JACS ARTICLES

Published on Web 05/25/2010

NMR Characterization of Long-Range Order in Intrinsically Disordered Proteins

Loïc Salmon,[†,§] Gabrielle Nodet,[†,§] Valéry Ozenne,[†] Guowei Yin,[‡]
Malene Ringkjøbing Jensen,[†] Markus Zweckstetter,[‡] and Martin Blackledge*,[†]

Protein Dynamics and Flexibility, Institut de Biologie Structurale Jean-Pierre Ebel, CEA; CNRS; UJF UMR 5075, 41 Rue Jules Horowitz, Grenoble 38027, France, and NMR-Based Structural Biology, Max Planck Institute for Biophysical Chemistry, 37077 Göttingen, Germany

Received February 25, 2010; E-mail: martin.blackledge@ibs.fr

Abstract: Intrinsically disordered proteins (IDPs) are predicted to represent a significant fraction of the human genome, and the development of meaningful molecular descriptions of these proteins remains a key challenge for contemporary structural biology. In order to describe the conformational behavior of IDPs, a molecular representation of the disordered state based on diverse sources of structural data that often exhibit complex and very different averaging behavior is required. In this study, we propose a combination of paramagnetic relaxation enhancements (PREs) and residual dipolar couplings (RDCs) to define both long-range and local structural features of IDPs in solution. We demonstrate that ASTEROIDS, an ensemble selection algorithm, faithfully reproduces intramolecular contacts, even in the presence of highly diffuse, ill-defined target interactions. We also show that explicit modeling of spin-label mobility significantly improves the reproduction of experimental PRE data, even in the case of highly disordered proteins. Prediction of the effects of transient long-range contacts on RDC profiles reveals that weak intramolecular interactions can induce a severe distortion of the profiles that compromises the description of local conformational sampling if it is not correctly taken into account. We have developed a solution to this problem that involves efficiently combining RDC and PRE data to simultaneously determine long-range and local structure in highly flexible proteins. This combined analysis is shown to be essential for the accurate interpretation of experimental data from α-synuclein, an important IDP involved in human neurodegenerative disease, confirming the presence of long-range order between distant regions in the protein.

Introduction

The realization that a large fraction of functional proteins encoded by the human genome are intrinsically disordered or contain long disordered regions has revealed a fundamental limitation of classical structural biology.[1−4] Intrinsically disordered proteins (IDPs) are functional despite their lack of well-defined structure, imposing a new perspective on the relationship between primary protein sequence and function and necessitating the development of an entirely new set of experimental and analytical techniques.[5,6] The importance of developing new methodologies to study these proteins is underlined by the fact that IDPs are associated with many human diseases, including cancer, cardiovascular disease, amyloidosis, neurodegenerative disease, and diabetes.

NMR spectroscopy is exquisitely suited to the study of IDPs,[7] primarily because heteronuclear chemical shift assignment remains possible even for very large disordered proteins.[8] NMR analysis can then be used to precisely study the specific local conformational preferences that encode biological function.[9−11] In spite of their highly dynamic nature, IDPs also exhibit transient or persistent long-range tertiary structure that may be related to biological activity (e.g., via so-called fly-casting interactions[12]) or simply confer protection from proteolysis or amyloidosis. It is precisely the transient nature of such contacts that precludes straightforward NMR detection using standard techniques such as $^1H-^1H$ cross-relaxation. However, long-range information can be measured via the effects of dipolar relaxation between the observed spin and an unpaired electron, which can be artificially introduced into the protein by attaching a nitroxide group to a strategically placed cysteine mutant.[13,14]

[†] Institut de Biologie Structurale Jean-Pierre Ebel.
[‡] Max Planck Institute for Biophysical Chemistry.
[§] These authors contributed equally.

(1) Uversky, V. N. *Protein Sci.* **2002**, *11*, 739–756.
(2) Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; Obradovic, Z. *Biochemistry* **2002**, *41*, 6573–6582.
(3) Tompa, P. *Trends. Biochem. Sci.* **2002**, *27*, 527–533.
(4) Dyson, H. J.; Wright, P. E. *Curr. Opin. Struct. Biol.* **2002**, *12*, 54–60.
(5) Mittag, T.; Forman-Kay, J. D. *Curr. Opin. Struct. Biol.* **2007**, *17*, 3–14.
(6) Eliezer, D. *Curr. Opin. Struct. Biol.* **2009**, *19*, 23–30.
(7) Dyson, H. J.; Wright, P. E. *Chem. Rev.* **2004**, *104*, 3607–3622.
(8) Mukrasch, M. D.; Bibow, S.; Korukottu, J.; Jeganathan, S.; Biernat, J.; Griesinger, C.; Mandelkow, E. M.; Zweckstetter, M. *PLoS Biol.* **2009**, *7*, 399–414.
(9) Meier, S.; Blackledge, M.; Grzesiek, S. *J. Chem. Phys.* **2008**, *128*, 052204.
(10) Wright, P. E.; Dyson, H. J. *Curr. Opin. Struct. Biol.* **2009**, *19*, 31–38.
(11) Jensen, M. R.; Markwick, P.; Griesinger, C.; Zweckstetter, M.; Meier, S.; Grzesiek, S.; Bernado, P.; Blackledge, M. *Structure* **2009**, *17*, 1169–1185.
(12) Shoemaker, B. A.; Portman, J. J.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8868–8873.
(13) Gillespie, J. R.; Shortle, D. *J. Mol. Biol.* **1997**, *268*, 158–169.
(14) Clore, G. M.; Tang, C.; Iwahara, J. *Curr. Opin. Struct. Biol.* **2007**, *17*, 603–616.

10.1021/ja101645g © 2010 American Chemical Society

The gyromagnetic ratio of the electron spin is sufficiently high that the observed line broadening due to the paramagnetic relaxation enhancement (PRE) affords sensitive long-range probes of intra- and intermolecular distances and distance distribution functions. The interpretation of experimental PREs can be relatively straightforward in the case of folded proteins, where an assumption of a static probe localized at a single point in space can be applied to extract approximate distance constraints.[15] It has also been shown that simple modeling of spin-label side-chain mobility in terms of an average over three positions can significantly improve the accuracy of the distance information.[16] Detailed information about transient encounter complexes and their role in protein−protein interactions can also be extracted by combining paramagnetic effects and ensemble-averaged restrained molecular dynamics (MD).[17−19]

In the case of partially folded and unfolded proteins, paramagnetic effects are particularly powerful, as the interactions are sufficiently strong to allow the identification of fluctuating, weakly populated tertiary structural contacts. In this case, the treatment of the intrinsic dynamics of the system is of considerable importance. PREs have thus been interpreted in terms of average distance restraints between the unpaired electron and the observed spin, and these distances have been incorporated directly as constraints into restrained MD or ensemble-averaged restrained MD approaches.[20−27] Explicit relaxation rates can also be incorporated as constraints,[28] and more recently, PREs have been interpreted in terms of probability distributions.[26,29,30] PREs can also be used to select representative ensembles from a large pool of possible conformers.[31−33]

In this study, we have applied to the interpretation of PRE data from disordered proteins a recently introduced approach for modeling highly dynamic and disordered systems that derives explicit molecular ensembles on the basis of experimental data. Ensemble selection is based on the creation of a large number of conformers using an amino acid-specific random coil database known as *flexible-meccano*.[34] *Flexible-meccano* allows for very efficient restraint-free sampling of the available conformational space and was initially demonstrated and refined to provide structural ensembles in agreement with experimentally measured NMR and small-angle X-ray scattering (SAXS) data.[35−42] In parallel, the ensemble selection algorithm ASTEROIDS has been developed to directly determine appropriate regions of conformational space populated by the IDP through selection of conformers from the *flexible-meccano* ensemble using inferential analysis of experimental NMR data.[43] To date, the approach has been applied to experimental measurements that depend essentially on local structural behavior, such as residual dipolar couplings (RDCs) and chemical shifts.[44] Here we have adapted the approach to incorporate the interpretation of PREs. In order to allow for flexibility of the spin label with respect to the backbone conformation, explicit rotameric libraries that have been parametrized against experimental electron spin resonance (ESR) measurements and MD simulations[45] are used to map the allowed position of the electron spin. We then account for the dynamics of the electron spin within this envelope by evoking a model for the autocorrelation function of the relaxation-active interaction that was originally proposed for the interpretation of $^1H-^1H$ cross-relaxation effects.[46] This allows the motion of the relaxation-active dipole−dipole interaction between the electron spin and the observed nucleus to be modeled for each conformer in the ensemble.

