Theor Chem Account (2008) 121:197–200
DOI 10.1007/s00214-008-0465-0

REGULAR ARTICLE

Prediction of NMR order parameters in proteins using weighted protein contact-number model

Shao-Wei Huang · Chien-Hua Shih · Chih-Peng Lin · Jenn-Kang Hwang

Received: 22 April 2008 / Accepted: 4 July 2008 / Published online: 1 August 2008

Abstract In the NMR experiment, the protein backbone motion can be described by the N–H order parameters. Though protein dynamics is determined by a complex network of atomic interactions, we show that the order parameter of residues can be determined using a very simple method, the weighted protein contact number model. We computed for each Cα atom the number of neighboring Cα atoms weighted by the inverse distance squared between them. We show that the weighted contact number of each residue is directly related to its order parameter. Despite the simplicity of this model, it performs better than the other method. Since we can compute the order parameters directly from the topological properties (such as protein contact number) of protein structures, our study underscores a very direct link between protein topological structure and its dynamics.

Keywords NMR order parameter · Weighted protein contact number · Protein dynamics · Prediction

Introduction

Given protein structures, the knowledge of protein dynamics is useful in suggesting potential protein active sites [1] or molecular recognition sites [2], or in understanding the mechanisms of enzymatic reactions [3]. With the increasing number of protein structures of unknown function deposited in Protein Data Bank (PDB) [4], it becomes increasingly important to develop efficient method to compute average dynamical properties of protein in a high-throughput fashion. Protein dynamics consists of a wide range of motional behavior arising from a complex network of atomic interactions. Molecular dynamics (MD) [3,5–10], taking into the atomic interactions through empirical force field, have proved to be a powerful method to compute protein dynamics, but it is impractical for large proteins due to the high computational cost [11].

Recently, Zhang and Bruschweiler [12] expressed the backbone S^2 order parameter as a function of close contacts between the amide proton and the carbonyl oxygen of the preceding amino acid and the surrounding protein atoms, i.e.,

$$S_i^2 = \tanh\left[\alpha \sum_k e^{-r^{O}_{i-1,k}/\rho} + \beta e^{-r^{H}_{i,k}/\rho}\right] + \gamma \quad (1)$$

where $r^{O}_{i-1,k}$ is the distance between the carbonyl oxygen of residue $i-1$ and heavy atom k and $r^{O}_{i-1,k}$ is the distance between the amide proton and heavy atom k. The parameter α, β and γ were determined empirically and ρ is set to 1 Å. We will refer to this method the contact model (CM). Despite the simplificity of the CM, it provides for many proteins a very accurate prediction of NMR order parameter. Later, to take into account of motional correlation effects, Bruschweiler and coworker developed a hybrid between the CM and the elastic network model (ENM) [13,14] referred to as reorientational contact-weighted elastic network model (rCENM) [15]. Here, we present a simple model to compute N–H backbone order parameters, which considers essentially only the contacting Cα atoms. However, despite its simplicity, this model performs better than other method.

Methods

We define the weighted protein contact number as

$$\nu_i = \sum_{j\neq i}^{N} 1/r_{ij}^2 \quad (2)$$

S.-W. Huang (✉) · C.-H. Shih · C.-P. Lin · J.-K. Hwang
Institute of Bioinformatics, National Chiao Tung University, HsinChu 30050, Taiwan, ROC
e-mail: swhwang.orz@gmail.com

Table 1 Comparison of the correlation coefficient between predicted and experimental N–H S^2 order parameters

Protein	PDB	WCN	WCN*	CM	rCENM
βARK1 PH domain	1bak	0.83	0.80	0.53	0.84
Calbindin	4icb	0.75	0.67	0.65	0.72
CspA	3mef	0.78	0.73	0.71	–
Frenolicin acyl carrier protein	1or5	0.85	0.79	0.89	0.87
Lysozyme	1jef	0.83	0.67	0.72	0.68
P85α SH2 domain	1bfj	0.86	0.74	0.79	–
Ubiquitin	1ubq	0.96	0.89	0.96	0.97
Ketosteroid isomerase	8cho	0.82	0.79	0.57	0.78
Taotomerase	4ota	0.51	0.55	0.44	–
Interleukin-4	1hik	0.71	0.72	0.81	0.81
Average correlation coefficient		0.79	0.73	0.71	–‡

