J|A|C|S COMMUNICATIONS

Published on Web 04/13/2009

Autocorrelation Analysis of NOESY Data Provides Residue Compactness for Folded and Unfolded Proteins

Andreas Schedlbauer, Nicolas Coudevylle, Renate Auer, Karin Kloiber, Martin Tollinger, and Robert Konrat*

Institute of Biomolecular Structural Chemistry, Max F. Perutz Laboratories, University of Vienna, Campus Vienna Biocenter 5, A-1030 Vienna, Austria

Received September 18, 2008; E-mail: Robert.Konrat@univie.ac.at

The nuclear Overhauser effect (NOE) is the most important and most informative experimental parameter for structure determination by NMR in solution, as its sensitive dependence on internuclear distances can be recast into 3D structural information about proton−proton distance relationships.[1] In this manuscript we use unassigned NOESY spectra to probe the structural compactness of individual residues in 3D protein structures. The approach is based on our NMR autocorrelation analysis for studying protein foldedness in solution.[2] It merely requires a 1D protein ^{1}H spectrum $S(\omega)$ to compute the autocorrelation function $C(\omega)$. It was shown that $C(\omega)$ is related to the distribution function of resonance frequency differences and provides an efficient way to observe protein structural stability in solution. Additionally, the relationship between $C(\omega)$ and the topological complexity, quantified by the relative contact order,[3] revealed that this approach can also be used to qualitatively assess a protein's fold.[2] Here we apply this methodology to the analysis of protein 3D NOESY spectra. Interproton NOEs observed in NOESY spectra provide information about resonance frequencies of spins located in the surrounding and thus the structural complexity of the spatial environment at a particular site in the protein.

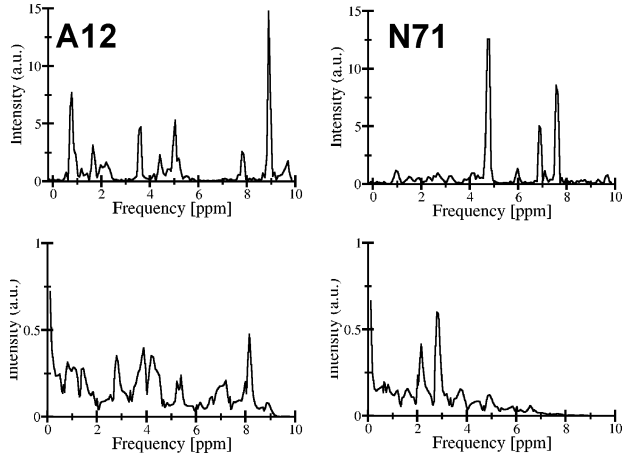

Figure 1. Experimental traces from a 3D ^{15}N NOESY-HSQC (top) and autocorrelation functions (obtained by self-convolution, see text, bottom) obtained for A12 (left) and Asn71 (right) of the 17.7 kDa mitochondrial matrix protein cyclophilin D (CypD).

Our strategy is as follows: Individual traces $S_i(\omega)$ from a 3D ^{15}N NOESY-HSQC are taken along the indirect F_1 (NOE) dimension at frequency positions of the different amide groups (backbone residue position i) of the protein. The $S_i(\omega)$ trace is then convoluted with itself yielding the autocorrelation function $C_i(\omega)$ of the NOESY-trace for residue position i. Figure 1 shows experimental 3D ^{15}N NOESY-HSQC traces (Ala12 and Asn71) obtained for the protein cyclophilin D (CypD), a 17.7 kDa mitochondrial matrix protein involved in opening of the mitochondrial permeability transition pore.[4] For comparison, the individual $C_i(\omega)$ are normalized ($\omega = 0$) and numerically smoothed (± 2). It can be seen that significantly different NOESY-HSQC traces and respective autocorrelation functions are found for residues located on the surface (Asn71) or in the hydrophobic core (Ala12) of CypD. Particularly noteworthy is the fact that in general the autocorrelation function $C_i(\omega)$ has a better signal-to-noise than the corresponding spectral data set. For quantification we take the spectral entropy as an information measure, which is defined as the integral:

$$\int C(\omega) \cdot \ln C(\omega)\, d\omega$$

Since individual NOESY traces are also determined by intramolecular NOEs and thus depend on the amino acid type, we calculated amino acid specific average entropy values. For the analysis of the spatial environment of a given residue, the differential entropy value S_ν (deviation from the amino acid type specific average entropy value) is used. Significantly different (differential) entropy values S_ν can be expected for surface exposed residues or residues in hydrophobic core regions of the protein. In Figure 2 experimental differential S_ν values obtained for CypD are shown as a function of residue position.