The observation that RDCs can be measured in disordered proteins has been followed by the rapid development of techniques for interpreting experimental data in terms of local structure.[38,40,41,47−60] Comparison of experimental data with

[15] Battiste, J. L.; Wagner, G. *Biochemistry* **2000**, *39*, 5355–5365.
[16] Iwahara, J.; Schwieters, C. D.; Clore, G. M. *J. Am. Chem. Soc.* **2004**, *126*, 5879–5896.
[17] Tang, C.; Schwieters, C. D.; Clore, G. M. *Nature* **2007**, *449*, 1078–1082.
[18] Volkov, A. N.; Worrall, J. A.; Holtzmann, E.; Ubbink, M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 18945–18950.
[19] Bashir, Q.; Volkov, A. N.; Ullmann, G. M.; Ubbink, M. *J. Am. Chem. Soc.* **2010**, *132*, 241–247.
[20] Gillespie, J. R.; Shortle, D. *J. Mol. Biol.* **1997**, *268*, 170–184.
[21] Lindorff-Larsen, K.; Kristjansdottir, S.; Teilum, K.; Fieber, W.; Dobson, C. M.; Poulsen, F. M.; Vendruscolo, M. *J. Am. Chem. Soc.* **2004**, *126*, 3291–3299.
[22] Dedmon, M. M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. *J. Am. Chem. Soc.* **2005**, *127*, 476–477.
[23] Bertoncini, C. W.; Jung, Y. S.; Fernandez, C. O.; Hoyer, W.; Griesinger, C.; Jovin, T. M.; Zweckstetter, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 1430–1435.
[24] Kristjansdottir, S.; Lindorff-Larsen, K.; Fieber, W.; Dobson, C. M.; Vendruscolo, M.; Poulsen, F. M. *J. Mol. Biol.* **2005**, *347*, 1053–1062.
[25] Song, J.; Guo, L. W.; Muradov, H.; Artemyev, N. O.; Ruoho, A. E.; Markley, J. L. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 1505–1510.
[26] Allison, J. R.; Varnai, P.; Dobson, C. M.; Vendruscolo, M. *J. Am. Chem. Soc.* **2009**, *131*, 18314–18326.
[27] Ganguly, D.; Chen, J. *J. Mol. Biol.* **2009**, *390*, 467–477.
[28] Huang, J.-R.; Grzesiek, S. *J. Am. Chem. Soc.* **2010**, *132*, 694–705.
[29] Felitsky, D. J.; Lietzow, M. A.; Dyson, H. J.; Wright, P. E. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 6278–6283.
[30] Xue, Y.; Podkorytov, I. S.; Rao, D. K.; Benjamin, N.; Sun, H.; Skrynnikov, N. R. *Protein Sci.* **2009**, *18*, 1401–1424.
[31] 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*, 1494–1510.
[32] Marsh, J. A.; Forman-Kay, J. D. *J. Mol. Biol.* **2009**, *391*, 359–374.
[33] Cho, M. K.; Nodet, G.; Kim, H. Y.; Jensen, M. R.; Bernado, P.; Fernandez, C. O.; Becker, S.; Blackledge, M.; Zweckstetter, M. *Protein Sci.* **2009**, *18*, 1840–1846.
[34] Bernado, P.; Blanchard, L.; Timmins, P.; Marion, D.; Ruigrok, R. W. H.; Blackledge, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17002–17007.
[35] Bernado, P.; Bertoncini, C.; Griesinger, C.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* **2005**, *127*, 17968–17969.
[36] Dames, S. A.; Aregger, R.; Vajpai, N.; Bernado, P.; Blackledge, M.; Grzesiek, S. *J. Am. Chem. Soc.* **2006**, *128*, 13508–13514.
[37] Skora, L.; Cho, M. K.; Kim, H.-Y.; Fernandez, C.; Blackledge, M.; Zweckstetter, M. *Angew. Chem., Int. Ed.* **2006**, *45*, 7012–7015.
[38] Mukrasch, M. D.; Markwick, P. R. L.; Biernat, J.; von Bergen, M.; Bernado, P.; Griesinger, C.; Mandelkow, E.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* **2007**, *129*, 5235–5243.
[39] Meier, S.; Grzesiek, S.; Blackledge, M. *J. Am. Chem. Soc.* **2007**, *129*, 9799–9807.
[40] Jensen, M. R.; Houben, K.; Lescop, E.; Blanchard, L.; Ruigrok, R. W. H.; Blackledge, M. *J. Am. Chem. Soc.* **2008**, *130*, 8055–8061.
[41] Wells, M.; Tidow, H.; Rutherford, T. J.; Markwick, P.; Jensen, M. R.; Mylonas, E.; Svergun, D. I.; Blackledge, M.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 5762–5767.
[42] Bernado, P.; Blackledge, M. *Biophys. J.* **2009**, *97*, 2839–2845.
[43] Nodet, G.; Salmon, L.; Ozenne, V.; Meier, S.; Jensen, M. R.; Blackledge, M. *J. Am. Chem. Soc.* **2009**, *131*, 16968–16975.
[44] Jensen, M. R.; Salmon, L.; Nodet, G.; Blackledge, M. *J. Am. Chem. Soc.* **2010**, *132*, 1270–1272.
[45] Sezer, D.; Freed, J. H.; Roux, B. *J. Phys. Chem. B* **2008**, *112*, 5755–5767.
[46] Brüschweiler, R.; Roux, B.; Blackledge, M.; Griesinger, C.; Karplus, M.; Ernst, R. R. *J. Am. Chem. Soc.* **1992**, *114*, 2289–2302.
[47] Shortle, D.; Ackerman, M. S. *Science* **2001**, *293*, 487–489.
[48] Alexandrescu, A. T.; Kammerer, R. A. *Protein Sci.* **2003**, *12*, 2132–2140.
[49] Mohana-Borges, R.; Goto, N. K.; Kroon, G. J. A.; Dyson, H. J.; Wright, P. E. *J. Mol. Biol.* **2004**, *340*, 1131–1142.
[50] Fieber, W.; Kristjansdottir, S.; Poulsen, F. M. *J. Mol. Biol.* **2004**, *339*, 1191–1199.
[51] Meier, S.; Güthe, S.; Kiefhaber, T.; Grzesiek, S. *J. Mol. Biol.* **2004**, *344*, 1051–1069.

predictions from calculated ensembles of random-coil conformers has indicated that RDCs are sensitive to amino acid-specific backbone dihedral angle distributions. The ability to define random-coil RDC values has led to first the identification and then the quantification of the level of secondary structure propensity in IDPs, initially by comparison with ensemble averages reporting on different sampling regimes[35−42] and more recently by using RDCs to determine conformational sampling on an amino acid-specific basis using ASTEROIDS.[43] In the latter case, a highly efficient local alignment window (LAW) approach to the simulation of RDCs was used to account for local-sampling and near-neighbor effects.[43,59] This demonstrated that in order to correctly define the conformational behavior for a LAW with a length of 15 amino acids, at least 200 structures are needed to average the RDCs.[43] In addition, it was noted that in contrast to chemical shifts and scalar couplings, RDCs are also sensitive to the degree and nature of transient long-range order, and even in the absence of specific contacts, it was found to be necessary to combine the local prediction from the LAWs with a generic baseline profile along the primary sequence that accounts for the chainlike nature of the protein.

In this study, ASTEROIDS and *flexible-meccano* were adapted to allow for transient long-range order and combined with experimental PREs to determine an ensemble description of α-synuclein, a paradigm of the IDP family, whose conformational properties in free solution have been characterized extensively using NMR spectroscopy and associated biophysical techniques.[22,23,26,61−66] We demonstrate that even in the presence of highly diffuse, ill-defined target interactions, explicit modeling of spin-label mobility significantly improves the prediction of experimental data not used in the analysis. We also show that even weak intramolecular interactions can induce a severe distortion of the expected RDC values that compromises the description of local conformational sampling if not correctly taken into account. The expected modulation of the RDCs is parametrized in such a way that it can be analytically introduced into the predicted RDC profile, and we demonstrate that incorporation of long-range contacts from the PRE-derived ensemble significantly improves the prediction of experimental RDCs from α-synuclein.[23] This novel approach allows for the direct and efficient introduction of long-range contacts into ensemble-averaged RDCs and provides for the simple and powerful combination of RDCs and PREs into a single ensemble description.

Theoretical Aspects

Dynamic Averaging of PREs. IDPs are highly flexible on diverse time scales, and this flexibility must be taken into account in the analysis of the measured PREs. The transverse relaxation rate due to the presence of the unpaired electron, Γ_2, can be expressed as follows:[67]

$$\Gamma_2 = \frac{2}{5}\left(\frac{\mu_0}{4\pi}\right)^2 \gamma_{\mathrm{H}}{}^2 g_{\mathrm{e}}{}^2 \mu_{\mathrm{B}}{}^2 s_{\mathrm{e}}(s_{\mathrm{e}}+1)[4J(0)+3J(\omega_{\mathrm{H}})] \qquad (1)$$

where g_e is the electron g-factor, γ_H is the gyromagnetic ratio of the observed nucleus (proton), s_e is the electron spin, ω_H is the proton frequency, μ_B is the Bohr magneton, and μ_0 is the permittivity of free space. It has been shown[14,46] that the spectral density function $J(\omega)$ can be described using a model-free expression of the order parameter comprising the orientational and distance-dependent components of the internal motion, both of which strongly depend on the motion of the spin label with respect to the observed nuclear spin:

$$J(\omega) = \langle r_{\mathrm{H-e}}{}^{-6}\rangle \left[\frac{S^2_{\mathrm{H-e}}\tau_{\mathrm{c}}}{1+\omega^2\tau_{\mathrm{c}}{}^2} + \frac{(1-S^2_{\mathrm{H-e}})\tau_{\mathrm{e}}}{1+\omega^2\tau_{\mathrm{e}}{}^2}\right] \qquad (2)$$

where the order parameter $S^2_{\mathrm{H-e}}$ describes the motion of the dipolar interaction vector, $\tau_c = \tau_r\tau_s/(\tau_r + \tau_s)$ is defined in terms of the electron spin and rotational correlation times τ_s and τ_r, respectively, τ_e is given by the expression $\tau_e = 1/(\tau_i{}^{-1} + \tau_r{}^{-1} + \tau_s{}^{-1})$, in which τ_i represents the effective correlation time of the spin label, and $r_{\mathrm{H-e}}$ is the instantaneous distance between the proton and electron spins. The order parameter can be usefully decomposed into radial and angular components as

$$S^2_{\mathrm{H-e}} = S^2_{\mathrm{ang}}S^2_{\mathrm{rad}} \qquad (3a)$$

where

$$S^2_{\mathrm{rad}} = \langle r_{\mathrm{H-e}}{}^{-6}\rangle^{-1}\langle r_{\mathrm{H-e}}{}^{-3}\rangle^2 \qquad (3b)$$

and

$$S^2_{\mathrm{ang}} = \frac{4\pi}{5}\sum_{m=-2}^{2}|\langle Y_2^m(\Omega^{\mathrm{mol}})\rangle|^2 \qquad (3c)$$

in which Ω^{mol} refers to the orientation of the interaction vector in the frame of the *flexible-meccano* conformer. These expressions are used to calculate the effective transverse relaxation rate for each backbone conformation produced with the *flexible-meccano* algorithm.

