J|A|C|S COMMUNICATIONS

Published on Web 01/11/2010

Defining Conformational Ensembles of Intrinsically Disordered and Partially Folded Proteins Directly from Chemical Shifts

Malene Ringkjøbing Jensen,* Loïc Salmon, Gabrielle Nodet, 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

Received November 25, 2009; E-mail: malene.ringkjobing-jensen@ibs.fr; martin.blackledge@ibs.fr

A significant fraction of proteins - over 40% of the human proteome - are not folded, or are only partially folded in their functional form.[1] These intrinsically disordered proteins (IDPs) are strongly implicated in important human pathologies such as cancer and neurodegenerative disease but fall beyond the reach of the tools developed for classical structural biology due to their extreme structural flexibility.[2,3] Many IDPs undergo disorder-to-order transitions upon interaction with physiological partners, where molecular recognition is accompanied by local folding upon binding.[4,5] However the relationship between intrinsic conformational propensity and the structure adopted by the protein in its bound form remains poorly understood. For these reasons the development of meaningful descriptions of the conformational behavior of IDPs, and their relationship to protein function and malfunction, represents a key challenge for contemporary structural biology.

Nuclear magnetic resonance (NMR) spectroscopy reports on structural propensities at atomic resolution, on time scales varying over many orders of magnitude, and is therefore probably the most powerful biophysical tool for studying IDPs.[6] NMR inherently provides time- and ensemble-averaged structurally dependent experimental measurements and, as such, is exquisitely suited to the study of conformationally heterogeneous and flexible systems.[7] The dynamic averaging properties of NMR observables are well understood, rendering their exploitation particularly appropriate for the development of atomic resolution ensemble descriptions of flexible or unfolded proteins.[8−12]

Chemical shifts measured in IDPs report on the population weighted average over an entire ensemble of interchanging conformers, exchanging on time scales faster than the millisecond range. These readily measured parameters are nevertheless highly sensitive probes of the local protein conformation,[13−15] as has been demonstrated by the recent determination of three-dimensional structures of entire globular proteins using chemical shifts as sole experimental constraints.[16,17] The dependences of $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts on backbone ϕ/ψ dihedral angles have been routinely used to identify the position of secondary structure and to estimate the level of secondary structural propensity within folded and unfolded proteins.[18−26] In this study we combine ensemble descriptions of unfolded proteins,[27] with a state-of-the-art chemical shift prediction algorithm that has underpinned the successful determination of folded proteins from chemical shifts.[28] This powerful combination is used to explore the possibility of using chemical shifts alone to map the local backbone conformational sampling of intrinsically disordered and partially folded proteins.

$^{13}C\alpha$, $^{13}C\beta$, $^{13}C'$, and ^{15}N chemical shifts exhibit different dependences on the backbone ϕ/ψ dihedral angles and are therefore sensitive probes of conformational sampling in disordered proteins.[29,30] These sensitivities are complementary in terms of the mapping of different regions of Ramachandran space, suggesting that their combination may allow the resolution of site-specific backbone conformational behavior. $^{13}C\alpha$ and $^{13}C\beta$ secondary shifts report essentially on the Ramachandran space sampled by the observed amino acid, while both $^{13}C'$ and ^{15}N are also sensitive to the sampling properties of the neighboring amino acids. To exploit this complementarity we employ an explicit ensemble description of unfolded proteins (*Flexible-Meccano*) that has been used in combination with residual dipolar coupling (RDCs),[28] scalar couplings,[31] and small angle scattering data[12,28] to describe conformational sampling in IDPs and chemically denatured proteins.

