JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Communication

pubs.acs.org/JACS

Characterization of the Conformational Equilibrium between the Two Major Substates of RNase A Using NMR Chemical Shifts

Carlo Camilloni,[†] Paul Robustelli,[†,‡] Alfonso De Simone,[†,§] Andrea Cavalli,[†] and Michele Vendruscolo*[,†]

[†]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, U.K.

[‡]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032, United States

[§]Division of Molecular Biosciences, Imperial College, South Kensington Campus, London SW7 2AZ, U.K.

Supporting Information

ABSTRACT: Following the recognition that NMR chemical shifts can be used for protein structure determination, rapid advances have recently been made in methods for extending this strategy for proteins and protein complexes of increasing size and complexity. A remaining major challenge is to develop approaches to exploit the information contained in the chemical shifts about conformational fluctuations in native states of proteins. In this work we show that it is possible to determine an ensemble of conformations representing the free energy surface of RNase A using chemical shifts as replica-averaged restraints in molecular dynamics simulations. Analysis of this surface indicates that chemical shifts can be used to characterize the conformational equilibrium between the two major substates of this protein.

Growing evidence indicates that chemical shifts can provide sufficient information for the determination of the three-dimensional structure of small-size proteins.[1,2] In addition, as chemical shifts are obtained from time- and ensemble-averaged measurements they contain information about the conformational fluctuations of proteins. This aspect has long been recognized, and several methods have been developed to exploit this information. It has been shown that chemical shifts can be used to predict S^2 order parameters and to obtain a score for the flexibility of different regions of an amino acid sequence folded in its native state.[3] It has also been shown that when the chemical shifts are calculated from ensembles of structures that represent simultaneously the structure and the dynamics of proteins, the agreement with the experimental chemical shifts improves with respect to the case in which a single average structure is considered.[4,5] In the present study we investigate whether chemical shifts can be used to generate structural ensembles by adopting strategies similar to those used with other NMR observables, including S^2 order parameters, RDCs, and PREs.[6−9] Our results indicate that this method can lead to characterization of the two major substates of ribonuclease A (RNase A), as well as of the conformational fluctuations between them.

In order to probe the amount of information about conformational fluctuations contained in the chemical shifts, we implemented chemical shifts as structural restraints in replica-averaged molecular dynamics simulations.[6,9] This use of the chemical shifts is obtained by the addition to the force field of the E_{CS} energy term, defined as

$$E_{CS} = \sum_{i=1}^{N} \sum_{j=1}^{6} E_{ij}(\delta_{ij}^{calc} - \delta_{ij}^{\exp}) \quad (1)$$

where E_{ij} is a chemical shift based energy term for an atom of type j (i.e., Hα, HN, N, Cα, Cβ, and C′) in the ith-residue of a protein of length N.[10,11] The overall E_{CS} energy term is calculated by considering the difference between the experimental chemical shifts, δ_{ij}^{exp}, of the atoms and their corresponding calculated values, δ_{ij}^{calc}, which are obtained as averages

$$\delta_{calc}^{ij} = \frac{1}{M} \sum_{k=1}^{M} \delta_{calc}^{ij,k} \quad (2)$$

where the index k runs over the M replicas used in the simulations. The averaging procedure over M replicas in the simulations makes it possible to account for the time- and ensemble-averaging intrinsic in the NMR measurements.[6−9] In eq 2, chemical shifts were calculated using the CamShift method,[10] which provides the chemical shifts corresponding to a given conformation through a differentiable function of its atomic coordinates. This feature enables the calculation of forces starting from eq 1 that are used to bias the trajectory toward a region of the conformational space where the chemical shifts match the experimental ones. Since the E_{ij} term in eq 1 involves primarily interatomic distances, its calculation and its differentiation are particularly rapid, given that such distances are already calculated at each time step for the estimation of the force field terms. This use of chemical shifts as structural restraints in molecular dynamics simulations has been already introduced, but without replica averaging,[11] to determine the structures of small globular proteins. The inclusion of replica-averaged restraints discussed here has the objective of minimizing the problem of over-restraining.[12] This problem appears in the presence of structural fluctuations or a relatively large amplitude, when imposing structural restraints on a single conformation can result in distortions to the underlying physical force fields used in the structure calculation. The resulting conformations are in these cases poor