‡ The average correlation coefficient over the seven structures for WCN, WCN*, CM and rCENM are 0.82, 0.76, 0.73 and 0.81, respectively
The experimental backbone N–H order parameter data: β ARK1 PH domain [16], Calbindin [20], CspA [22], Frenolicin acyl carrier protein [23], Lysozyme [18], P85α SH2 domain of phosphoinostide 3-kinase [25], Ubiquitin [17], Ketosteroid isomerase [26], Taotomerase [27], and Interleukin-4 [28]

where r_{ij} is the distance between Cα atoms of residue i and j. This equation essentially defines the contributions of neighboring Cα atoms to the ith residue—the contribution of each surrounding atom j to the central atom i is scaled down by the factor $1/r_{ij}^2$. To confine the computed order parameter S^2 to be lying between 0 and 1, we apply the hyperbolic tangent function to ν_i.

$$S_i^2 \sim \tanh^2 \nu_i \tag{3}$$

This is the main result of this work. We will refer to this method as the weighted contact-number (WCN) model.

Dataset

We used the datasets of Zhang and Bruschweiler [12] and Ming and Bruschweiler [15]. However, our dataset is not completely identical with the original one, since we could not find some of the order parameters that are consistent with those of the original dataset, and we also added some new ones from the current literature. Our current dataset is larger than the original dataset and comprises βARK1 PH domain (1bak), calbindin (4icb), CspA (3mef), frenolicin acyl carrier protein (1or5), lysozyme (1jef), P85α SH2 domain (1bfj), ubiquitin (1ubq), ketosteroid isomerase (8cho), tautomerase (4ota), and interleukin-4 (1hik).

Results and discussion

Table 1 summarizes the Pearson correlation coefficients between the NMR and the computed S^2 order parameters by the WCN, CM, rCENM as well as the WCN* model (see below). The WCN model generally performs better than the CM model—the average correlation coefficient between the NMR and computed S^2 order parameters is $\bar{c} = 0.79$, while that of the CM is $\bar{c} = 0.71$. In the case of rCENM, since its source code is not available, we can only compare their results available from the literature. For these seven structures, the performance of the WCN model ($\bar{c} = 0.82$) is comparable to that of the more complicated hybrid approach rCENM ($\bar{c} = 0.81$) for the seven structures.

To check the effects of the additional information of side-chain groups on the computed order parameters, we compared the computed S^2 values with (denoted as WCN*) and without (i.e., the WCN model) the side-chain information in Table 1. However, the inclusion of the side-chain groups deteriorates the performance of the WCN model ($\bar{c}$ goes down from 0.79 to 0.73). The reason for this is not clear. It may be that the flexible side-chain conformations, though conveying more detailed information about the atomic environments, introduce undesirable noises that overshadow the supposedly useful information of the former in the computation of the order parameters.

Figure 1 compares the experimental and the computed order-parameter profiles by the WCN for some of the proteins mentioned above. βARK1 PH domain has the same topology as other PH domains, which are characterized by several β-strands forming a β-sandwich flanked on one side by an extended C-terminal α-helix that behaves as a molten helix [16]. The Pearson correlation coefficient between the computed N–H order parameter S^2_{WCN} and the experimental one S^2_{NMR} is 0.83. Ubiquitin [17] is a small single-domain protein with 76 residues containing both an α-helix and β-sheets. The agreement between S^2_{WCN} and S^2_{NMR} of ubiquitin is excellent ($\bar{c} = 0.96$).