An efficient selection algorithm (ASTEROIDS)[32] is used to assemble a 200-strong subensemble of structures out of a much larger pool, which is in agreement with the experimental $^{13}C\alpha$, $^{13}C\beta$, $^{13}C'$, and ^{15}N chemical shifts. Selection starts from a large pool of conformers (typically 10 000 structures) constructed by *Flexible-Meccano* using standard random coil ϕ/ψ backbone dihedral angles. The program SPARTA is used to calculate chemical shifts for each member of the ensemble. The selection procedure involves two steps: an iteration step where each residue is treated independently, and a final step where full structures are selected. The first iteration step consists of the selection of 200 ϕ/ψ values for each residue that are in agreement with the $^{13}C\alpha$, $^{13}C\beta$, and $^{13}C'$ chemical shifts. This step is repeated five times to obtain 1000 ϕ/ψ values for each residue. A new ensemble of structures is created using *Flexible-Meccano*, but this time using the selected 1000 ϕ/ψ values for each residue. ASTEROIDS is applied again for each residue (see Supporting Information) independently to select 5×200 ϕ/ψ values from the new pool of structures. This iterative procedure is repeated until no further improvement in the fitting of the chemical shifts of the individual residues can be obtained. Step two of the selection procedure is then applied using $^{13}C\alpha$, $^{13}C\beta$, $^{13}C'$, and ^{15}N chemical shifts, where entire structures (200 conformers) are selected from the pool of structures generated during the previous iterations. While chemical shifts are expected to report only on local conformations, other experimental data such as residual dipolar couplings (RDCs) and paramagnetic relaxation enhancements (PREs) or SAXS report on long-range order. Therefore, the selection of entire structures will allow combined fitting of several types of experimental data.

No assumptions are made in terms of the secondary structural propensity, as the first ensemble contains only unfolded structures derived from the statistical coil database. Local conformational bias is recognized on the basis of chemical shift, and resulting propensities are used to assemble the new database for the subsequent iteration. In this way the algorithm automatically provides the appropriate backbone dihedral angles for the

 10.1021/ja909973n © 2010 American Chemical Society

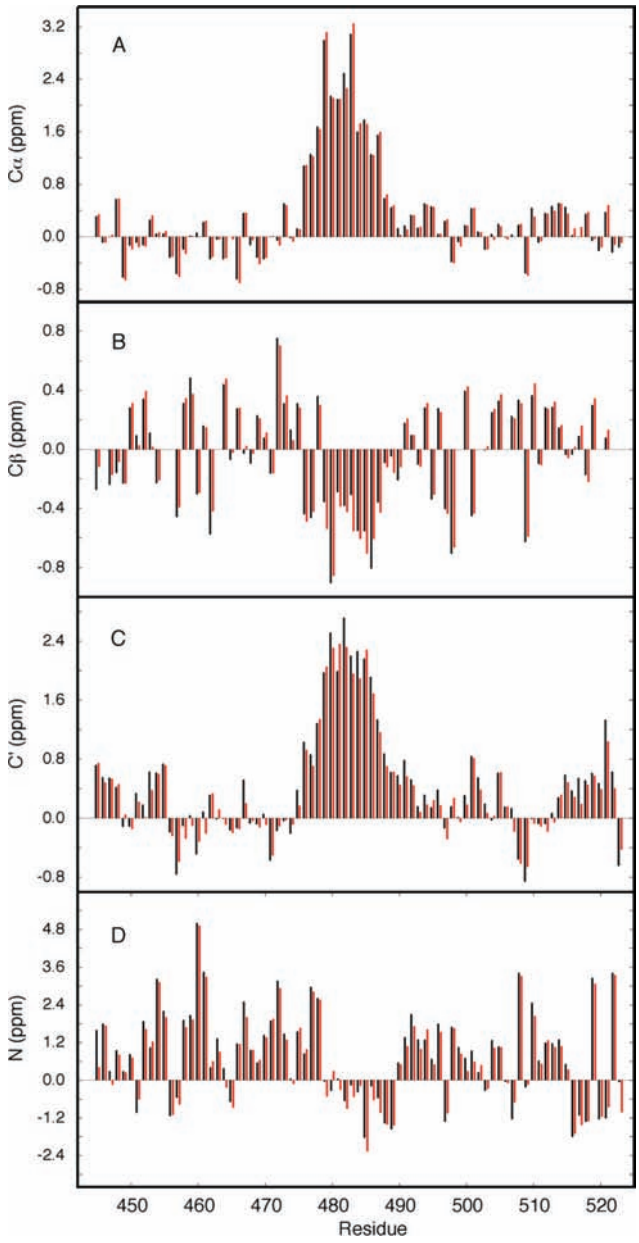

Figure 1. Reproduction of experimental secondary chemical shifts (random coil values from RefDB[25] were used) from an ensemble of 200 structures determined using the ASTEROIDS algorithm (3 iterations). Chemical shifts were calculated using the program SPARTA. *Flexible-Meccano* was used to calculate ensembles of structures of the protein, and iterative selection of 200-strong subensembles provided a final ensemble in agreement with experimental shifts. Black: experimental secondary chemical shifts. Red: secondary chemical shifts averaged over the final ensemble. (A) α carbon, (B) *β* carbon, (C) carbonyl, (D) amide nitrogen.