Received: November 21, 2011
Published: February 9, 2012

   dx.doi.org/10.1021/ja210951z | J. Am. Chem. Soc. 2012, 134, 3968−3971

representatives of the average structure and, when presented as a structural ensemble, reflect the magnitude of the errors in the structure calculation protocol rather than a representation of the range of conformations that are populated in solution because of thermal fluctuations. On the other hand, when addressing the problem of over-restraining with replica-averaged restraints, it is important to optimize the protocol to minimize problems of overfitting,[12] which can occur if too many replicas are used relative to the information content of the experimental restraints. If too many replicas are used, the total number of degrees of freedom of all the molecules simulated exceeds the number of experimental data, and there are many ways of satisfying the restraints, making it impossible to identify the correct solution. To establish an optimal balance between over-restraining and overfitting given the information content contained in the CamShift restraints, the method was tested using 1, 2, and 4 replicas (see below and Supporting Information (SI)). For comparison, these replica simulations were also repeated without enforcing the chemical shift restraints with an otherwise identical protocol (see below and SI).

In this work we considered the case of RNase A, since this protein exhibits interdomain motions of fairly large amplitude and therefore represents a challenging test for any method aimed at characterizing the conformational fluctuations of proteins. Before using experimentally measured chemical shifts, to carry out a rigorous test of the approach that we propose we applied the test of the reference ensemble.[12,13] In this method, a 'reference ensemble' of conformations is generated at first by unrestrained molecular dynamics simulations. Chemical shifts are then calculated from the structures of this reference ensemble and employed as structural restraints to generate a 'restrained ensemble'. In this test, the force fields used in the reference and restrained cases are different, and thus they give rise to distinct free energy surfaces, and it is only the use of the chemical shift restraints that makes them similar (see Figure 1).

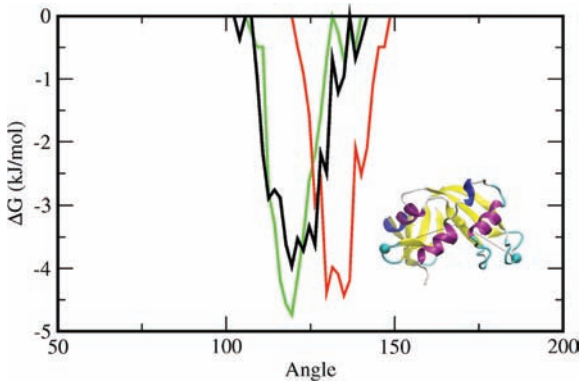

Figure 1. Comparison of the free energy profiles of the reference (black), unrestrained (red), and restrained (green) ensembles of RNase A. Free energy profiles are calculated as a function of a hinge angle defined by the three Cα atoms of residues T45, K91, and P114, which are shown in the inset.

An advantage of using this reference ensemble test is that it allows for a stringent validation analysis in which the atomic coordinates of the conformations in the reference ensemble are known exactly and therefore the accuracy of the conformations in the restrained ensemble can be assessed with great confidence. The reference ensemble was generated by extracting an ensemble of conformations from a long series of annealing cycles of molecular dynamics using the Gromos96 force field (see SI for further details). Chemical shifts were then back calculated using Sparta+[14] from this ensemble and used as restraints in two new molecular dynamics simulations with the Amber99SB force field using a modified version of Gromacs.[15,16] The use of two different methods for back-calculating chemical shifts (Sparta+ for obtaining the chemical shifts from the reference ensemble, and CamShift to generate the restrained ensemble) reproduces the effects of introducing errors in the structure-based predictions of chemical shifts, as the differences between the chemical shifts calculated by Sparta+ and CamShift for the reference ensemble are of the same magnitude of the typical errors made by these prediction methods (Table S1).

We first assessed the ability of the restrained and unrestrained simulations to reproduce the chemical shift values corresponding to the reference ensemble (Table S2). The results indicate that the chemical shift values from the restrained simulations match more closely those of the reference simulations than those of the unrestrained simulations, thus showing that the chemical shift restraints are effective in driving the trajectories to satisfy the restraints.