The electron spin label is attached to the molecule via a thiol-reactive methanethiosulfonate (MTSL) attached to a cysteine side chain. MTSL conformations are built explicitly for each *flexible-meccano* backbone conformer by randomly sampling known rotameric descriptions.[45] Only conformations that do not result in steric overlap with the remainder of the chain are retained in the *N*-conformer ensemble that is used to represent the position of the side chain. Thus, for each backbone

(52) Ohnishi, S.; Lee, A. L.; Edgell, M. H.; Shortle, D. *Biochemistry* **2004**, *43*, 4064–4070.
(53) Sallum, C. O.; Martel, D. M.; Fournier, R. S.; Matousek, W. M.; Alexandrescu, A. T. *Biochemistry* **2005**, *44*, 6392–6403.
(54) Ding, K.; Louis, J. M.; Gronenborn, A. M. *J. Mol. Biol.* **2004**, *335*, 1299–1307.
(55) Louhivuori, M.; Pääkkönen, K.; Fredriksson, K.; Permi, P.; Lounila, J.; Annila, A. *J. Am. Chem. Soc.* **2003**, *125*, 15647–15650.
(56) Jha, A. K.; Colubri, A.; Freed, K.; Sosnick, T. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13099–13105.
(57) Obolensky, O. I.; Schlepckow, K.; Schwalbe, H.; Solov'yov, A. V. *J. Biomol. NMR* **2007**, *39*, 1–16.
(58) Betancourt, M. R. *J. Phys. Chem. B* **2008**, *112*, 5058–5069.
(59) Marsh, J. A.; Baker, J. M. R.; Tollinger, M.; Forman-Kay, J. D. *J. Am. Chem. Soc.* **2008**, *130*, 7804–7805.
(60) Jensen, M. R.; Blackledge, M. *J. Am. Chem. Soc.* **2008**, *130*, 11266–11267.
(61) Eliezer, D.; Kutluay, E.; Bussell, R., Jr.; Browne, G. *J. Mol. Biol.* **2001**, *307*, 1061–1073.
(62) Fernandez, C. O.; Hoyer, W.; Zweckstetter, M.; Jares-Erijman, E. A.; Subramaniam, V.; Griesinger, C.; Jovin, T. M. *EMBO J.* **2004**, *23*, 2039–2046.
(63) Sung, Y.-h.; Eliezer, D. *J. Mol. Biol.* **2007**, *372*, 689–707.
(64) Wu, K. P.; Kim, S.; Fela, D. A.; Baum, J. *J. Mol. Biol.* **2008**, *378*, 1104–1115.
(65) Li, C.; Lutz, E. A.; Slade, K. M.; Ruf, R. A.; Wang, G. F.; Pielak, G. J. *Biochemistry* **2009**, *48*, 8578–8584.
(66) Lendel, C.; Damberg, P. *J. Biomol. NMR* **2009**, *44*, 35–42.

(67) Solomon, I. *Phys. Rev.* **1955**, *99*, 559–565.

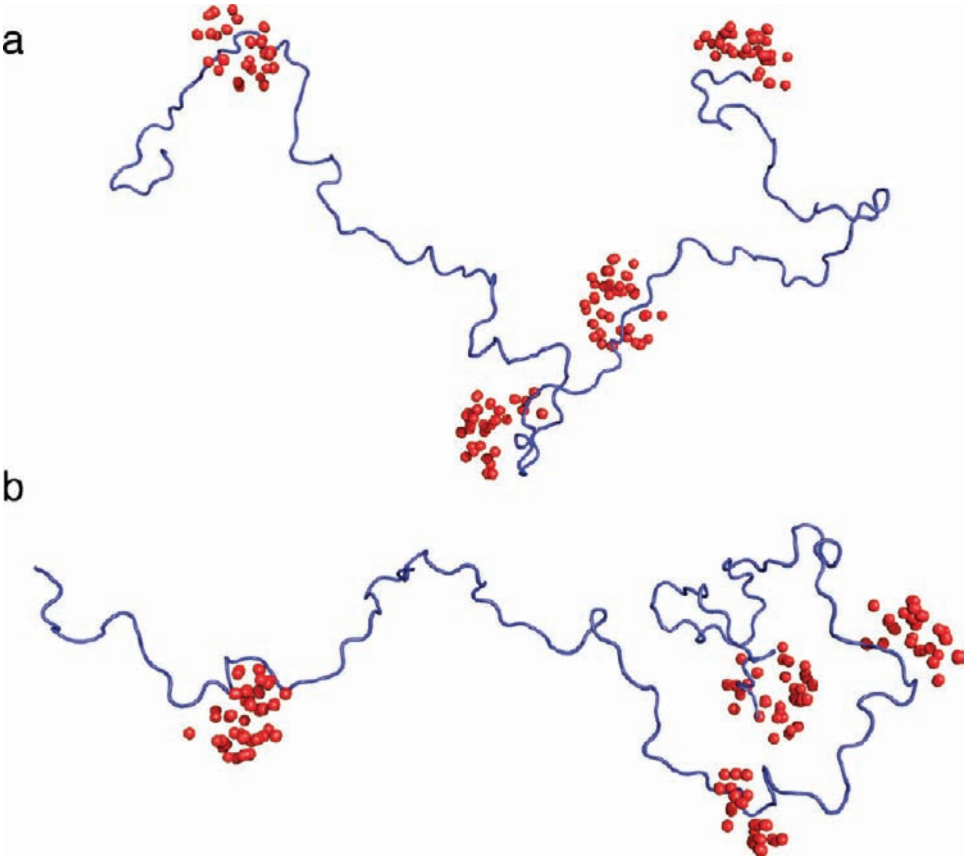

Figure 1. Representation of the possible nitroxide spin label positions relative to the backbone of individual structures calculated using the conformational sampling algorithm *flexible-meccano*. Two representative conformers are shown. The positions of the heavy atoms are represented by the blue ribbon, while allowed MTSL side-chain positions are shown in red for each of four paramagnetic probes used in the α-synuclein study (amino acids 18, 76, 90, and 140). Previously proposed MTSL rotameric libraries[45] were randomly sampled for a total of 600 conformers for each site. Each position was retained and included in the averaging procedure if no steric clashes were found with respect to the given backbone conformation.

conformation, the MTSL side chain is represented by a population-weighted sampling of the available rotameric states. The effective relaxation rate for each amide proton is taken as the average of the rates $\Gamma_{2,c}^{fm}$ for the N retained *flexible-meccano* conformers:

$$\Gamma_2^{\mathrm{total}} = \frac{1}{N}\sum_{c=1}^{N}\Gamma_{2,c}^{fm} \qquad (4)$$

Effective intensities are then calculated as described in Methods.

The assumption made here are that the interconversion between different side-chain conformations is independent of (and faster than) the interconversion between different discrete conformers. In common with previous applications,[23,28] we estimated τ_c to be 5 ns, and the internal motion describing the sampling of the different side-chain conformations was assumed to have a correlation time of 500 ps. This is in broad agreement with values derived from earlier MD/ESR-based studies,[68] and we note that changing the internal correlation time by a factor of 2 in either direction had no noticeable influence on the resulting analysis.

Figure 1 shows the possible positions of the spin label for each of four paramagnetic probes attached to cysteine mutants of the protein α-synuclein in two *flexible-meccano* conformers (amino acids 18, 76, 90, and 140, which are the positions used in the experimental study).[23] The spin label can clearly occupy a large volume space that could potentially affect the effective relaxation behavior of the observed spins.

Results and Discussion

Our aim in this study was to analyze the effects of long-range transient contacts on experimentally observable NMR parameters from unfolded proteins and to develop a formalism that allows their use for the meaningful characterization of both local and long-range structure in these highly flexible systems. In order to do this, we initially used molecular simulations to investigate the expected effects in systems with either one or two dominant long-range contacts. Although these simulated systems were intentionally oversimplified for the sake of clarity, the application of the observed results to more complex networks of long-range transient interactions is expected to be straightforward.

Paramagnetic Relaxation Enhancement in Highly Disordered Systems: Simulation. We initially determined whether it is possible to detect weakly specific long-range interactions via the combined ASTEROIDS and *flexible-meccano* analysis applied to simulated PREs. Figure 2 shows PREs calculated for a simulated model protein of 100 amino acids with paramagnetic spin labels attached at positions 20, 40, 60, and 80 (red bars). In Figure 2a, each conformer contains a contact between 41−50 and 81−90. The definition of a contact is given in Methods. The solid line shows the expected broadening in the absence of *specific* contacts (the reference ensemble where all conformers are allowed). We note that the effective broadening, even in the absence of specific contacts, is quite significant

[68] Sezer, D.; Freed, J. H.; Roux, B. *J. Chem. Phys.* **2008**, *128*, 165106.

