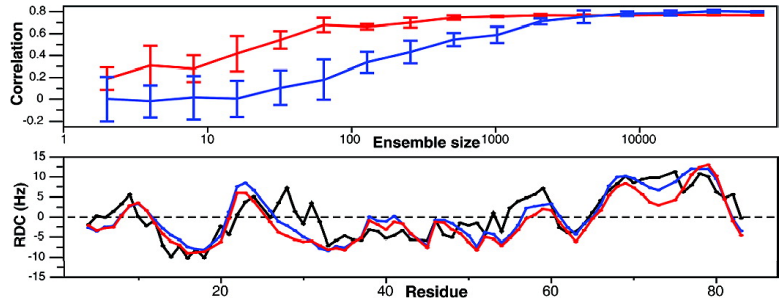
Communication

Calculation of Residual Dipolar Couplings from Disordered State Ensembles Using Local Alignment

Joseph A. Marsh, Jennifer M. R. Baker, Martin Tollinger, and Julie D. Forman-Kay

J. Am. Chem. Soc., **2008**, 130 (25), 7804-7805 • DOI: 10.1021/ja802220c • Publication Date (Web): 31 May 2008



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S COMMUNICATIONS

Published on Web 05/31/2008

Calculation of Residual Dipolar Couplings from Disordered State Ensembles Using Local Alignment

Joseph A. Marsh, Jennifer M. R. Baker, Martin Tollinger,[§] and Julie D. Forman-Kay*

Molecular Structure & Function, Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada, and Department of Biochemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada

Received March 25, 2008; E-mail: forman@sickkids.ca

Residual dipolar couplings (RDCs) in folded proteins are well understood and have been extensively utilized as structural restraints.[1] However, the origin of RDCs in disordered states of proteins and the structural information they provide are still topics of considerable confusion. Although the observation of nonzero RDCs in a highly denatured state was originally thought to provide evidence for long-range native-like structure,[2] it was later shown that RDCs would arise from a random flight chain.[3] Subsequently, RDCs in denatured proteins were found to correlate very well with secondary structure propensities, leading to the proposal that RDCs in a disordered state arise from transient alignment of short stretches of the polypeptide chain.[4] A similar theoretical model used alignment of chain segments to describe random coil RDCs.[5] Two recent studies showed that the RDCs in some disordered proteins could be reasonably predicted using simple statistical coil models and assuming global alignment of each conformer according to shape.[6,7] By comparing simulated and experimental RDCs, it has been possible to probe aspects of local and tertiary structure in disordered states.[8–10] However, these models require averaging over many thousands of conformers, suggesting that the direct use of RDCs as restraints in structural calculations of disordered states could be very computationally intensive. In this report, we describe the use of local alignment tensors for the calculation of RDCs from statistical coil models of disordered proteins that incorporate experimental secondary structure propensities. We show that reasonable agreement with experimental RDCs can be obtained with far fewer structures than when global alignment is used.

In our method, rather than aligning entire structures (i.e., global alignment), we split each structure into a series of short incrementally overlapping fragments. A fragment size of 15 residues was found to be optimal; however, varying it from 10 to 20 residues did not have a major effect. Thus we treat a 100 residue conformer as 86 different fragments which differ by one-residue increments. We assume that the alignment tensor can be approximated by the gyration tensor, similar to previous studies.[6,7] We calculate the gyration tensor for each fragment and calculate RDCs according to equation 1, where A_a and A_r are the axial and rhombic components of the alignment tensor, θ and ϕ describe the orientation of the bond vector in the alignment frame and K is a scaling factor that includes the dependence of experimental RDCs on the degree of alignment. The same K is used for all fragments from all conformers within a given ensemble. Importantly, we ignore the three residues at the N- and C-termini of each fragment as these significantly reduce the agreement with experimental RDCs.

$$\mathrm{RDC} = K\left(\frac{1}{2}A_a(3\cos^2\theta - 1) + \frac{3}{4}A_r\sin^2\theta\cos 2\varphi\right) \qquad (1)$$

[§] Present address: Department of Biomolecular Structural Chemistry, Max F. Perutz Laboratories, A-1030 Vienna, Austria.

