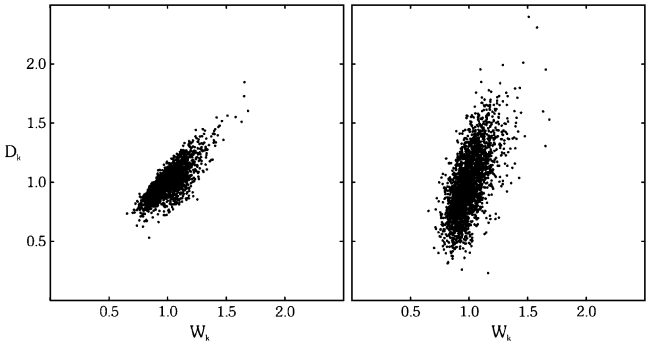

Figure 2. Correlation between the relative magnitude of alignment W_k and conformation's relative contribution D_k to RDCs, averaged over all residues, for the $C^{\alpha}CO$ (right) and HN (left) bond directions ($R = 0.78$ and 0.64, respectively). RDC values in the range from −0.5 to 0.5 Hz were excluded to avoid the larger errors arising from a small denominator in eq 6.

protein G. It is very challenging to predict very precisely RDC values because the values are highly sensitive to the direction of alignment and affected by a number of interactions, various motions, and solution conditions. The simulation method with its approximate expressions of steric and electrostatic forces may fail to reproduce exactly the experimental data, but certainly it describes adequately the overall alignment phenomenon to show the consequences of ensemble averaging.

We calculated each conformation's relative contribution to RDCs averaged over all residues n, i.e.

$$D_k = \frac{1}{n}\sum_{i=1}^{n} \mathrm{RDC}_i^{\,k}/\langle \mathrm{RDC}_i \rangle. \qquad (6)$$

As expected from the analytical model we find that there is a correlation between the alignment of a conformation and its contribution to the average RDC signal (Figure 2). The correspondence is not exact because a RDC value does not *only* depend on the magnitude of alignment that was considered analytically. Also the direction of the internuclear vector, denoted by $c_i^k c_j^k$ in eq 2, varies from one conformation to another. The variation is for example larger for HN than for $C^{\alpha}CO$ bonds because the amides fluctuate intrinsically more than the backbone carbon−carbon bonds. Furthermore, the protein conformations are not ideal ellipsoids, and the direction of alignment, denoted by S_{ij}^k in eq 2, also varies from one conformation to another, leading to additional dispersion in the internuclear vector directions. The distributions of the general degree of order and the RDC contributions obtained as projections of data in Figure 2 show that contributions from various conformations indeed vary. Apparently the variation is a result of many small factors because the distributions are smooth and approximately bell-shaped. To illustrate the causes of variations it is instructive to inspect the subpopulations that depart most from the average.

Inspection of Conformations

On the basis of the relative magnitudes of alignment we can attempt to identify conformations that give the largest and smallest contributions to RDCs. We find distinct subpopulations of two charged residues in the GB1 ensemble whose alignment next to a charged cylindrical nematogen was simulated (Figure 3).

The underlying reason for the dispersion of the relative weights is the variation in $\Delta G_k(\Omega)$ caused by the motions of