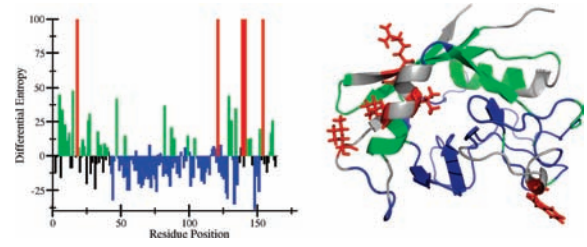

Figure 2. Differential spectral entropy S_ν as function of residue position for CypD. Differential S_ν values are defined as differences to average residue type-specific S_ν values. Large positive differential S_ν values are indicative of residue positions comprising loosely packed side chains, whereas negative differential S_ν values are indicative of core regions of the protein structure with tightly interacting side chains. The location of residues displaying positive (red, green) and negative (blue) differential S_ν entropy values are mapped onto the 3D structures of CypD. For residues displaying very large differential S_ν values (red), the side chains are also indicated.

The structure of CypD consists of nine β-strands forming a β-barrel structure, and it is best viewed as a sandwich of two roughly orthogonal β-sheets, in which the first sheet is composed of strands β_1 to β_3 (L5-A12, Q15-L24, L51-S52) and β_9 (I156-Q163), whereas the second sheet is formed by strands β_4 to β_7 (R55-I57, M61-A64, V97-M100, F112-C115). β_8 is part of both β-sheets. As can be seen from Figure 2 there is a significant correlation between differential entropy values S_ν and the location in the 3D structure. Residues displaying negative S_ν values (shown in blue) are consistently found in the second β-sheet, whereas residues with positive S_ν values (shown in green) are found either in the first β-sheet or in the two α-helices (capping the β-barrel). Residues displaying maximal positive S_ν values are found in exposed areas with

10.1021/ja8074067 CCC: $40.75

no significant secondary structure elements and their side chains pointing toward to the solvent. These findings indicate that the spectral entropy obtained from individual NOESY traces provides a sensible measure for the spatial surrounding of individual sites in a protein.

An experimental example illustrating the validity of the approach for intrinsically unstructured/disordered proteins is given with Osteopontin (OPN), a cytokine and cell attachment protein implicated in tumorigenesis.[5,6] Multidimensional NMR spectroscopy has shown that quail OPN is highly flexible in solution and devoid of any significant tertiary structure.[7] ^{13}C chemical shift analysis, however, provided evidence for the existence of local structural elements (α-helix: Leu58-Leu65; β-strand: around Thr119 and Val152-Ile158). Figure 3 shows good agreement between residual structure (small S_ν values) and reduced mobility (shorter ^{15}N T_2 and larger/positive η values), indicating that the proposed autocorrelation approach provides reliable information about local compaction of the polypeptide chain in unfolded protein states.

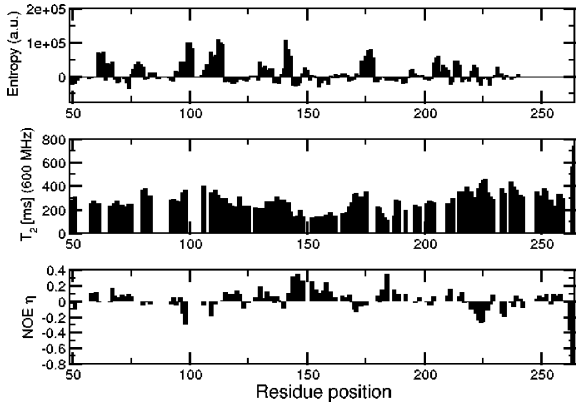

Figure 3. Differential spectral entropy S_ν (top), NMR spin relaxation (^{15}N T_2) (middle), and ^{15}N−^{1}H^N heteronuclear NOE, η, (bottom) as function of residue position for Osteopontin. Positive differential entropy values S_ν are indicative of residues with loosely packed side chains and reduced local flexibility.

Overall, large positive differential S_ν values indicative of residue positions comprising loosely packed side chains are paralleled by elevated T_2 and negative η values. Flexible parts (elevated S_ν values) are found within the following residue stretches 95−110, 130−140, 170−180, and 210−230, whereas considerable compaction of the polypeptide chain is observed for 110−130, 145−170, and 180−200. This correlation offers valuable possibilities for the studies of intrinsically unfolded proteins. Deviations occur because S_ν values probe spatial surroundings and not only local restrictions in mobility. While conventional NMR spin relaxation studies only provide information about the local restriction of motions (possibly as a consequence of local hydrogen bonding), the NOESY spectral entropy provides additional and novel information about transiently formed inter-residue contacts. The NOESY spectral entropy approach thus nicely complements existing NMR methodology for studying structural ensembles of natively unfolded proteins (e.g., RDC and PRE).[10,11] An obvious application of the approach is to use this information as a guiding tool for the identification of spin label attachement sites in PRE experiments of natively unfolded proteins.