The correlation coefficient of the prediction for lysozyme [18] is 0.83. The order parameter of Pro-70 located on the

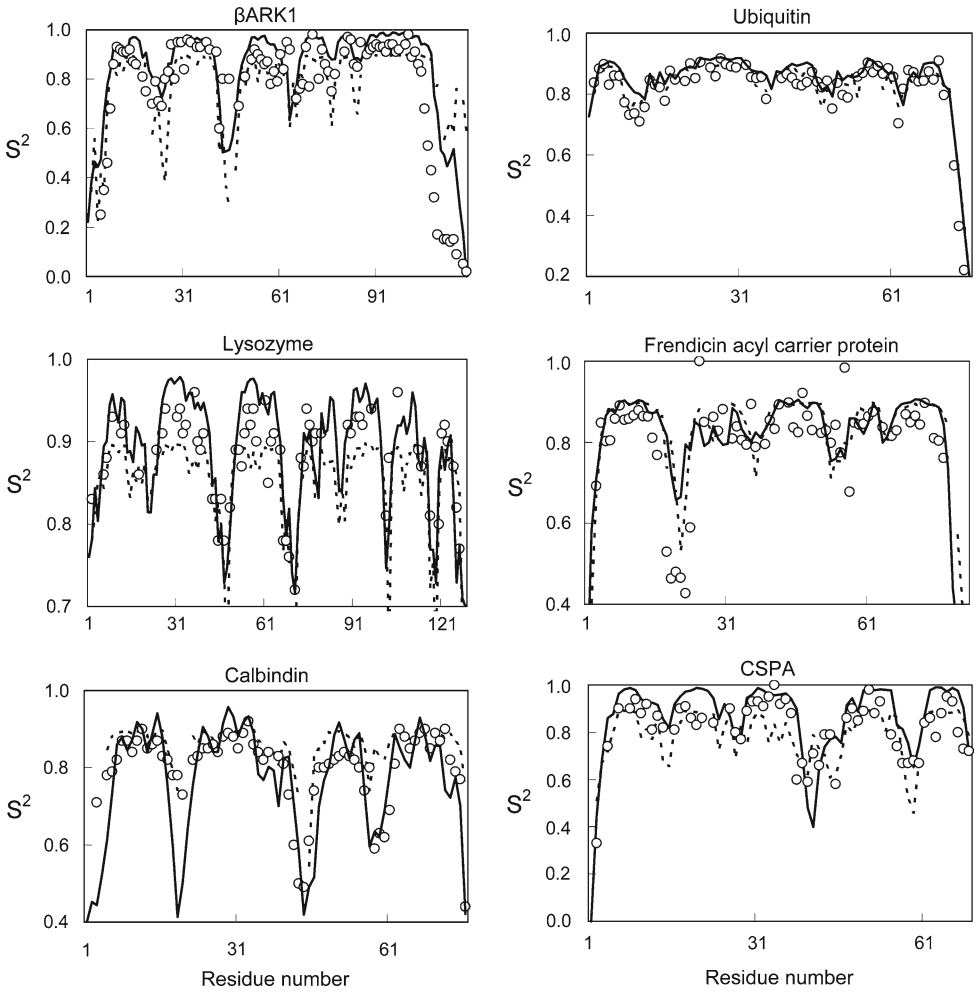

Fig. 1 Experimental (*open circle*) and predicted order parameter by WCN (*solid line*) and CM (*dotted line*) for β ARK1, Ubiquitin, Calbindin, lysozyme, Frendicin acyl carrier protein, and CspA

most flexible loop is not available from NMR relaxation. According to the X-ray structure of lysozyme, Pro-70 has the sixth highest temperature factor (28.37) in the whole protein [19] (the four residues having the highest temperature factors are on the C-terminal region). Our prediction also shows that Pro-70 is the most flexible except few residues located on the C-terminal.