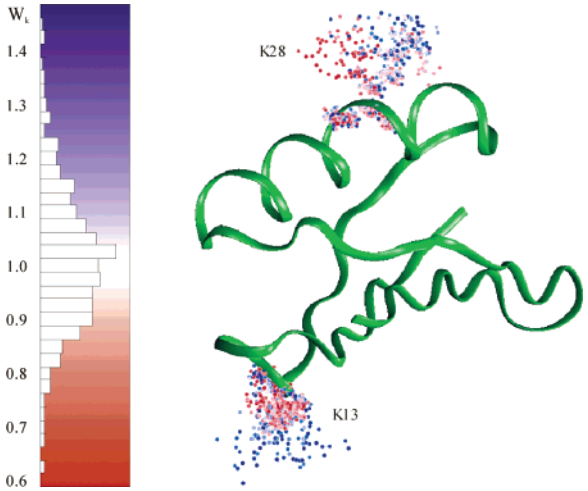

Figure 3. Dispersion of GB1 side chains K13 and K28. Heavy atoms of representative conformations are color coded according to the histogram that gives the relative contribution of a conformation (W_k) to the RDC signal. The conformations shown in blue are emphasized, and those in red are down scaled in the RDC data. The dynamics of all charged side chains cause variation in the net electric dipole moment of the protein that couples with the repulsive electric potential gradient generated by the liquid crystalline particle.

the charged side chains that perturb the net electric dipole moment along with the shape perturbations. The variation in $\Delta G_k(\Omega)$ is smooth overall because there are numerous permutations of distinct side-chain rotamers in GB1, each giving a rise to a slightly different alignment. Certain conformations of the charged side chains give rise to a stronger alignment, and hence these conformations make an above average contribution to the observed RDC signal. Likewise other side-chain conformations lead to a weaker alignment, and hence these conformations make a below average contribution to the observed RDC signal. Neutral side chains are only subject to the short-range van der Waals interactions, and their motions correlate even less with the alignment. These subpopulations are hard to visualize. There are many protruding residues; hence, only few rotamer combinations give the highest and lowest W_k values. Consistently when we simulate how steric interactions give rise to GB1 alignment, we find no apparent relationship between the charged side-chain conformations and the contributions to the RDC signal. We note that the net dipole moment of GB3 is about 2.5 times larger that that of GB1. It is not as susceptible to fluctuations as that of GB1, and therefore for GB3 the relative weight perturbations are smaller. Consistently with our reasoning the prediction of RDCs in Pf1 medium using one model structure is better for GB3 than for GB1.[37]

The ubiquitin C-terminal tail serves to illustrate the dynamic preferences resulting solely from the van der Waals interactions in a neutral nematic medium (Figure 4). The relative contributions of the outstretched tail conformations to the RDC signals are larger than those of the compact ones. The reason for this is that the elongated molecules fit closer to the obstructing liquid crystalline particle than the compact ones. The distribution of weights reflects the conformational probabilities of the tail; for example, fully extended conformations are rare. In other parts of the molecule we observe no noteworthy deviations. The tail seems to be the only mobile moiety that alone can influence the alignment to any noticeable extent.

The visualization of an ensemble using the relative weights serves to illustrate the effects of dynamics but fails to quantize them. When assuming, incorrectly, isotropic populations, i.e.,

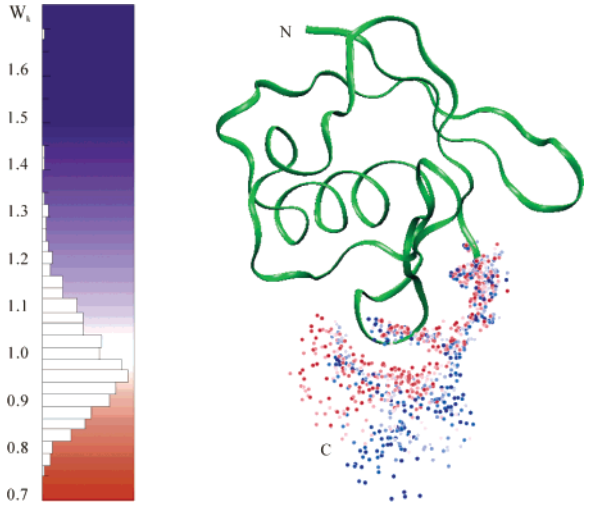

Figure 4. Dispersion of ubiquitin C-terminal tail conformations. Heavy atoms of example conformations are color coded according to the histogram that gives the relative contribution of a conformation (W_k) to the RDC signal. The dispersion in tail residues results in a varying molecular alignment due to the van der Waals coupling with the obstructing nematogen. The elongated conformations (blue) are favored at the expense of the compact ones (red).