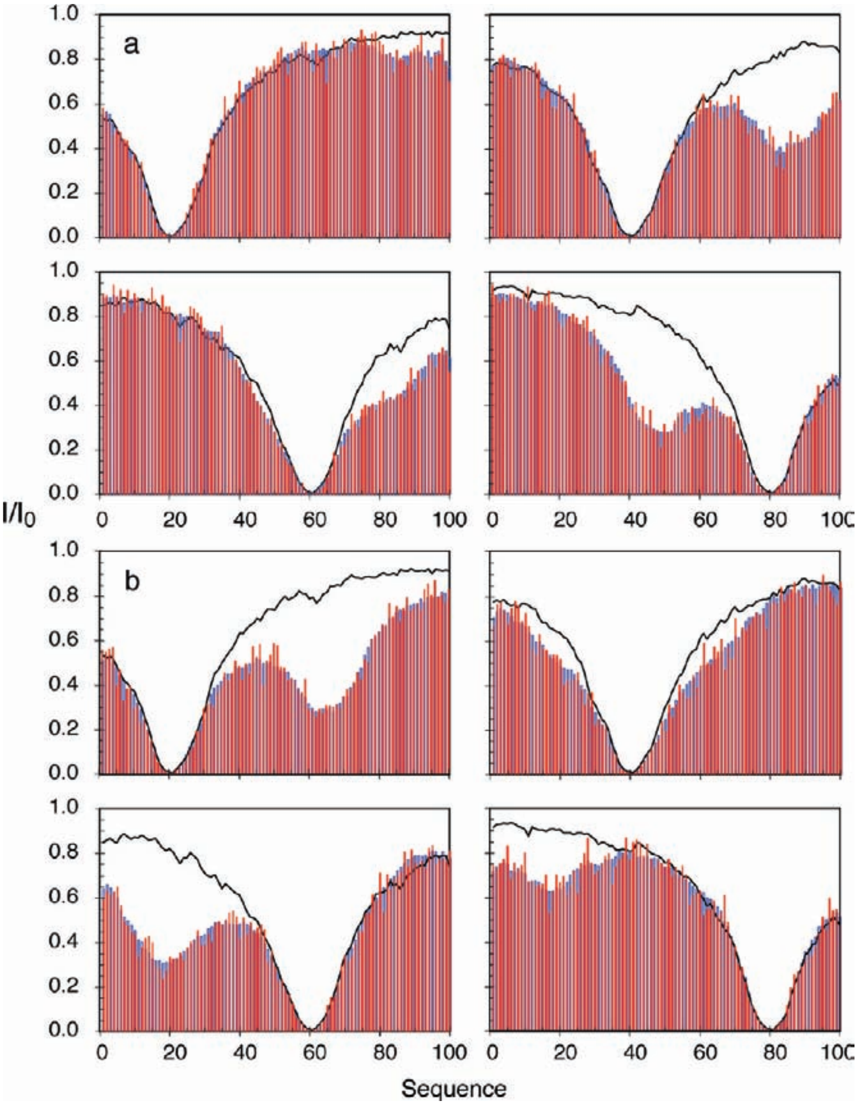

Figure 2. Reproduction of simulated sample PRE data for ensembles containing specific contacts using the ASTEROIDS ensemble selection algorithm.[43] (a) Blue: data averaged over the target ensemble in which each conformer has a contact between 41−50 and 81−90. Red: data averaged over an ensemble of 80 structures selected using ASTEROIDS. The four boxes show the PRE data for simulated spin labels at residues 20 (top left), 40 (top right), 60 (bottom left), and 80 (bottom right). Lines show the PREs calculated from a control ensemble with no specific contacts. (b) Blue: as in (a) for a target ensemble in which each conformer has a contact between 11−20 and 61−70. Red: data averaged over an ensemble of 80 structures selected using ASTEROIDS.

as a result of the large volume space sampled by the spin label. Figure 2b shows a similar representation of an ensemble with contacts between positions 11−20 and 61−70. The ASTEROIDS algorithm targeting these simulated PREs was then used to select 80-member conformational ensembles from a pool of 10 000 structures without specific contacts calculated using the *flexible-meccano* Monte Carlo sampling approach (see Methods). The resulting ensembles reproduced the simulated PREs well, as shown by the blue bars in Figure 2. It should be noted that these simulations used examples that were quite demanding, with 20% of the chain involved in weakly specific contacts. These simulations nevertheless represent a reasonable reproduction of the situation that one may encounter when studying intrinsically disordered or partially folded proteins, with long-range interactions occurring between strands carrying complementary electrostatic charge or containing hydrophobic side-chains. It was therefore of interest to determine whether the broad averaging effects predicted from such a simulation would allow the extraction of meaningful information concerning the long-range contacts.

ASTEROIDS Reproduces the Overall Biophysical Features of the Target Ensemble. Figure 3 shows the effective contacts present in the ASTEROIDS ensembles that matched the simulated data. This representation compares interatomic (C^{α}) distances present in the reference ensemble with those in the selected ensemble (see Methods). The contacts that were used to simulate the data are well identified in both cases. The exact values of the distances were not reproduced (the distances were underestimated), but this is not considered a serious drawback in view of the ill-defined nature of the contact. We also compared the overall distributions of the selected ensembles relative to the reference ensemble. Figure 4 shows that the ASTEROIDS ensemble of structures selected using the simulated PREs from the ensemble containing contacts between regions 11−20 and 61−70 (Figure 2b) reproduced the distribution of the radiii of gyration (R_g) for members of the target ensemble quite closely. The average R_g of the ASTEROIDS ensembles increased slightly with increasing number of structures, from 21.3 Å for the 80-member ensemble to 21.7 Å for the 160-member ensemble, compared with 22.6 Å for the

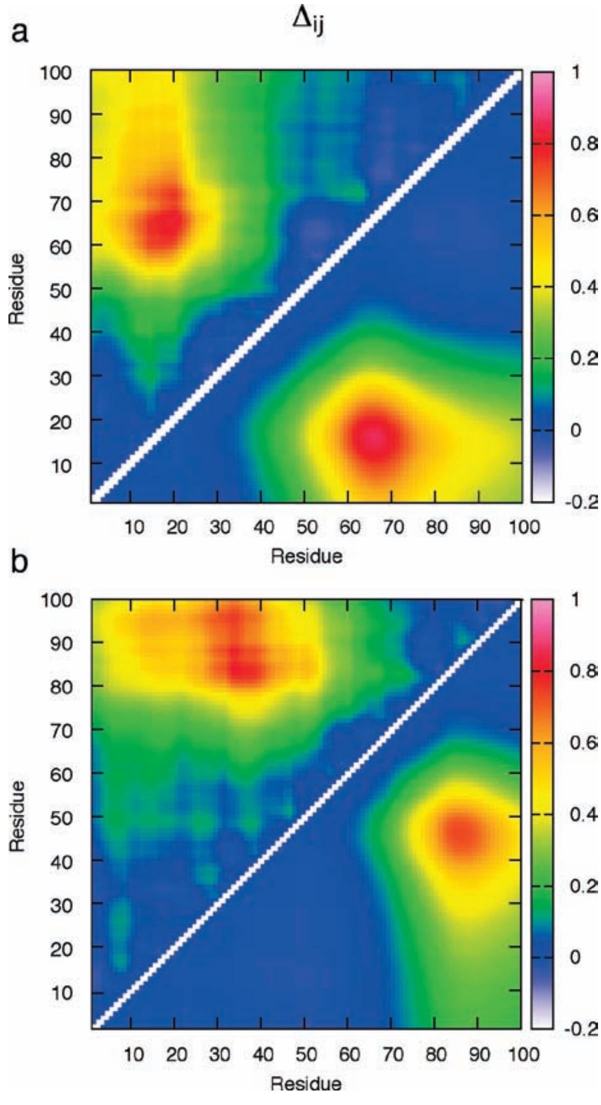

Figure 3. Contact maps showing chain proximity in the ensembles selected using ASTEROIDS on the basis of the data shown in Figure 2 (above the diagonal) in comparison with target ensembles (below the diagonal). In (a), the contact was between 11−20 and 61−70, while for (b), the contact was between 41−50 and 81−90. The scale for the data above the diagonal in each panel has been multiplied by a factor of 0.50 for ease of identification of the contact.

target ensemble. The previously noted tendency of PRE-based analysis to produce unrealistically compact ensembles of unfolded states, although present, was apparently less pronounced using the combined ASTEROIDS and *flexible-meccano* approach than in the case of restrained MD-based studies.[26−28] The exact origin of this observation is not clear and will require further comparative studies, but the improvement may be related to the explicit modeling of side-chain flexibility or to the fact that this approach uses the data to select representative ensembles rather than fitting the conformational sampling directly to the data.

We tested the ability of the combined ASTEROIDS and *flexible-meccano* approach to reproduce more than one contact. Clearly, the accuracy of this reproduction depends strongly on the number of paramagnetic probes and their specific distribution in the protein as well as the nature of the contacts (diffuse or well-defined). We performed an additional simulation, in this case for a protein containing 200 amino acids, where the target ensemble consisted of conformers with a contact between 11−20 and 61−70 or between 141−150 and 181−190. Simulated data from eight paramagnetic probes allowed ASTEROIDS to accurately and unambiguously find both contacts (Figure 5). The simulated target and fitted data from the eight sites are shown in Figure S1 in the Supporting Information.

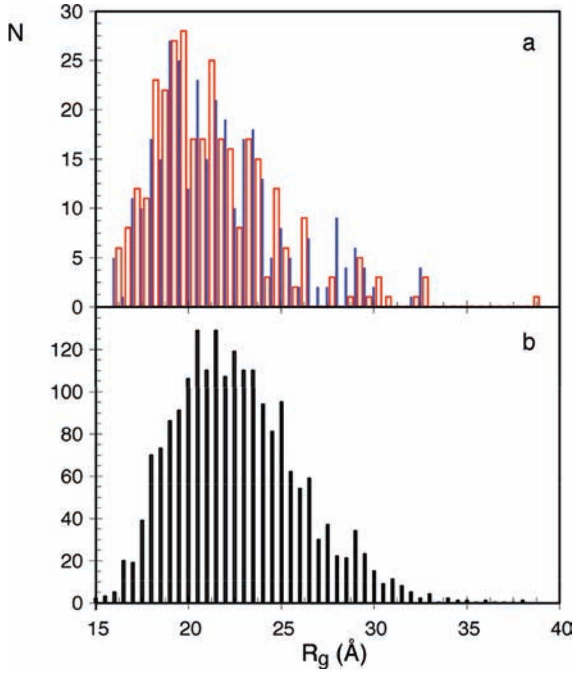

Figure 4. Ability of the ASTEROIDS approach to accurately reproduce the distribution of radii of gyration (R_g) in the selected ensembles. (a) Histogram showing the overall dimensions of the structures in ASTEROIDS ensembles selected on the basis of PREs shown in Figure 2b (contacts between 11−20 and 61−70). Blue: distribution of R_g in ensembles of size 80 (average R_g = 21.3 Å). Red: distribution of radii of gyration in ensembles of size 160 (average R_g = 21.7 Å). (b) Distribution of R_g for a set of 2000 structures from the target ensembles in which all of the structures contain a contact between 11−20 and 61−70 (average R_g = 22.6 Å).