construction of entire secondary structural elements, as well as determining local conformational sampling in the unfolded domains.

This analysis is applied here to the study of N_{TAIL}, the C-terminal domain of the Sendai virus nucleoprotein. The molecular recognition element of N_{TAIL} has been shown, using detailed analysis of multiple RDCs, to contain a conformationally fluctuating helical element at its center.[33,34] This protein is a particularly appropriate model with which to test the approach, as it contains both partially structured and fully disordered elements. Figure 1 shows the agreement between the experimental and the calculated secondary chemical

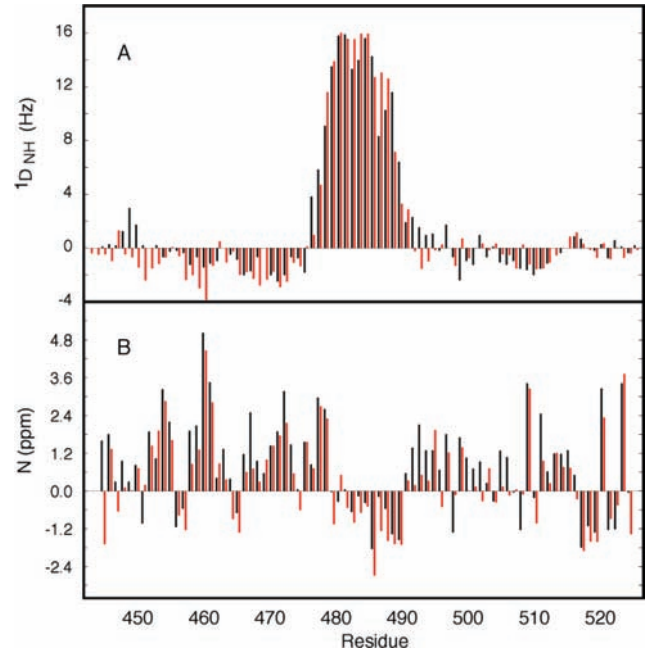

Figure 2. Reproduction of independent parameters by the ensemble based on chemical shift selection. (A) $^{15}N-^{1}H$ residual dipolar couplings (RDCs) measured in sterically aligned N_{TAIL}.[33] 50 000 conformers were calculated using the amino acid specific description of N_{TAIL} determined from the chemical shifts. RDCs were calculated using the program PALES directly from the ensemble and averaged. Simulated RDCs (red) were scaled uniformly to best match experiment (black). (B) Reproduction of ^{15}N secondary chemical shifts using an ensemble determined from only $^{13}C\alpha$, $^{13}C\beta$, $^{13}C'$ shifts (black: experiment, red: simulation).

shifts in N_{TAIL} after application of the ASTEROIDS algorithm. Excellent agreement with experimental shifts is observed throughout the protein.

The simple observation that data can be reproduced by a specific conformational ensemble does not necessarily guarantee that the ensemble is physically realistic. It is therefore essential to be able to cross-validate this approach with independent experimental data. Figure 2A shows the agreement between experimental $^{15}N-^{1}H$ RDCs measured from partially aligned N_{TAIL} compared to those calculated using an ensemble obtained from chemical shift derived conformational sampling. RDCs were calculated using the program PALES.[35] The agreement is striking, in both the folded and unfolded regions of the protein, demonstrating the ability of the algorithm to unambiguously interpret chemical shifts in terms of local conformational propensity. The level of helical structure agrees very closely with the helical description that was derived from analysis of RDCs,[28] indicating that the method is also quantitative. The ensemble dimensions also agree with those found in the previous studies (data not shown).[28] In a further test of consistency we have repeated the analysis in the absence of the ^{15}N chemical shifts and compared these shifts to predicted values (Figure 2B). Although this implies removing 25% of the data, experimental values are still reasonably reproduced (rmsd = 0.77 ppm compared to 1.15 ppm for the standard coil distribution).