equal weights, we do expect errors. However, we find that in practice these errors are smaller than the precision by which the RDCs can be measured experimentally. In particular, in terms of a normalized root-mean-square deviation,[3] the q-value, we find for HN and $C^{\alpha}CO$ backbone RDCs $q < 0.02$ for ubiquitin, GB1, and GB3 when we compare the RDCs with equal weights to those with the predicted steric or electrostatic alignments. Interesting also is that a perfectly refined structure, which is available by substituting the average coordinates $\langle c_i c_j \rangle$ by $c_i c_j$, will give errors $q = 0.07$ for ubiquitin and 0.10 for GB1 using backbone HN and $C^{\alpha}CO$. This means that a structure refinement is ultimately limited by the fact that the data comes from an ensemble whose members align differently from one another. We find no such principle limit if all conformations were to align the same; that is, a perfectly refined structure would reproduce the data.

The reason for very small errors when using isotropic population weights originates from a nearly complete lack of correlation between bond fluctuations and the alignment of protein; effectively making eqs 2 and 3 equivalent. According to Figure 4 we find only a very weak connection between the W_k values and RDCs of ubiquitin C-terminal residues. Similarly in the case of electrostatic alignment of GB1 we see some connection between the W_k values only for certain mobile charged side chains. It would require at least 2 orders of magnitude more samples from the conformational space to quantify the weak correlation between a bond vector and the overall alignment. The reasons for the lack of correlation are inherent to protein properties and will be shortly discussed.

Finally we demonstrate the weak alignment of a protein that is subject to a two-state conformational exchange in addition to the fast low-amplitude internal fluctuations mentioned above. The regulatory domain of cardiac troponin C (cTnC) in its calcium-loaded form is known to sway between the open and closed state.[40] The open state is stable in the complex with troponin I, and the closed state, in the absence of calcium ions.

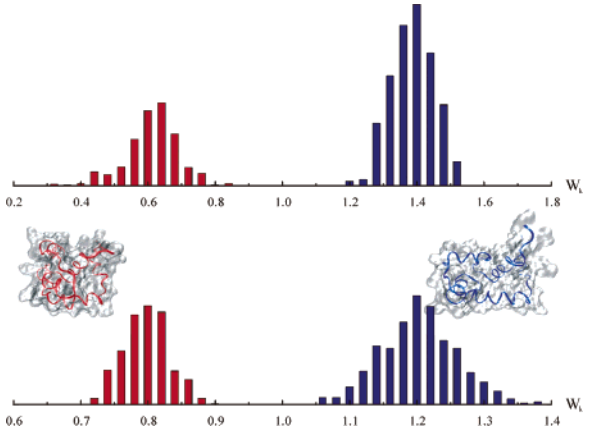

Figure 5. Relative contributions (W_k) of cardiac troponin C open (right) and closed (left) conformations to the RDC signals in the case of steric (above) and electrostatic (below) alignment.

For simplicity we assume the two states to be equally populated and any intermediate state to be transient and negligible.

The overall distribution of relative weights is composed of two subdistributions for the open- and closed-state conformations obtained by MD (Figure 5).

The results from the steric alignment meet the expectations based on the analytical model (see Supporting Information). The elongated open state aligns more strongly than the compact closed state, and its population variation is smaller than that of the globular one. The conformational change from the closed to the open state induced by Ca-ion binding results in substantial change in the charge distribution. This is reflected in the electrostatic alignments. The closed-state conformations align more strongly because their net electric dipole moments are approximately orthogonal to the long axes and thus along the normal of the charged obstruction. Consequently, the variation in RDC contributions is also reduced. The stronger moments of the open-state conformations are approximately parallel to the molecular long axes. Thus, the steric and electrostatic effects compete for the alignment to reduce and to disperse it.