Paramagnetic Relaxation Enhancement in Highly Disordered Systems: Experimental Data. In order to test the ensemble selection procedure further, we applied this approach to an experimental data set measured by Bertoncini et al.[23] for the intrinsically disordered protein α-synuclein. We employed these experimental data to determine how the use of an explicit flexible side-chain description of the spin label compares to using a fixed single position for each *flexible-meccano* conformer. In order to do this, we used ASTEROIDS to select ensembles based on the PRE data from cysteine mutants 18, 90, and 140 and then used these ensembles to predict the PREs measured for the spin label at position 76. It should be noted that this involved removing 25% of the available experimental data. The ensembles determined using a flexible side-chain description and a static side-chain description both fit the experimental data from the three "active" labels to within the experimental uncertainty, with the flexible side-chain model affording a slightly better fit (data not shown). More importantly, the reproduction of the "passive" data (i.e., the data not used in the ensemble selection) was systematically and significantly better when the flexible side-chain model was employed: the root-mean-square deviation (rmsd) for the flexible side-chain model was 0.17 ± 0.01, compared with an rmsd of 0.24 ± 0.02 for the static description. An example is shown in Figure 6, where the data reproductions of the PREs induced by the spin label at position 76 are compared for the two descriptions. This

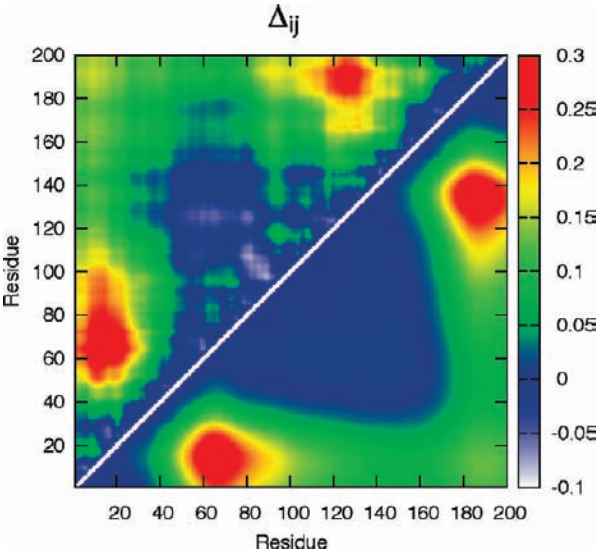

Figure 5. Contact map showing chain proximity in the presence of two contacts. Above the diagonal: contact map for an ASTEROIDS ensemble selected to reproduce simulated PRE data averaged over an ensemble in which each 200 amino acid conformer has a contact between 11−20 and 61−70 or between 141−150 and 181−190. In this case, eight PRE sites were simulated (sequence numbers 22, 44, 66, 88, 110, 132, 154, and 176). Below the diagonal: contact map for the target ensemble used to simulate the PRE data. The scale for the data above the diagonal has been multiplied by a factor of 0.66 for ease of identification of the contact.

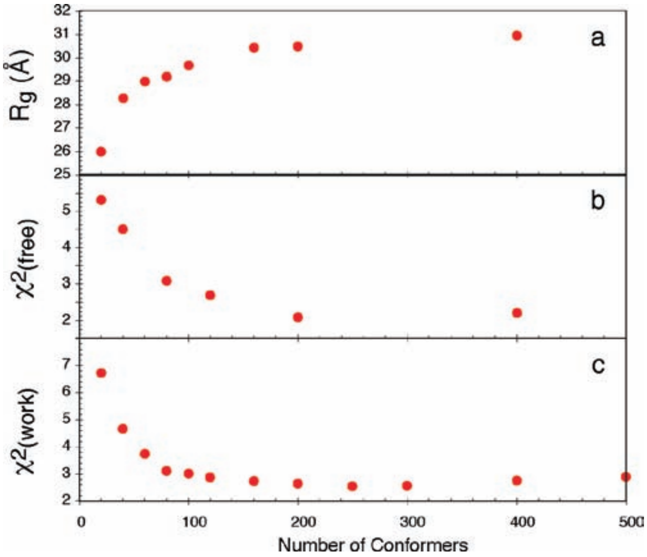

Figure 7. Ensemble characteristics as a function of selected ensemble size, targeting experimental PRE data measured in α-synuclein. (a) Average radius of gyration as a function of the number of structures in the selected ensemble. (b) χ^2 for the passive data as a function of the number of structures in the selected ensemble. The passive data in this case consists of the entire A76C data set. Only data from A18C, A90C, and A140C were used in the ensemble selection for the cross-validated reproduction of the "passive" data set. (c) χ^2 for the active data as a function of the number of structures in the selected ensemble.

example was chosen at random and is representative of the observed improvement. This result demonstrates the importance of incorporating local MTSL side-chain dynamics into the ensemble interpretation of the PREs, even for highly dynamic systems. These motions are predicted to occur on a relaxation-active time scale[45] and therefore require the use of the model-free or equivalent description that can explicitly account for the effect of local motions on the spectral density function. If fast motions of the spin label relative to the backbone are not included in the analysis, time-scale-dependent modulation of the observed relaxation interaction may be aliased into the effective intramolecular distance distribution.

The quality of the cross-validated data reproduction using the dynamic description allowed us to use this approach to probe the optimal number of structures required to describe the ensemble. This number depends on the complexity of the system (including the number of long-range contacts) as well as the number and position of the spin labels, but in this case, both the active and passive χ^2 values indicated that ensembles of ∼200 structures were appropriate (Figure 7). This was supported by analysis of the effective radius of gyration, which rises until it reaches a plateau at approximately the same number of structures. Figure 8 shows the data reproduction when data from all four sites were included in the analysis; also shown is the resulting contact map comparing average interatomic distances in the ensemble with those from a control ensemble in which no selection on the basis of experimental data was made. In line with previous studies, a long-range contact between the C- and N-terminal domains was observed as well as a weaker contact between the so-called NAC region (residues 65−95) and the C-terminal domain.[22,23,35]

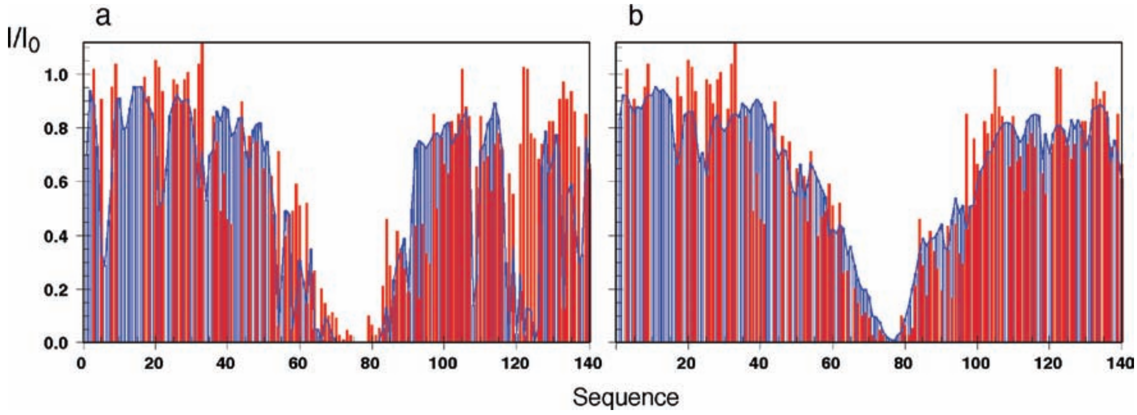

Figure 6. Cross-validation of "passive" α-synuclein PRE data. Only data from A18C, A90C, and A140C were used in the ensemble selection. (a) Example of the reproduction of the PRE data from the A76C site using the static position of the C^β atom as a representation of the average position of the spin label. (b) Example of the reproduction of the PRE data from the A76C site using the explicit MTSL side-chain dynamic averaging model described in the text. In both cases, the experimental PREs are shown in red and the calculated ratios in blue.