The best results were obtained in the 2-replica case (Table S2). In order to assess whether the better agreement with the imposed restraints corresponds to a better description of the structural heterogeneity of the reference ensemble we compared the distributions of the distances between Cα atoms. We built a 124 × 124 S matrix[12] (Figure S1) in which each element S_{ij} represents the difference between the distributions P^{ref} and P^{sim} of the distances in the reference and simulated ensembles,

$$S_{ij} = \frac{1}{2}\sum_{k} |P^{ref}_{ij,k} - P^{sim}_{ij,k}| \qquad (3)$$

where k runs over the bins used to characterized the distributions; the subscripts ij indicate that the P^{ref} and P^{sim} distributions refer to the Cα atom pair of residues i and j. The S matrix provides an accurate description of both local and nonlocal structural and dynamical similarity for two ensembles. An S_{ij} of 0 corresponds to two identical distributions while an S_{ij} of 1 corresponds to two nonoverlapping distributions. The average dissimilarity changes from 0.46 for the unrestrained simulations to 0.39 for the restrained ones.

The differences of the distance distributions between the unrestrained and the reference ensembles are related to the differences between the two force fields (Gromos96 and Amber99SB) that we used to generate them, and are located mainly in the loop regions of the RNase A (A20-Y25, R33-K37, K66-T70, and S89-P93). After the addition of the restraints there is an overall improvement across the entire protein, including in the loop regions, in agreement with the reference ensemble. The results for the S matrix indicate that observable functions of the distances between atoms, if calculated on the restrained ensemble, are in better agreement with the reference ensemble than if calculated on the unrestrained ensemble. In order to illustrate this point we show the free energy landscape as a function of an angle describing the relative motion of the two domains of RNase A calculated on the reference, the restrained and the unrestrained ensembles (Figure 1). The addition of the chemical shift restraints averaged over two replicas resulted in a change of the free energy landscape of

dx.doi.org/10.1021/ja210951z | J. Am. Chem. Soc. 2012, 134, 3968−3971

RNase A, from the one centered at 130° to the one centered at 120° in excellent agreement with the free energy landscape of the reference ensemble.

In order to verify whether the chemical shift restraints generate an ensemble of conformation that does not depend significantly on the underlying force field used, we carried out a second series of restrained molecular dynamics simulations, this time using the Amber03 force field, instead of the Amber99SB force field. We found that the Amber03 and Amber99SB restrained ensembles have an *S* score of 0.19 (Table S3), which is essentially at the level of statistical errors, which is estimated to be about 0.14 by comparing the first and the second half of the reference ensemble trajectory (Table S3). These results indicate that the use of chemical shift restraints overcomes the differences in the force fields to generate an ensemble of structures maximally compatible with the given set of chemical shifts.

Having assessed in this way the ability of molecular dynamics simulations with replica-averaged chemical shift restraints to characterize conformational fluctuations, we applied the method to RNase A using experimental chemical shifts (BMRB 4031). The approach that we discuss in this work enabled us to obtain an ensemble of conformations describing the native state dynamics of this protein (Figure 2). We then

Figure 2. Superposition of 100 low-energy structures of the RNase A ensemble generated by using chemical shift restraints (BMRB 4031). The structure of RNase A consists of two antiparallel β-sheets, V1 and V2, which form a characteristic V-shaped motif;[17] colors represent different secondary structure types.

looked at the distribution of the hinge angle[17] (see Figure 1) in the structures of RNase A present in the PDB, both in the free state (Figure 3a, red bars) and in the bound state (Figure 3a, black bars). The distribution of this angle in the ensemble of structures that we determined (Figure 3b) covers the range of values observed in the PDB. In Figure S2 this motion is represented through the first eigenvector of the principal component analysis of the ensemble. These results confirm that the type of motion that we obtained for this angle is consistent with the conformational variability observed in the available structures of RNase A and suggest the presence of a conformational selection mechanism[18] for this protein.