An analysis of the ϕ/ψ distribution of the selected conformers outside the helical element reveals that the fully disordered regions of the protein have an overall tendency to sample less *β*-extended $\{\phi/\psi \approx -135°/135°\}$ and more (on average 5%) polyproline II $\{\phi/\psi \approx -75°/150°\}$ than is present in standard random coil

databases. These trends are in qualitative agreement with observations based on complementary spectroscopic techniques.[36−38] We are currently applying similar analyses to chemical shifts from more proteins to determine general trends for backbone conformational propensities of IDPs.

The ability of chemical shifts to reproduce conformational sampling was tested using extensive simulation. Ensembles of a model unfolded sequence were created using the standard ϕ/ψ database, an extended database sampling more β-sheet and polyproline II regions, or a database sampling more α-helical conformations. The chemical shifts of these ensembles were calculated with SPARTA as described above. The three sets of chemical shifts (standard, extended, and helix) were subjected to ASTEROIDS for selection of a subensemble of 200 structures from a pool of conformers generated using the standard database. These simulations demonstrate that it is possible to obtain a standard coil, more extended sampling, or a more helical sampling directly from chemical shifts, to within 5% accuracy (results not shown).

The ability to describe conformational sampling on the basis of chemical shifts alone is important for the development of atomic resolution descriptions of IDPs. The approach presented here makes no assumption concerning the true conformational properties of the molecule, starting with a standard statistical coil description of backbone conformational sampling, and refining this iteratively until convergence is reached compared to the experimental data. This allows the identification and characterization of entire secondary structural elements and their associated populations, as well as providing indications of the subtle detail of local conformational sampling in unfolded proteins. The approach is entirely compatible with recently presented ensemble selection algorithms based on the use of complementary structural restraints such as RDCs or ^3J scalar couplings, providing a tool for the development of a unified conformational model of partially ordered states. Possibly more exciting, this technique raises the prospect of probing the conformational behavior of unfolded proteins under conditions where additional parameters cannot be easily measured but where chemical shifts are still accessible, for example in crowded or cellular environments.[39]

Acknowledgment. This work was supported by the Commisariat à l'Energie Atomique, the French CNRS, the Université Joseph Fourier, Grenoble, and through ANR Protein Motion PCV (2007) ANR-07-PCVI PROTEIN MOTION. M.R.J. is supported by Lundbeckfonden and is recipient of a long-term EMBO fellowship.