When considering the integrated distributions in the case of steric alignment, the RDCs are biased. The members of the open state carry higher weights than the compact conformations of the closed state. If this effect were neglected in the RDC data analysis, erroneous relative populations of 0.36 and 0.64 would be concluded for the closed and open states, respectively. When electrostatic effects also influence the alignment, the free energies of the two states are more comparable, and the dominance of the open state is less.

In this and similar[19,20,41] cases where the exchange takes place between conformations that differ substantially by their shapes or charge distributions, the variation in W_k is considerable. Its informative value ought to be taken into account when interpreting RDC data as a linear combination of two or more states to elude conformational preferences. This is possible when the corresponding structures are available for alignment simulations.

Discussion

Our results demonstrate that weak alignment measurements of dynamic ensembles may give an emphasis on some conformations relative to others due to differences in their degree of alignment. The fundamental reason for this is that the alignment

(40) Li, M. X.; Wang, X.; Sykes, B. D. *J. Muscle Res. Cell Motil.* **2004**, *25*, 559−579.

(41) Jha, A. K.; Colubri, A.; Freed, K. F.; Sosnick, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13099−13104.

phenomenon features free energy changes just as any other interaction phenomenon. The coupling between the solute and the nematogen is known for ordinary dense liquid crystals,[42,43] and the effects of weak alignment on conformational ensembles have also been addressed in the case of a random coil peptide[15] and a flexible oligosaccharide.[44] Also, in the special case when the alignment in a very dilute medium is brought about by an attraction,[45] the adhered subpopulation is detected, or when a specific binding dominates the alignment, essentially only the bound conformation is detected.[46] Here we have demonstrated that the dilute liquid crystals and axial matrices[47] routinely used in biomolecular NMR spectroscopy also favor certain protein conformations. The variation is of course not as big as in the special cases noted above, but it can still be of importance. This variation in relative weights can also be regarded as dynamic orientation-dependent deformations (ODD). As demonstrated for a well-structured biological macromolecule, the ODD effects are very small if not negligible, for example compared to crystal packing artifacts. We expect that the nematogen is likewise subject to dynamic deformations imposed by the solute. This and other effects, for example plausible adjustments in hydration, are likely to absorb some of the free energy dependence on directional and positional coordinates that were in this study attributed exclusively to the protein to reduce the span of population dispersion.

In the folded regions of a protein, the dynamic emphasis in RDCs is well below the experimental precision of protein structure determination. We were able to identify a couple of charged residues from GB1 and a few from the flexible ubiquitin tail that correlate very weakly with the alignment, but most individual internuclear vectors remain uncorrelated. The sites of largest conformational perturbations localize to the very molecular determinants that are the primary causes for the conformation-dependent alignments. The effects are inherently confined to the dynamic sites of least structural precision typically found at the molecular surfaces. This makes it is very difficult to observe ODDs.

A biological macromolecule houses a number of groups that interact with the LC-particle as practically the whole surface influences the alignment. As a result a single moiety is effectively decoupled from the aligning medium. In other words the correlation between the side-chain orientations and the alignment is very weak. The internuclear vectors examined by RDCs are nearly uncorrelated with the fluctuations of the alignments.[53] This indicates that the protein surface is liquidlike. Consequently, in the RDC signal, that is the average over all conformations, the variation in the aligned populations evens out. This is the reason the q values, for the B domains of protein G and ubiquitin, were much smaller than one might anticipate on the basis of variation in the aligned populations. Therefore, when considering structure determination, the RDC parameter is remarkably robust. This conclusion is similar to that obtained for the pairwise distances that are available from Patterson maps obtained by X-ray crystallography or from NOE spectra measured by NMR spectroscopy.[48] Also for the most flexible moieties RDC data combined from multiple alignment media[49] will largely reconstitute[50−52] the isotropic situation in essence with equal weights. We conclude that in most cases the conformational emphases are inconsequential for the structure determination, refinement, and cross-validation of native proteins.