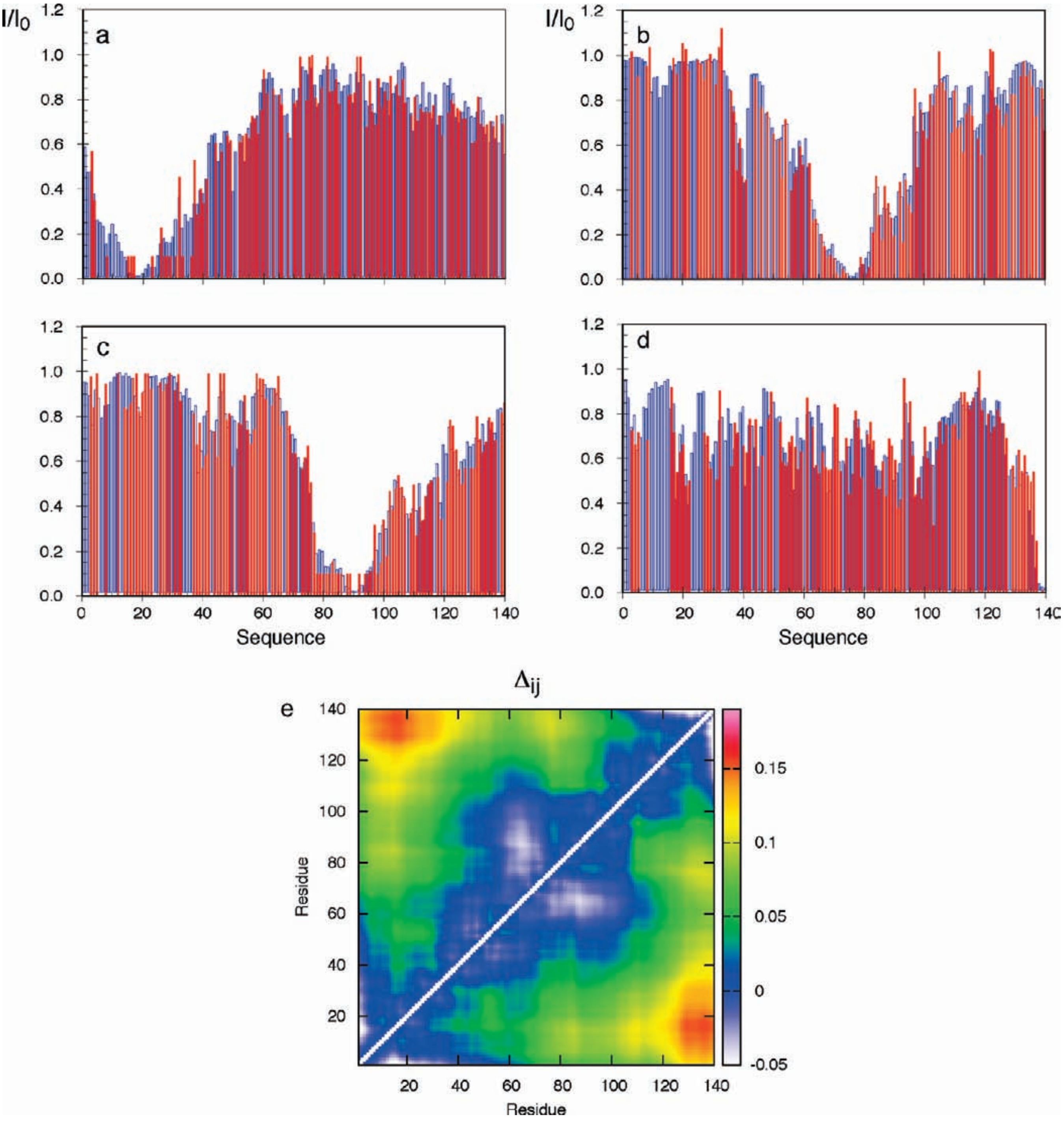

Figure 8. Reproduction of PRE data measured for α-synuclein. (a−d) Comparison of (red) experimental and (blue) ensemble-averaged data for an example calculation. (e) Resulting contact map showing the relative proximity of different parts of the chain.

Effects of Weak Long-Range Contacts on RDCs Measured in Highly Disordered Systems. In order to obtain a unified representation of the behavior of disordered proteins in solution, it is necessary to incorporate data from different sources that exhibit different structural and dynamic dependences. Here we investigate the effects of weak long-range contacts on the expected values of RDCs that are generally assumed to report mainly on local conformational propensities in disordered chains, and we propose appropriate guidelines for combining PREs and RDCs when using ensemble descriptions of flexible proteins.

The *flexible-meccano* approach was used to predict RDCs from 100 000-member ensembles of the 100 amino acid model sequence in the presence of weakly defined long-range contacts (Figure 9). The expected profiles when no specific contacts were present are also shown (Figure 9a). Figure 9b−g shows profiles of the expected ^{15}N−^{1}H^{N} ($^{1}D_{NH}$) and ^{13}C$^{\alpha}$−^{1}H$^{\alpha}$ ($^{1}D_{C\alpha H\alpha}$) RDCs when a contact between two 20 amino acid strands (e.g., regions 1−20 and 81−100) was present. The effect of even such diffuse long-range contacts is surprisingly strong, resulting in significant quenching of the RDC values in regions between the two contact regions and some reinforcement of RDCs in the region of the contacting parts of the chain. Amino acids in all regions had essentially identical conformational sampling in all cases, but the RDCs were very different, indicating very clearly that

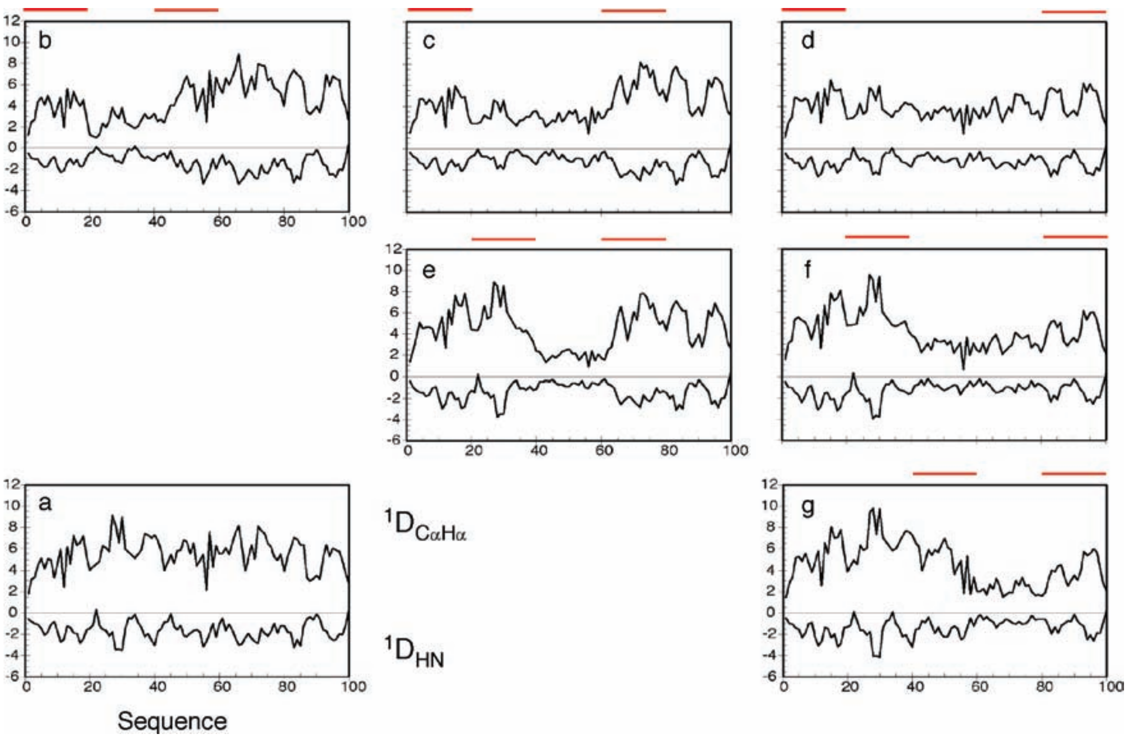

Figure 9. Simulation of $^1D_{NH}$ and $^1D_{C\alpha H\alpha}$ RDC profiles for a disordered protein with an arbitrary sequence in the presence of contacts between different sections of the chain. (a) Profile of couplings in the absence of specific contacts. The program PALES was used to calculate RDCs for each conformer; 100 000 conformers were used in this average and the ones shown in panels (b−g). (b−g) Profiles of couplings in the presence of contacts between regions i and j: (b) $i = 1-20$, $j = 41-60$; (c) $i = 1-20$, $j = 61-80$; (d) $i = 1-20$, $j = 81-100$; (e) $i = 21-40$, $j = 61-80$; (f) $i = 21-40$, $j = 81-100$; (g) $i = 41-60$, $j = 81-100$. The two continuous red bars above each plot indicate the positions of the contacting regions.

caution needs to be exercised when interpreting RDCs uniquely in terms of local structure if long-range contacts are also present. This would potentially lead to significant error in the cases shown in Figure 9.

In order to further clarify the origin of these effects, the same analysis was carried out for a homopolymer (polyvaline), resulting in the expected bell-shaped curve for the ensembles without contact-specific selection (Figure 10a) and clear modifications occurring for the ensembles with specific contacts (Figure 10b−g). The effect of diffuse long-range contacts is apparently to superpose a more complex baseline upon the local structure of the expected RDCs. This baseline has peaks in the interacting regions and a trough in the intervening region. We believe that the effect has a similar origin as that found in the presence of helical elements in disordered chains, where $^1D_{NH}$ values become positive as a result of the effective average alignment of the $^{15}N-^1H^N$ bond vectors with the average chain direction and thereby the magnetic field.[60] The same effect may occur here, although in this case, the helix has a very long period in terms of amino acids and therefore would create a very broad inverted curve relative to the bell-shaped curve, whose shallowness depends on the distance between the interacting segments, as observed from the numerical simulation.

Parametrization of the Effect of Long-Range Contacts on RDCs in Disordered Systems. It has previously been shown that RDCs from unfolded chains with no specific interacting regions can be expressed in terms of the product of a generic baseline, b_{ml}, and RDCs derived from sampling of conformational space that can be defined using short local alignment windows (LAWs):[59,43]

$$D_{ml} = |b_{ml}|D_{ml}^{\mathrm{LAW}} \qquad (5)$$

where m and l represent the pair of nuclei (e.g., N and H^N). In Figure 10, the red curves were obtained using the parametrization of a generic baseline expression that reproduces the numerically predicted baselines shown for the polyvaline chain (see Methods for the full expression). This can be described as a combination of the baseline expression for no specific contacts (a hyperbolic cosine function introduced previously[43]) with Gaussian curves between the contact points. Importantly, the curves depend only on the position of the contacts and the length of the chain.

This expression can then be combined with RDCs predicted using LAWs accounting for short-range conformational behavior. This is illustrated in Figure 11, where the LAW-derived profile (Figure 11b), which was calculated using 200 structures, is combined with the baseline predicted for long-range contacts between segments 41−60 and 81−100 (Figure 11a). The prediction agrees essentially identically with the explicit simulations calculated using 100 000 conformers containing the required contact (Figure 11c). In the case of more than one contact (as shown in Figure 5, for example), the baseline effects are combined as shown in eq 11 and can again be shown to accurately reproduce the effects simulated from explicit averages over 100 000 conformers containing these contacts (see Figure S2 in the Supporting Information).