Calbindin D_{9k} [20] is composed of four α-helices, the N-terminal (E17-S24) and C-terminal Ca^{2+}-binding loops (D54-S62), and the linker loop. Our prediction correctly identifies the most mobile linker loop and the C-terminal Ca^{2+}-binding loop which have significant lower S^2 values. The rigid helical regions are also predicted to have higher order parameters. The experimental data does not give the S^2 value of the Pro-20 on the N-terminal Ca^{2+}-binding loop because of the limit of NMR relaxation experiment. However Pro-20 shows higher temperature factor than its neighboring residues in the X-ray structure [21], which is consistent with our prediction. Despite the missing data of Pro-20, the correlation coefficient is still high ($r = 0.79$).

Cold-shock protein from *E. coli* (CspA) [22] is a Greek-key β-barrel protein. The segment of residues Asn-39 to Tyr-42 between two β-strands is identified to be partially disordered in the crystallization environment [22]. Our method successfully predicts it to be the most mobile region in the protein except the N-terminal loop. However there is an disagreement between prediction and experiment on Asp-46 (S^2_{exp} : 0.58, S^2_{cal} : 0.84) which is not fitted well with any models in the NMR experiment [22]. The correlation coefficient increases to 0.78 (from 0.74) if the data of residue Asp-46, which is less reliable, is removed.

Frendicin ACP [23] is comprised of a three-helix bundle structure and have a high correlation coefficient between prediction and experiment ($r = 0.81$). The average value of the order parameters of the three helices is 0.844, which is consistent with our prediction that they have high S^2 values. We also correctly predict the C-terminal residues and the long loop (Gly-17 to Asp-23) connecting two helices have the first and second lowest average S^2 values respectively (0.358 and 0.492).

A recent study by Halle [24] showed that B factors (or atomic mean-square displacements) are inversely proportional to the number of noncovalent neighbor atoms within a certain cutoff radius. The main differences between Halle's

approach and the WCN model are: (1) the former assumes that every neighboring atom contributes equally, while, in the latter, the contribution of each atom is scaled down by its squared distance from the central atom; (2) and consequently, Halle's model needs to determine an optimal cutoff distance, while the WCN does not need one (Eq. 2).

Our results show that the backbone dynamics of protein structures can be directly inferred from the static structural properties without the assumption of any mechanical models. Since it is possible to compute quite accurate order parameters directly from the structural properties of proteins, our study underscores a direct link between protein topological structure and its dynamics. In addition, since the WCN model uses only Cα atoms, our results indicate that protein dynamics (such as the order parameters) can be determined without the knowledge of protein sequences. As increasing numbers of protein structures are solved in recent years, our method offers an efficient way to determine backbone motions with high accuracy and is practical in the study of protein function-dynamics relationship and structural genomics.

Acknowledgments This research was supported by National Science Council and ATU from Ministry of Education, Taiwan, ROC.