We then considered the motions in the active site of RNase A. This protein is known to be relatively rigid on time scales from ps to ns[19,20] (Table S4), while larger amplitude motions, which are important for its function, take place on a time scale from μs to ms.[18,20−25] These functional dynamics involve multiple residues belonging to the active site (H12, K41, H119, D121), the phosphate-binding subsite P0 (K66), and the

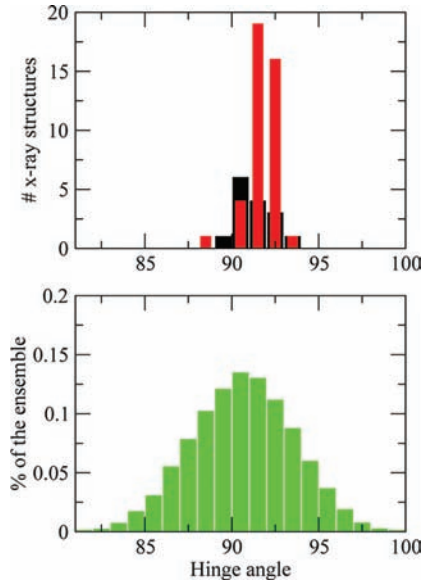

Figure 3. Comparison of the hinge angle distributions for all the X-ray structures of RNase A in the PDB (upper panel; red bars refer to the free state, and black bars to the bound state) and those in the ensemble that we determined using chemical shift restraints (lower panel).

binding subsites B1 (T45, D83) and B2 (Q69, E111).[19−21,24] The structures of RNase A deposited in the PDB can be grouped into two clusters, generally known as A and B states, depending on the conformations of these residues in the active site. We found that the structures in the ensemble that we determined could be clustered in the same way, resulting in a free energy landscape with two minima (Figure 4). This

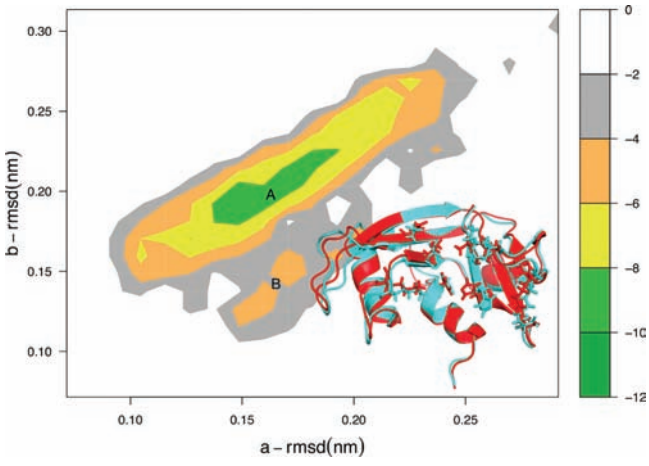

Figure 4. Conformational equilibrium of RNase A. By using chemical shift restraints we found the two most populated clusters of structures, corresponding to the A state (red) and the B state (cyan). The two order parameters used to represent the free energy landscape correspond to the root-mean-square distances between the conformations in the ensemble that we generated and those in the PDB in the A (a-rmsd) and B (b-rmsd) states, respectively.

analysis revealed that such minima cover 95% of the structures. The A state (Figure 4 red) can be further divided into an open-A state (hinge angle >90° and H119 in conformation A[24]), which represents 55% of the conformations in the ensemble, and a close-A state (hinge angle ≤90° and H119 in

dx.doi.org/10.1021/ja210951z | J. Am. Chem. Soc. 2012, 134, 3968−3971

conformation A), which represents 25% of the conformations in the ensemble. The B state (Figure 4 magenta) can be analogously divided into an open-B state (hinge angle >90° and H119 in conformation B[24]), which represents 10% of the conformations in the ensemble, and in a close-B state, which included 5% of the conformations in the ensemble (hinge angle ≤90° and H119 in conformation B) and whose presence has been identified by relaxation dispersion methods.[21] Finally, the remaining 5% of the ensemble is in higher energy conformations that cannot be directly linked to A or B. These results indicate that the approach that we presented is effective in extracting from the chemical shifts the information about the conformational fluctuations of RNase A.

In summary, by considering the case of RNase A we have described an approach in which chemical shifts are used to characterize the conformational fluctuations in the native states of proteins. The procedure that we used, in which chemical shifts are enforced as restraints averaged over multiple replicas in molecular dynamics simulations, is generally applicable to other NMR parameters and, indeed, can be used in combination with them.