Finally, the methodology was applied to a protein complex between Q83, a lipocalin protein highly overexpressed in fibroblasts transformed by the *v-myc* oncogene,[12] and the siderophore Enterobactin, an Fe(III)-chelator. Details are given in the Supporting Information. Interestingly, chemical shift changes were not limited to residues located in the neighborhood of the ligand but rather distributed over the backbone and thus pointing to a restructuring of the protein upon ligand binding. This plasticity was also seen in the changes of μs−ms dynamics (probed by CPMG), as both increased and decreased mobilities were observed. Interestingly, the residues experiencing an increased μs−ms dynamics suggest an additional hitherto unknown ligand binding site in the lipocaline protein family. Figure 4 illustrates that also changes in S_ν values were not localized to the ligand binding site but observed at distant sites. Specifically, residues located in one-half of the β-barrel and experiencing a reduction in conformational flexibility (Figure 4, blue) display significant S_ν changes (Figure 4, green). These findings nicely illustrate the potential of the methodology to identify structural adaption processes induced by ligand binding. Since NOESY data are typically available in a protein ligand structural analysis, this information can be obtained at no additional cost.

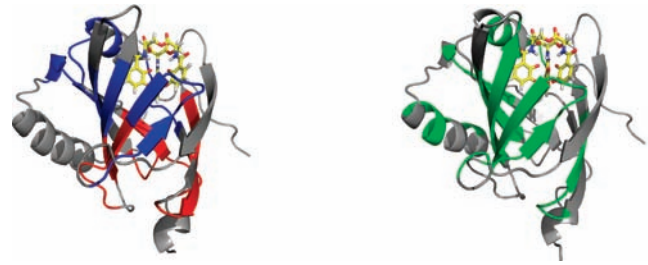

Figure 4. Differential conformational dynamics and differential entropy in Q83 upon binding of enterobactin. (Left) Differential exchange contributions to ^{15}N T_2 relaxation (μs−ms mobility; red: increased and blue: decreased flexibility). (Right) Residues displaying differential entropy S_ν values are color coded in green.

In summary, we presented a novel analysis tool for protein NOESY data based on spectral entropy calculated from self-convoluted 3D NOESY-HSQC data. Data obtained on stably folded as well as intrinsically unfolded proteins indicated that the approach provides information about local compaction of the polypeptide chain with fruitful applications to characterizations of dynamical conformational ensembles of intrinsically unstructured proteins. Applications to protein complexes provide information about changes in side-chain packing upon ligand binding and the existence of potential allosteric regulation effects. The inherent sensitivity of the autocorrelation offers unique possibilities. Experimental studies of protein and RNA kinetics using this novel approach are currently underway in our laboratory.

Acknowledgment. R.A. is a recipient of a DOC-fFORTE fellowship, and K.K. of an APART-fellowship of the Austrian Academy of Sciences. N.C. is a recipient of a Lise-Meitner FWF fellowship. This work was partly supported by grants (P19428 to M.T. and SFB17 to R.K.) and via generous start-up funding from the Austrian Science Foundation (FWF) (to R.K.).

Supporting Information Available: Description and comparison between chemical shift differences, ^{15}N CPMG data, and differential entropy. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Neuhaus, D.; Williamson, M. P. *The Nuclear Overhauser Effect in Structural and Conformational Analysis*; Verlag Chemie: New York, 1989.
(2) Hoffmann, B.; Eichmüller, C.; Steinhauser, O.; Konrat, R. *Methods Enzymol.* **2005**, *394*, 142–175.
(3) Plaxco, K. W.; Simons, K. T.; Baker, D. *J. Mol. Biol.* **1998**, (277), 985–994.
(4) Crompton, M. *J. Physiol.* **2000**, (1), 11–21.
(5) Bellahcène, A.; Castronovo, V.; Ogbureke, K. U. E.; Fisher, L. W.; Fedarko, N. S. *Nat. Rev. Cancer* **2008**, *8*, 212–226.
(6) Rittling, S. R.; Chambers, A. F. *Br. J. Cancer* **2004**, *90*, 1877–1881.
(7) Schedlbauer, A.; Ozdowy, P.; Kontaxis, G.; Hartl, M.; Bister, K.; Konrat, R. *Biomol. NMR Assignments* **2008**, *2*, 29–31.
(8) Mittag, T.; Forman-Kay, J. D. *Curr. Opin. Struct. Biol.* **2007**, *17*, 3–14.
(9) Dyson, H. J.; Wright, P. E. *Meth. Enzymol.* **2001**, *339*, 258–270.
(10) Blackledge, M. *Prog. NMR Spectrosc.* **2005**, *46*, 23–61.
(11) Lietzow, M. A.; Jamin, M.; Dyson, H. J.; Wright, P. E. *J. Mol. Biol.* **2002**, *322*, 655–662.
(12) Hartl, M.; Matt, T.; Schüler, W.; Siemeister, G.; Kontaxis, G.; Kloiber, K.; Konrat, R.; Bister, K. *J. Mol. Biol.* **2003**, *333*, 33–46.

JA8074067