As has been pointed out, it is more difficult to predict the molecular alignment when also the electrostatic interactions come into the play.[37] It may well be that some of this difficulty stems from the presence of an ensemble of conformations. Albeit being quite similar, the conformations nevertheless align somewhat differently especially when steric and electrostatic interactions compete for the alignment. There may not be a model to display all moieties that affect the alignment in the average manner.

The lack of correlation between an internuclear vector direction and alignment of a conformation explains why, in practice, for folded proteins a common alignment tensor and a refined structure can account for the observed RDCs even if the contributions of conformations vary. The lack of correlation seems to prevent one from detecting the variation of the alignment decisively. However, it should be noted that the common alignment tensor S_{ij} may in some cases differ from the average $\langle S_{ij} \rangle$ over all conformations, and hence, in such cases the representative structure $c_i c_j$ may not contain the average vector directions $\langle c_i c_j \rangle$. This distinction appears indifferent according to eq 3 where the two quantities are in the form of a product to account for the measured RDCs, but may be of consequence when the underlying dispersion of bond vector directions is of interest. The W_k distribution reflects the sampling of alignment directions by conformations. In other words, it is a powder pattern that eludes the components of the fictitious common alignment tensor S_{ij} noted in eq 3 for the whole ensemble. Indeed, we expect that the effects of conformation-dependent alignment will manifest in studies of dynamics.

Even today, many years after its discovery, the weak alignment method continues to be an inspiring phenomenon. When properly analyzed, it is a valuable asset to examine conformational preferences and also to probe dynamics at molecular surfaces, at the very sites where interactions give rise to biological functions.

Acknowledgment. We thank Drs. Ad Bax, Juhani Lounila, and Michele Vendruscolo for valuable discussions, and the Danish Center for Scientific Computing for providing computational resources. This work was supported by the Academy of Finland and the National Technology Agency of Finland.

Supporting Information Available: An analytical model describing a conformation's contribution to the RDC signal; a complete author list for ref 32. This material is available free of charge via the Internet at http://pubs.acs.org.

JA0576334

(42) Emsley, J. W.; Luckhurst, G. R.; Stockley, C. P. *Proc. R. Soc. London A* **1982**, *381*, 117−138.
(43) Lounila, J.; Diehl, P. *J. Magn. Reson.* **1984**, *56*, 254−261.
(44) Berthault, P.; Jeannerat, D.; Camerel, F.; Alvarez Salgado, F.; Boulard, Y.; Gabriel, J. C.; Desvaux, H. *Carbohydr. Res.* **2003**, *12*, 1771−1785.
(45) Ojennus, D. D.; Mitton-Fry, R. M.; Wuttke D. S. *J. Biomol NMR* **1999**, *14*, 175−179.
(46) Shimizu, H.; Donohue-Rolfe A.; Homans, S. W. *J. Am. Chem. Soc.* **1999**, *121*, 5815−5816.
(47) Tycko, R. Blanco, F. J.; Ishii, Y. *J. Am. Chem. Soc.* **2000**, *122*, 9340−9341.
(48) Tang, K. E. S.; Dill, K. A. *Int. J. Quantum Chem.* **1999**, *75*, 147−164.
(49) Ramirez, B. E.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 9106−9107.
(50) Clore, G. M.; Schwieters, C. D. *Biochemistry* **2004**, *43*, 10678−10691.
(51) Clore, G. M.; Schwieters, C. D. *J. Am. Chem. Soc.* **2004**, *126*, 2923−2938.
(52) Clore, G. M.; Schwieters, C. D. *J. Mol. Biol.* **2006**, *355*, 879−886.
(53) X. Salvatella, B. Richter, and M. Vendruscolo have arrived at similar conclusions (unpublished results).