Table 1. Correlation of Experimental RDCs with RDCs Calculated Using Global or Local Alignment of 500 000 Structures Generated with and without Secondary Structure Propensity (SSP) Bias

		experimental vs calculated *r*		
protein	SSP	global	local	global vs local *r*
drkN SH3 domain	+	0.64	0.67	0.93
	−	0.22	0.33	0.92
acid-denatured ACBP	+	0.80	0.77	0.98
	−	0.13	0.00	0.91
Staphylococcal nuclease Δ131Δ	+	0.76	0.77	0.99
	−	0.46	0.56	0.95
CFTR R region	+	0.55	0.61	0.91
	−	0.18	0.19	0.86

We applied our method to four disordered proteins of different sizes for which chemical shifts and $^1H-^{15}N$ backbone RDCs are available: the Δ131Δ fragment of staphylococcal nuclease (131 residues)[2], acid-denatured (pH 2.3) ACBP (86 residues),[11,12] the unfolded state of the drkN SH3 domain (59 residues),[13] and the R region of CFTR (191 residues).[14] RDCs from the latter two proteins are from systems currently under study in our laboratory and experimental details of their measurement are described in the Supporting Information. We simulated disordered state ensembles using the program TraDES,[15] both with secondary structure distributions according to their experimental chemical shift-derived secondary structure propensity (SSP) values,[16] and without any secondary structure bias. The resulting ensembles are not biased by any specific tertiary contacts, but are slightly more compact than expected for random coils. We generated many independent ensembles of varying size and calculated RDCs from each ensemble using both our local alignment method and global alignment with the program PALES.[17] The averaged RDCs calculated with both global and local alignment are scaled to maximize the agreement with experimental RDCs.

Table 1 shows the Pearson correlation (r) values between experimental RDCs and RDCs calculated from very large ensembles using both local and global alignment. The inclusion of SSP bias dramatically improves the correlation for all four proteins. Table 1 also gives the correlation between RDCs calculated with global and local alignment for each type of ensemble. The two alignment methods lead to calculated RDCs that are very similar to each other and have similar correlations to experimental RDCs, demonstrating that both local and global alignment methods are similarly effective at predicting RDCs when used with very large ensembles. Plots of simulated and experimental RDCs are given in the Supporting Information.

Jha et al. previously attempted to calculate RDCs for a single protein using a similar local alignment method and found worse agreement with experimental RDCs than when using global alignment.[7] However, they used a much shorter fragment size (7

 10.1021/ja802220c CCC: $40.75

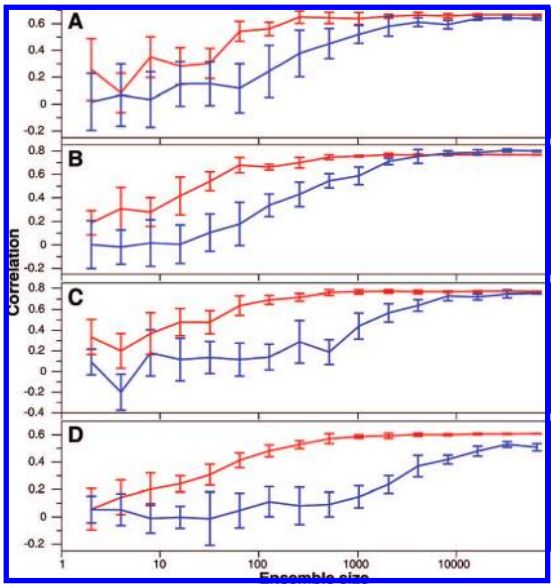


Figure 1. Comparison of RDCs calculated using local and global alignment. Pearson correlation (r) values between experimental and simulated RDCs (global alignment, blue; local alignment, red) from ensembles of different sizes are shown for (A) drkN SH3 domain, (B) acid-denatured ACBP, (C) staphylococcal nuclease Δ131Δ, and (D) CFTR R region. Error bars represent standard deviations from at least 5 replicas.

residues) and did not ignore terminal residues of the fragments. We also observe worse agreement with experimental RDCs using such short fragments (not shown). We speculate that the fragment size should be longer than any secondary structure elements.

Figure 1 shows the correlation between experimental RDCs and RDCs calculated from SSP-biased ensembles of varying size. We see that local alignment shows much stronger correlation with experimental RDCs from smaller ensembles, whereas for larger ensembles the two methods converge upon similar values. This can be explained by fewer conformers being required to approach the average structure of a small fragment than for a full-length protein.

The success of the local alignment method and the improvement of simulated RDCs upon inclusion of secondary structure strongly suggest that RDCs in disordered states of proteins are dominated by local structure, consistent with previous work showing that RDCs correlate well with secondary structure.[4] These results do not necessarily imply, however, that physical alignment only involves short protein segments, since local and global alignment perform similarly with very large ensembles. In the Supporting Information, we show that for disordered chains containing single stable secondary structural elements, the average local alignment tensor of the local structural element is very similar to the global alignment tensor. This demonstrates how global and local alignment can lead to similar RDCs.