Supporting Information Available: Complete ref 17. Materials and methods, describing use of SPARTA, application of ASTEROIDS for ensemble selection. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Uversky, V. N. *Protein Sci.* **2002**, *11*, 739–756.
(2) Dyson, H. J.; Wright, P. E. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197–208.
(3) Eliezer, D. *Curr. Opin. Struct. Biol.* **2009**, *19*, 23–30.
(4) Tompa, P.; Fuxreiter, M. *Trends Biochem. Sci.* **2008**, *33*, 2–8.
(5) Sugase, K.; Dyson, H. J.; Wright, P. E. *Nature* **2007**, *447*, 1021–1025.
(6) Dyson, H. J.; Wright, P. E. *Chem. Rev.* **2004**, *104*, 3607–3622.
(7) Mittag, T.; Forman-Kay, J. D. *Curr. Opin. Struct. Biol.* **2007**, *17*, 3–14.
(8) Jha, A. K.; Colubri, A.; Freed, K.; Sosnick, T. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13099–13105.
(9) Marsh, J. A.; Forman-Kay, J. D. *J. Mol. Biol.* **2009**, *391*, 359–374.
(10) Jensen, M. R.; Markwick, P.; Griesinger, C.; Zweckstetter, M.; Meier, S.; Grzesiek, S.; Bernado, P.; Blackledge, M. *Structure* **2009**, *17*, 1169–1185.
(11) 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.
(12) Krzeminski, M.; Fuentes, G.; Boelens, R.; Bonvin, A. M. J. J. *Proteins* **2009**, *74*, 895–904.
(13) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *J. Mol. Biol.* **1991**, *222*, 311–333.
(14) Spera, S.; Bax, A. *J. Am. Chem. Soc.* **1991**, *113*, 5490–5492.
(15) Osapay, K.; Case, D. A. *J. Biomol. NMR* **1994**, *4*, 215–230.
(16) Cavalli, A.; Salvatella, X.; Dobson, C. M.; Vendruscolo, M. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 9615–9620.
(17) Shen, Y.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 4685–4690.
(18) Schwarzinger, S.; Kroon, G. J. A.; Foss, T. R.; Chung, J.; Wright, P. E.; Dyson, H. J. *J. Am. Chem. Soc.* **2001**, *123*, 2970–2978.
(19) Yao, J.; Chung, J.; Eliezer, D.; Wright, P. E.; Dyson, H. J. *Biochemistry.* **2001**, *40*, 3561–3571.
(20) Wang, Y. J.; Jardetzky, O. *J. Am. Chem. Soc.* **2002**, *124*, 14075–14084.
(21) Eghbalnia, H. R.; Wang, L.; Bahrami, A.; Assadi, A.; Markley, J. L. *J. Biomol. NMR* **2005**, *32*, 71–81.
(22) Marsh, J. A.; Singh, V. K.; Jia, Z. C.; Forman-Kay, J. D. *Protein Sci.* **2006**, *15*, 2795–2804.
(23) Modig, K.; Jürgensen, V. W.; Lindorff-Larsen, K.; Fieber, W.; Bohr, H. G.; Poulsen, F. M. *FEBS Lett.* **2007**, *581*, 4965–4971.
(24) De Simone, A.; Cavalli, A.; Hsu, S. T.; Vranken, W.; Vendruscolo, M. *J. Am. Chem. Soc.* **2009**, *131*, 16332–16333.
(25) Zhang, H. Y.; Neal, S.; Wishart, D. S. *J. Biomol. NMR* **2003**, *25*, 173–195.
(26) Peti, W.; Smith, L. J.; Redfield, C.; Schwalbe, H. *J. Biomol. NMR* **2001**, *19*, 153–165.
(27) 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.
(28) Shen, Y.; Bax, A. *J. Biomol. NMR* **2007**, *38*, 289–302.
(29) Alexandrescu, A. T.; Abeygunawardana, C.; Shortle, D. *Biochemistry* **1994**, *33*, 1063–1072.
(30) Zhang, O. W.; Kay, L. E.; Olivier, J. P.; Forman-Kay, J. D. *J. Biomol. NMR* **1994**, *4*, 845–858.
(31) Meier, S.; Grzesiek, S.; Blackledge, M. *J. Am. Chem. Soc.* **2007**, *129*, 9799–9807.
(32) Nodet, G.; Salmon, L.; Ozenne, V.; Meier, S.; Jensen, M. R.; Blackledge, M. *J. Am. Chem. Soc.* **2009**, *131*, 16968–16975.
(33) Jensen, M. R.; Houben, K.; Lescop, E.; Blanchard, L.; Ruigrok, R. W. H.; Blackledge, M. *J. Am. Chem. Soc.* **2008**, *130*, 8055–8061.
(34) Jensen, M. R.; Blackledge, M. *J. Am. Chem. Soc.* **2008**, *130*, 11266–11267.
(35) Zweckstetter, M.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 3791–3792.
(36) Maiti, N. C.; Apetri, M. M.; Zagorski, M. G.; Carey, P. R.; Anderson, V. E. *J. Am. Chem. Soc.* **2004**, *126*, 2399–2408.
(37) Shi, Z. S.; Chen, K.; Liu, Z. G.; Kallenbach, N. R. *Chem. Rev.* **2006**, *106*, 1877–1897.
(38) Woody, R. W. *J. Am. Chem. Soc.* **2009**, *131*, 8234–8245.
(39) Serber, Z.; Dötsch, V. *Biochemistry* **2001**, *40*, 14317–14323.

JA909973N