References

1. Yang LW, Bahar I (2005) Coupling between catalytic site and collective dynamics: a requirement for mechanochemical activity of enzymes. Structure 13:893–904
2. Mukherjee M, Dutta K, White MA, Cowburn D, Fox RO (2006) NMR solution structure and backbone dynamics of domain III of the E protein of tick-borne Langat flavivirus suggests a potential site for molecular recognition. Protein Sci 15:1342–1355
3. Warshel A (2002) Molecular dynamics simulations of biological reactions. Acc Chem Res 35:385–395
4. Berman H, Henrick K, Nakamura H, Markley JL (2007) The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. Nucleic Acids Res 35:D301–303
5. Levitt M, Warshel A (1975) Computer simulation of protein folding. Nature 253:694–698
6. McCammon JA, Gelin BR, Karplus M (1977) Dynamics of folded proteins. Nature 267:585–590
7. Warshel A (1976) Bicycle-pedal model for the first step in the vision process. Nature 260:679–683
8. Warshel A, Parson WW (2001) Dynamics of biochemical and biophysical reactions: insight from computer simulations. Q Rev Biophys 34:563–679
9. Karplus M, McCammon JA (2002) Molecular dynamics simulations of biomolecules. Nat Struct Biol 9:646–652
10. Showalter SA, Bruschweiler R (2007) Validation of molecular dynamics simulations of biomolecules using NMR spin relaxation as benchmarks: application to the AMBER99SB force field. J Chem Theory Comput 3:961–975
11. Rueda M, Ferrer-Costa C, Meyer T, Perez A, Camps J, Hospital A, Gelpi JL, Orozco M (2007) A consensus view of protein dynamics. Proc Natl Acad Sci USA 104:796–801
12. Zhang F, Bruschweiler R (2002) Contact model for the prediction of NMR N–H order parameters in globular proteins. J Am Chem Soc 124:12654–12655
13. Tirion MM (1996) Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. Phys Rev Lett 77:1905–1908
14. Bahar I, Atilgan AR, Erman B (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Fold Des 2:173–181
15. Ming D, Bruschweiler R (2006) Reorientational contact-weighted elastic network model for the prediction of protein dynamics: comparison with NMR relaxation. Biophys J 90:3382–3388
16. Pfeiffer S, Fushman D, Cowburn D (2001) Simulated and NMR-derived backbone dynamics of a protein with significant flexibility: a comparison of spectral densities for the betaARK1 PH domain. J Am Chem Soc 123:3021–3036
17. Tjandra N, Feller SE, Pastor RW, Bax A (1995) Rotational diffusion anisotropy of human ubiquitin from 15N NMR relaxation. J Am Chem Soc 117:12562–12566
18. Buck M, Boyd J, Redfield C, MacKenzie DA, Jeenes DJ, Archer DB, Dobson CM (1995) Structural determinants of protein dynamics: analysis of 15N NMR relaxation measurements for main-chain and side-chain nuclei of hen egg white lysozyme. Biochemistry 34:4041–4055
19. Harata K, Muraki M (1997) X-ray structure of turkey-egg lysozyme complex with tri-N-acetylchitotriose. Lack of binding ability at subsite A. Acta Crystallogr D Biol Crystallogr 53:650–657
20. Akke M, Skelton NJ, Kordel J, Palmer AG 3rd, Chazin WJ (1993) Effects of ion binding on the backbone dynamics of calbindin D9k determined by 15N NMR relaxation. Biochemistry 32:9832–9844
21. Svensson LA, Thulin E, Forsen S (1992) Proline cis-trans isomers in calbindin D9k observed by X-ray crystallography. J Mol Biol 223:601–606
22. Feng W, Tejero R, Zimmerman DE, Inouye M, Montelione GT (1998) Solution NMR structure and backbone dynamics of the major cold-shock protein (CspA) from Escherichia coli: evidence for conformational dynamics in the single-stranded RNA-binding site. Biochemistry 37:10881–10896
23. Li Q, Khosla C, Puglisi JD, Liu CW (2003) Solution structure and backbone dynamics of the holo form of the frenolicin acyl carrier protein. Biochemistry 42:4648–4657
24. Halle B (2002) Flexibility and packing in proteins. Proc Natl Acad Sci USA 99:1274–1279
25. Kristensen SM, Siegal G, Sankar A, Driscoll PC (2000) Backbone dynamics of the C-terminal SH2 domain of the p85[alpha] subunit of phosphoinositide 3-kinase: effect of phosphotyrosine-peptide binding and characterization of slow conformational exchange processes. J Mol Biol 299:771–788
26. Yun S, Jang DS, Kim DH, Choi KY, Lee HC (2001) 15N NMR relaxation studies of backbone dynamics in free and steroid-bound Delta 5-3-ketosteroid isomerase from Pseudomonas testosteroni. Biochemistry 40:3967–3973
27. Stivers JT, Abeygunawardana C, Mildvan AS (1996) 15N NMR relaxation studies of free and inhibitor-bound 4-oxalocrotonate tautomerase: backbone dynamics and entropy changes of an enzyme upon inhibitor binding. Biochemistry 35:16036–16047
28. Redfield C, Boyd J, Smith LJ, Smith RA, Dobson CM (1992) Loop mobility in a four-helix-bundle protein: 15N NMR relaxation measurements on human interleukin-4. Biochemistry 31:10431–10437