Simultaneous Analysis of PRE-Derived Long-Range Contacts and RDC-Derived Local Information. The above results show that it is possible in principle to combine PRE-derived long-range information with RDC-derived local information while accounting for possibly significant long-range effects on RDCs and preserving a relatively small number of structures. This latter point is of particular importance when using ensemble selection approaches. In order to test this possibility further, we analyzed

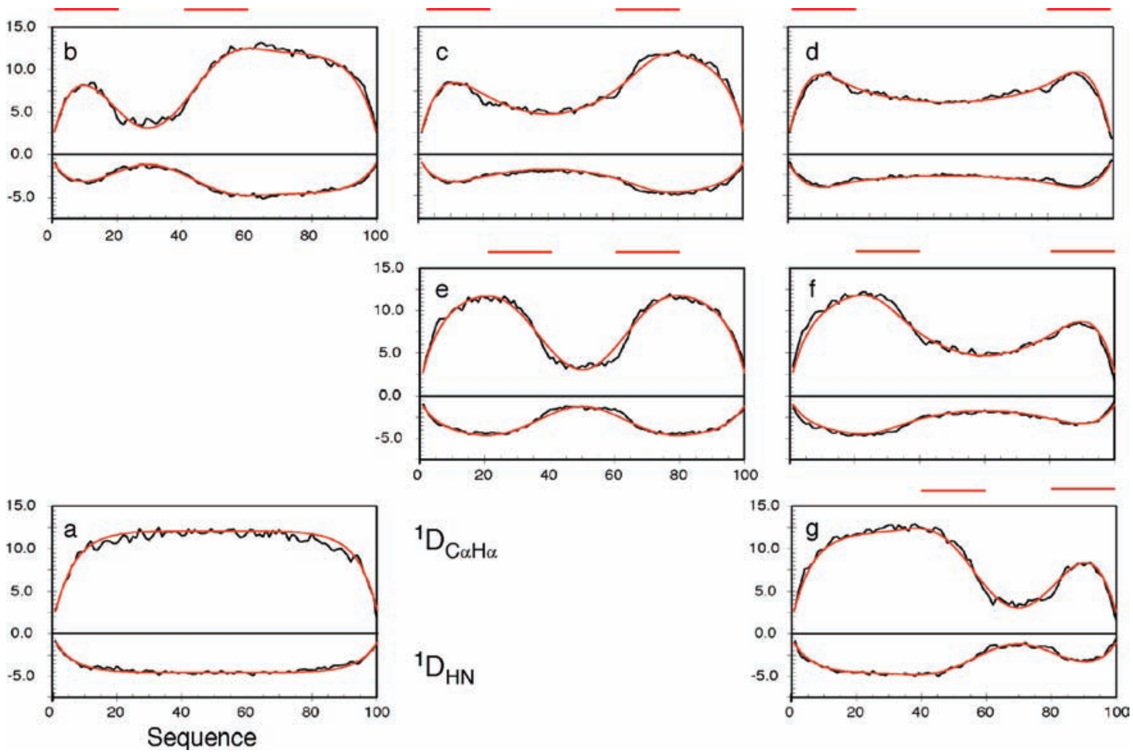

Figure 10. Simulation of RDC profiles for a homopolymer (polyvaline) in the presence of contacts between different sections of the chain. (a) Profile of calculated couplings in the absence of specific contacts. The program PALES was used to calculate RDCs from each conformer; 100 000 conformers were used in this average and the ones shown in panels (b−g). (b−g) Profiles of couplings in the presence of contacts between regions *i* and *j*: (b) $i = 1-20$, $j = 41-60$; (c) $i = 1-20$, $j = 61-80$; (d) $i = 1-20$, $j = 81-100$; (e) $i = 21-40$, $j = 61-80$; (f) $i = 21-40$, $j = 81-100$; (g) $i = 41-60$, $j = 81-100$. The two continuous bars above each plot indicate the positions of the contacting regions. The red curves were computed using eq 11 with the contact positioned in the center of each region.

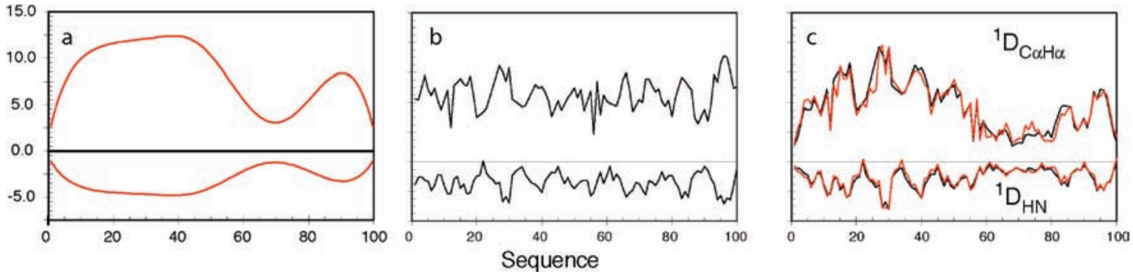

Figure 11. Example of the combination of analytically calculated baselines and RDCs averaged using the local alignment window (LAW) approach. (a) Baseline contribution calculated analytically using eq 11 for contacts between the regions centered on residues 50 and 90. (b) RDCs calculated using the previously proposed LAW approach with windows 15 amino acids in length; each RDC was averaged over 200 structures. (c) Combination of the baseline from (a) and the local RDCs from (b) (red curves) compared to the RDCs averaged over 100 000 full-length conformers in which each structure has a contact between 41−60 and 81−100 (black curves).

the ensembles presented in Figure 3, where the target contacts were between positions 11−20 and 61−70 and between positions 41−50 and 81−90. The contact matrices were analyzed to find the maximum of the difference between the PRE-derived ensemble and the reference ensemble containing no specific contacts (see Methods). The results are shown in Figure 12. In Figure 12a, the red and blue curves indicate the RDC baselines derived using this approach (calculated using eq 11), and the black curve shows the $^1D_{NH}$ RDCs calculated using the LAW approach. In Figure 12b,c, the combination of the baseline and the locally calculated RDCs is compared to RDCs calculated explicitly from 100 000 conformers, all of which fulfill the contact criterion. The good agreement demonstrates that one can combine PREs and RDCs in a meaningful way for the ensemble description of disordered proteins using experimental data.

Combining Experimental PREs and RDCs in α-Synuclein Validates RDC Baseline Analysis. Finally, we applied this analysis to the contact matrix determined on the basis of experimental PRE data from α-synuclein (shown in Figure 8e). Experimentally measured RDCs are shown in Figure 13a and compared to RDCs calculated from an explicit representation of full-length α-synuclein. The RDC baseline derived from analysis of the contact matrix is shown in Figure 13b, superimposed on the RDCs calculated using the LAW approach. The two curves were combined using eq 5, and the result is compared to the experimental data (after appropriate scaling) in Figure 13c. The RDC profile reproduces the experimental data significantly better than the ensemble derived in the absence of specific contacts (rmsd of 0.52 Hz compared with 0.78 Hz). This study therefore not only validates the predicted effects on

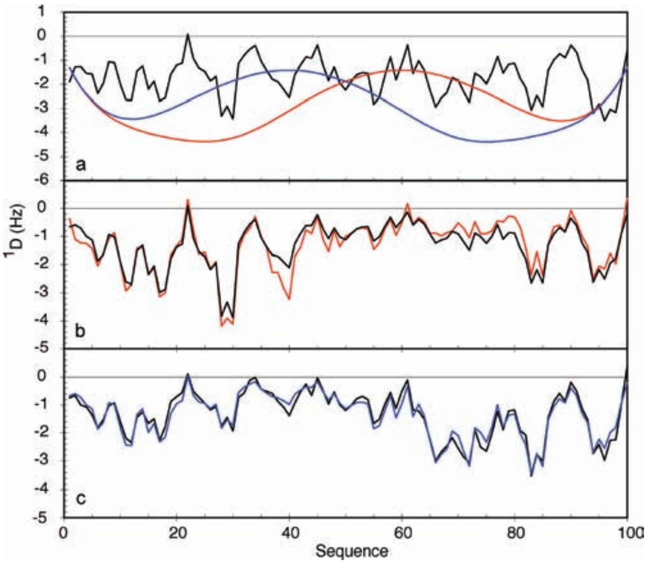

Figure 12. Example of a combined analysis of PREs and RDCs in the context of simulated data. PREs were used to determine long-range contacts. RDC profiles were calculated using baselines determined on the basis of PRE analysis and LAWs. Contacts were identified from distance matrices as described in the text. The reproduction of the PREs and the resulting distance matrix from this simulation are shown in Figures 2 and 3. (a) Black curve: LAW-averaged RDCs. Blue curve: RDC baseline extracted from the contact matrix shown in Figure 3a (contact between 11−20 and 61−70). Red curve: RDC baseline extracted from the contact matrix shown in Figure 3b (contact between 41−50 and 81−90). (b) Black curve: RDCs calculated from an explicit ensemble calculation using 100 000 conformers containing a contact between 41−50 and 81−90. Red curve: the combination of the LAW curve and red baseline curve shown in (a) (contact between regions 41−50 and 81−90). (c) Black curve: RDCs calculated from an explicit ensemble calculation using 100 000 conformers containing a contact between 11−20 and 61−70. Red curve: combination of the LAW curve and blue baseline curve shown in (a) (contact between regions 11−20 and 61−70).

RDC profiles due to long-range transient contacts in unfolded systems but also demonstrates that PREs and RDCs can be usefully combined in an experimental context. This provides further support for previously published observations that RDCs have been correctly reproduced only in the presence of long-range contacts.[35]

Conclusions

In order to understand the conformational behavior of IDPs, a molecular representation of the partially folded state is required. Because of the very large number of degrees of conformational freedom available to such a disordered system, this representation should be based on extensive sets of experimental data. Each experimental parameter is sensitive to different aspects of the structural and dynamic behavior of the disordered state and requires specific consideration of the relevant averaging properties of the physical interaction. In this study, we have taken another step toward the development of a unified molecular representation of the disordered state by combining complementary data sets with novel analytical tools designed to exploit the specific conformational sensitivity of the different experimental parameters.