■ ASSOCIATED CONTENT

Supporting Information

Methods, Tables S1−S4, Figures S1 and S2. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

mv245@cam.ac.uk

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This study was funded by FEBS (C.C.), EPSRC (A.D.S), Wellcome Trust (A.C. and M.V.), and BBSRC (A.C. and M.V.).

■ REFERENCES

(1) Cavalli, A.; Salvatella, X.; Dobson, C. M.; Vendruscolo, M. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 9615.

(2) Shen, Y.; Lange, O.; Delaglio, F.; Rossi, P.; Aramini, J. M.; Liu, G. H.; Eletsky, A.; Wu, Y. B.; Singarapu, K. K.; Lemak, A.; Ignatchenko, A.; Arrowsmith, C. H.; Szyperski, T.; Montelione, G. T.; Baker, D.; Bax, A. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 4685.

(3) Berjanskii, M. V.; Wishart, D. S. *J. Am. Chem. Soc.* **2005**, *127*, 14970.

(4) Jensen, M. R. J. M. R.; Salmon, L.; Nodet, G.; Blackledge, M. *J. Am. Chem. Soc.* **2010**, *132*, 1270.

(5) De Gortari, I.; Portella, G.; Salvatella, X.; Bajaj, V. S.; van der Wel, P. C. A.; Yates, J. R.; Segall, M. D.; Pickard, C. J.; Payne, M. C.; Vendruscolo, M. *J. Am. Chem. Soc.* **2010**, *132*, 5993.

(6) Clore, G. M.; Schwieters, C. D. *J. Am. Chem. Soc.* **2004**, *126*, 2923.

(7) Dedmon, M. M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. *J. Am. Chem. Soc.* **2005**, *127*, 476.

(8) Huang, J. R.; Grzesiek, S. *J. Am. Chem. Soc.* **2010**, *132*, 694.

(9) Lindorff-Larsen, K.; Best, R. B.; DePristo, M. A.; Dobson, C. M.; Vendruscolo, M. *Nature* **2005**, *433*, 128.

(10) Kohlhoff, K. J.; Robustelli, P.; Cavalli, A.; Salvatella, X.; Vendruscolo, M. *J. Am. Chem. Soc.* **2009**, *131*, 13894.

(11) Robustelli, P.; Kohlhoff, K.; Cavalli, A.; Vendruscolo, M. *Structure* **2010**, *18*, 923.

(12) Richter, B.; Gsponer, J.; Varnai, P.; Salvatella, X.; Vendruscolo, M. *J. Biomol. NMR* **2007**, *37*, 117.

(13) Kuriyan, J.; Petsko, G. A.; Levy, R. M.; Karplus, M. *J. Mol. Biol.* **1986**, *190*, 227.

(14) Shen, Y.; Bax, A. *J. Biomol. NMR* **2010**, *48*, 13.

(15) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theor. Comp.* **2008**, *4*, 435.

(16) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. *Comput. Phys. Commun.* **2009**, *180*, 1961.

(17) Vitagliano, L.; Merlino, A.; Zagari, A.; Mazzarella, L. *Proteins* **2002**, *46*, 97.

(18) Watt, E. D.; Shimada, H.; Kovrigin, E. L.; Loria, J. P. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 11981.

(19) Cole, R.; Loria, J. P. *Biochemistry* **2002**, *41*, 6072.

(20) Kovrigin, E. L.; Cole, R.; Loria, J. P. *Biochemistry* **2003**, *42*, 5279.

(21) Beach, H.; Cole, R.; Gill, M. L.; Loria, J. P. *J. Am. Chem. Soc.* **2005**, *127*, 9167.

(22) Doucet, N.; Khirich, G.; Kovrigin, E. L.; Loria, J. P. *Biochemistry* **2011**, *50*, 1723.

(23) Kovrigin, E. L.; Loria, J. P. *J. Am. Chem. Soc.* **2006**, *128*, 7724.

(24) Kovrigin, E. L.; Loria, J. P. *Biochemistry* **2006**, *45*, 2636.

(25) Doucet, N.; Watt, E. D.; Loria, J. P. *Biochemistry* **2009**, *48*, 7160.