It is unclear to what extent tertiary structure affects experimental RDCs in disordered proteins. There is strong evidence for significant tertiary structure in all four proteins used in this study[18–20] (unpublished results for the R region). The fact that we are able to reproduce experimental RDCs fairly well using such simple disordered state models suggests that local structure plays a dominant role in determining RDCs for these systems. However, there are still some significant differences between experimental and simulated RDCs (see Supporting Information) that likely reflect secondary or tertiary structure not present in our simple models. It was previously observed that including a long-range tertiary contact in structural models of α-synuclein significantly improved agreement between predicted and experimental RDCs.[8] It is possible that the tertiary contacts in those simulations led to some changes in local structure, thus inducing an effect on the simulated RDCs. However, it is also possible that the local alignment method would be insensitive to such tertiary structure. It should be possible to clarify this with further experiments and simulations.

The fact that reasonable agreement between predicted and experimental RDCs can be obtained from a fairly small number of conformers bodes well for the use of RDCs as direct restraints in disordered state structural calculations. It is currently feasible to use on the order of 100 simultaneous structures within our ENSEMBLE approach.[20] This number of structures provides fairly good agreement with experimental RDCs as shown in Figure 1. We are now in the process of incorporating RDCs within our ENSEMBLE characterizations of disordered proteins. This should not only improve our calculations and provide information about the relationship between disordered state structure and RDCs, but also increase our understanding of the structural nature of disordered proteins.

Acknowledgment. We thank Lewis Kay for useful discussions, Ranjith Muhandiram for assistance with NMR spectroscopy, and David Shortle for providing Δ131Δ chemical shifts. This work was supported by grants from the Canadian Institutes for Health Research. J.M. and J.B. received awards from the National Sciences and Engineering Research Council and the Canadian Cystic Fibrosis Foundation, respectively.

Supporting Information Available: Experimental and computational details; insight into the effectiveness of local alignment; plots of experimental and calculated RDCs. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Bax, A.; Grishaev, A. *Curr. Opin. Struct. Biol.* **2005**, *15*, 563–70.
(2) Shortle, D.; Ackerman, M. S. *Science* **2001**, *293*, 487–9.
(3) Louhivuori, M.; Paakkonen, K.; Fredriksson, K.; Permi, P.; Lounila, J.; Annila, A. *J. Am. Chem. Soc.* **2003**, *125*, 15647–50.
(4) Mohana-Borges, R.; Goto, N. K.; Kroon, G. J.; Dyson, H. J.; Wright, P. E. *J. Mol. Biol.* **2004**, *340*, 1131–42.
(5) Louhivuori, M.; Fredriksson, K.; Paakkonen, K.; Permi, P.; Annila, A. *J. Biomol. NMR* **2004**, *29*, 517–24.
(6) Bernado, P.; Blanchard, L.; Timmins, P.; Marion, D.; Ruigrok, R. W.; Blackledge, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17002–7.
(7) Jha, A. K.; Colubri, A.; Freed, K. F.; Sosnick, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13099–104.
(8) Bernado, P.; Bertoncini, C. W.; Griesinger, C.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* **2005**, *127*, 17968–9.
(9) Meier, S.; Grzesiek, S.; Blackledge, M. *J. Am. Chem. Soc.* **2007**, *129*, 9799–807.
(10) Mukrasch, M. D.; Markwick, P.; Biernat, J.; Bergen, M.; Bernado, P.; Griesinger, C.; Mandelkow, E.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* **2007**, *129*, 5235–43.
(11) Thomsen, J. K.; Kragelund, B. B.; Teilum, K.; Knudsen, J.; Poulsen, F. M. *J. Mol. Biol.* **2002**, *318*, 805–14.
(12) Fieber, W.; Kristjansdottir, S.; Poulsen, F. M. *J. Mol. Biol.* **2004**, *339*, 1191–9.
(13) Zhang, O.; Forman-Kay, J. D. *Biochemistry* **1995**, *34*, 6784–94.
(14) Baker, J. M.; Hudson, R. P.; Kanelis, V.; Choy, W. Y.; Thibodeau, P. H.; Thomas, P. J.; Forman-Kay, J. D. *Nat. Struct. Mol. Biol.* **2007**, *14*, 738–45.
(15) Feldman, H. J.; Hogue, C. W. *Proteins* **2000**, *39*, 112–31.
(16) Marsh, J. A.; Singh, V. K.; Jia, Z.; Forman-Kay, J. D. *Protein Sci.* **2006**, *15*, 2795–804.
(17) Zweckstetter, M.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 3791–2.
(18) Kristjansdottir, S.; Lindorff-Larsen, K.; Fieber, W.; Dobson, C. M.; Vendruscolo, M.; Poulsen, F. M. *J. Mol. Biol.* **2005**, *347*, 1053–62.
(19) Gillespie, J. R.; Shortle, D. *J. Mol. Biol.* **1997**, *268*, 170–84.
(20) Marsh, J. A.; Neale, C.; Jack, F. E.; Choy, W. Y.; Lee, A. Y.; Crowhurst, K. A.; Forman-Kay, J. D. *J. Mol. Biol.* **2007**, *367*, 1494–510.

JA802220C