Having recently demonstrated that multiple RDCs can be combined with an efficient ensemble selection algorithm (ASTEROIDS) to define local conformational sampling directly from the experimental data, we have extended the approach to incorporate the possible presence of long-range contacts. We have demonstrated the use of ASTEROIDS to analyze PREs and faithfully reproduce intramolecular proximity even in the presence of highly diffuse, ill-defined contacts that give rise to broad PRE profiles. We have also demonstrated that the combination of numerical and analytical modeling of spin-label mobility significantly improves the reproduction of the experimental data. The effects of long-range contacts on RDCs have been shown to produce severe distortion of RDC profiles predicted on the basis of local sampling alone. We have demonstrated that this distortion can be generally parametrized and combined with RDC prediction based on local sampling alone to provide an efficient and reliable tool for interpreting RDCs in flexible chains containing preferred long-range contacts.

We thus have shown that it is possible to combine NMR data that exhibit very different averaging properties and structural dependences in a meaningful way, providing the perspective of characterizing the essential local and long-range conformational characteristics of unfolded proteins using PREs and RDCs. In the example we provided, the reproduction of experimental RDCs from the protein α-synuclein was significantly improved when baseline effects derived from the PRE analysis were introduced into the analysis, demonstrating the feasibility of combining these experimental parameters into an informative ensemble description.

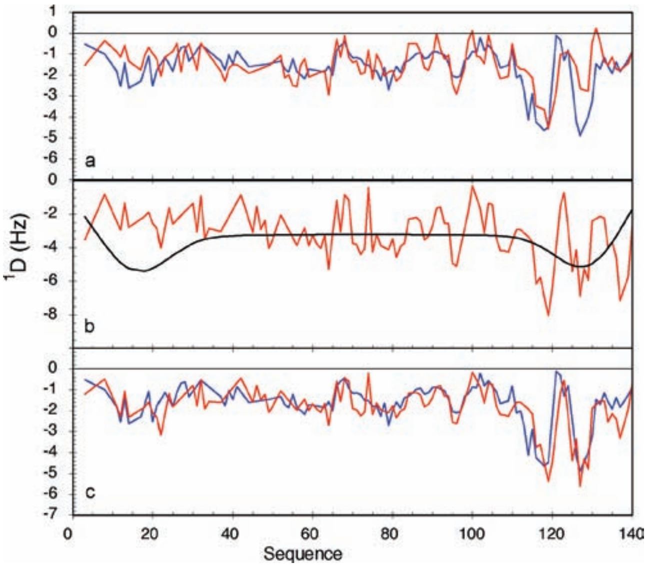

Figure 13. Example of a combined analysis of PREs and RDCs in the context of experimental data: comparison of experimental $^1D_{\mathrm{NH}}$ RDCs measured from α-synuclein aligned in PEG-hexanol with values obtained using the combination of LAW and baseline prediction from PRE analysis. (a) Comparison of experimental $^1D_{\mathrm{NH}}$ RDCs (blue) with couplings calculated using a standard *flexible-meccano* prediction (red). The rmsd between the two distributions was 0.78 Hz. (b) LAW-predicted RDCs (red) and effective baseline derived from the contact map shown in Figure 8e using eq 11 (black). (c) Combination of the curves shown in (b) (red) compared to the experimental $^1D_{\mathrm{NH}}$ RDCs (blue). The rmsd in this case was 0.52 Hz.

Methods

Experimental Data. Details of experimental measurements of RDCs and PREs have been published elsewhere.[23,33]

PRE Calculations with *Flexible-Meccano.* Sterically allowed MTSL side-chain conformations were sampled using previously published rotameric distributions[68] and built explicitly for each spin-label site of each *flexible-meccano* backbone; 600 side-chain conformers were calculated, and the sterically allowed conformers were retained. Relaxation effects were averaged over these conformers as described in Theoretical Aspects.

Definition of Contacts. We considered a contact to be present between two different parts of the polypeptide chain if the C$^\beta$ of an amino acid in one contiguous strand (e.g., residues 11−20) was located less than 15 Å from any C$^\beta$ in another contiguous strand (e.g., residues 51−60).

Contact Matrices. Average distances between sites were represented in terms of the metric Δ_{ij}, defined as

$$\Delta_{ij} = \log(\langle d_{ij}\rangle / \langle d_{ij}^0\rangle) \tag{6}$$

where d_{ij} is the distance between sites i and j in any given structure of the ASTEROIDS ensemble and d_{ij}^0 is the distance between sites i and j in any given structure of the reference ensemble (with no specific selection). This metric was used to highlight a higher propensity to form contacts than in a molecule that has no specific contacts. It should be noted that this representation of average interatomic distances naturally (and artificially) enhances contacts that are further apart in the chain, so the observed contacts are "smeared" away from the diagonal.

Contact matrices were analyzed to determine $l_{ij}^{\max}$, the maximum of the difference between the PRE-derived ensemble and the reference ensemble containing no specific contacts:

$$l_{ij}^{\max} = \max_{i,j\in[1,n]}(\Delta_{ij}) \tag{7}$$

The matrix was divided into segments of 5 × 5 amino acids and searched for the highest-populated segment fulfilling the following criterion:

$$0.9 l_{ij}^{\max} \leq \Delta_{ij} \leq l_{ij}^{\max} \tag{8}$$

This approach identified the highest-populated contacting region. The center of this region was then used to calculate the baseline effects on the RDC profile using eq 11.

RDC Calculations with *Flexible-Meccano* Using a Global Alignment Tensor. Simulated RDCs were calculated using the program *flexible-meccano* interfaced to PALES.[69] Profiles of RDCs in the presence of long-range order were simulated by retaining only conformers for which the desired contact was present.

RDC Calculations with *Flexible-Meccano* Using a Local Alignment Window. For calculations using a LAW, the RDC for the central amino acid of the local 15 amino acid segment was calculated for each individual structure.[43] For the terminal amino acids, seven alanines were added to the N- or C-terminus during the building of the protein to ensure that a 15 amino acid segment was always present. The resulting RDC profile along the primary sequence was calculated by averaging each value over the whole ensemble and multiplying by the corresponding scaled absolute value of the effective baseline given in eq 11. RDCs calculated using full-length descriptions of the protein were averaged over all conformers as previously described.[34]

ASTEROIDS Ensemble Selection. ASTEROIDS uses a previously described genetic algorithm to build a representative ensemble of structures of fixed size N from a large database. The algorithm selects an ensemble of N structures by comparing with experimental data using the following fitness function:

$$\chi^2_{\mathrm{ASTEROIDS}} = \sum_k (\Delta I^k_{\mathrm{calcd}} - \Delta I^k_{\mathrm{exptl}})^2 \tag{9}$$

where

$$\Delta I^k_{\mathrm{calcd}} = \frac{\Gamma^k_{2,\mathrm{red}} \exp(-\Gamma^k_{2,\mathrm{para}} t_{\mathrm{m}})}{\Gamma^k_{2,\mathrm{red}} + \Gamma^k_{2,\mathrm{para}}} \tag{10}$$

in which $\Gamma_{2,\mathrm{para}}$ is the paramagnetic component of the measured relaxation rate given in eq 4, $\Gamma_{2,\mathrm{red}}$ is the intrinsic transverse relaxation rate of the observed proton spin, and t_{m} is the mixing time, for which a value of 10 ms was used. The final ensemble is obtained from generations of ensembles that undergo evolution and selection using this fitness function. Each generation comprises 100 different ensembles of size N. Remaining parameters are treated as previously described.

Parametrization of a Generic RDC Baseline Expression for Transiently Contacting Chains. A generic RDC baseline expression for transiently contacting chains can be obtained by combining the baseline expression for no specific contacts (a hyperbolic cosine function introduced previously[43]) with a Gaussian curve between the contact points and then correcting this with Gaussian curves in the vicinity of the contacting points. Importantly, the Gaussian curves depend only on the position of the contacts and the length of the chain. This results in the following analytical expression for the baseline RDC, D_{ij}^{BL}:

$$\begin{aligned} D_{ij}^{\mathrm{BL}} = \{2b(L)\cosh[-a(L)(m - m_0)] - \\ c(L)\}\Big(1 - \sum_i \big\{G_i \mathrm{e}^{-(\mathrm{m}-n_{0,i})^2/2\sigma_i^2} + \\ H_i\big[(D_i + S_i)\mathrm{e}^{-(\mathrm{m}-n_{1,i}+\mathrm{S}_i/2)^2/2\delta^2} + \\ (D_i - S_i)\mathrm{e}^{-(\mathrm{m}-n_{2,i}+\mathrm{S}_i/2)^2/2\delta^2}\big]\big\}\Big) \end{aligned} \tag{11}$$

where L is the length of the chain, the contact occurs between positions n_1 and n_2, and the sum includes all of the independent contacts i. Other parameters are defined as follows: $m_0 = (L + 1)/2$, $n_0 = (n_1 + n_2)/2$, $D = |n_1 - n_2|$, and $S = n_0 - m_0$. The parametrizations of a, b, c, G, H, σ, and δ are given in the Supporting Information.

Acknowledgment. L.S. received a grant from the French Ministry of Education. This work was supported by the French Research Ministry through ANR Protein Motion PCV07_194985 PCVI 0013 and the Deutsche Forschungsgemeinschaft Heisenberg Scholarship Z.W. 71/2-1 and 3-1 (to M.Z). M.R.J. benefited from a long-term EMBO fellowship and Lundbeckfonden support.

Supporting Information Available: Figure S1 showing a reproduction of simulated sample PRE data for ensembles containing two specific contacts (produced using the ensemble selection algorithm ASTEROIDS) and the associated baseline effects; Figure S2 showing a comparison between RDCs calculated by ensemble averaging and the baseline contribution calculated using eq 11; Figure S3 showing RDCs measured in A76C cysteine mutant and wild-type α-synuclein; Figure S4 showing calculated and experimental ^3J scalar couplings from α-synuclein; and parametrization of a generic RDC baseline expression. This material is available free of charge via the Internet at http://pubs.acs.org.

JA101645G

[69] Zweckstetter, M.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 3791.