Romanian Reports in Physics, Vol. 64, No. 1, P. 116–126, 2012

PROTEINS SURFACE ROUGHNESS ANALYSIS. COMPARISON OF CRYSTALLOGRAPHIC AND NMR STRUCTURES

D. CRACIUN[1], L. PITULICE[2,3], A. CIORSAC[4], V. OSTAFE[2,3], A. ISVORAN[2,3]*

[1]West University of Timisoara, Teacher Training Department, 4 Blvd. V. Parvan, Timisoara, 300223, Romania
[2]West University of Timisoara, Department of Chemistry, 16 Pestalozzi, Timisoara, 300115, Romania
E-mail: aisvoran@cbg.uvt.ro
[3]Nicholas Georgescu-Roegen Interdisciplinary Research and Formation Platform, 16 Pestalozzi, Timisoara, 300115, Romania
[4]Politehnica University of Timisoara, Department of Physical Education and Sport, 2 P-ta Victoriei, 300006, Timisoara, Romania

Received November 3, 2010

Abstract. Within this study we compare the surface roughness for 15 different proteins using X-ray crystallographic and NMR structures, respectively. We have noticed higher surface fractal dimensions for NMR structures with a mean value of 2.36 ± 0.02 in comparison to the crystallographic ones, where the mean value was 2.20 ± 0.01. As the protein surface is a result of its three-dimensional folding, we have also analyzed the dependence of the surface fractal dimension on the protein radius of gyration and its molecular weight and we observed only a weak dependence. Also, our study reflects that the radius of gyration for every protein usually has a higher value when it is determined using the NMR file than using the crystallographic one, illustrating that the root mean square distance between monomers is higher for the solution structure reflecting its smaller degree of compactness. The latter is also reflected by the smaller correlation exponent between the radius of gyration and the molecular weight obtained for NMR structures (C = 1.15) in comparison with X-ray structures (C = 1.28). The advantages of using NMR data for studying protein roughness are also discussed.

Key words: proteins, surface roughness, fractal surface, radius of gyration.

1. INTRODUCTION

Proteins interact with a great variety of substrates and the corresponding binding sites are situated on the protein surface, thus protein surface interactions are very important for natural processes. Despite this importance and their great potential for applications ranging from nanotechnology to medicine, the basic

* Paper presented at the National Conference on Physics, September 23–25, 2010, Iaşi, Romania.

mechanisms determining affinity and specificity are still poorly understood [1] and there is a need for quantification of the protein surface properties. These properties are dictated by the relative position and specific features of exposed residues, being modulated by a variety of factors (hydrogen bonding, hydrophobicity, electrostatic interactions, etc) but sharing a unique feature, hydration [2] which is intimately related to the protein fluctuations [3]. Nuclear magnetic resonance (NMR) spectroscopy studies are used to analyze protein surface hydration and they have been revealed water residence times at protein surfaces within the subnanosecond regime [4].

For the protein surface interactions its surface roughness is an important property and it can be quantified by the fractal dimension [5]. Surface roughness means the presence of some elevations and depressions in a small surface area. A higher roughness indicates a much more complexity of the surface shape that may be correlated to specificity of the interactions [6]. Also, a higher roughness allow a lot of van der Waals contacts to occur between the protein and its ligand, therefore the rougher is the binding site the stronger is the interaction meaning that roughness contributes to binding.

Protein surfaces have been shown to be fractal with different fractal dimensions at macro- and micro-scale [5–14]. As we know, all these results have been obtained using crystal structures. Crystal structures can be affected by inter-molecular packing or crystal lattice packing, but in solution the molecule can adopt many conformations. Also, soluble proteins perform their function in solution and it is important to obtain detailed information concerning their structure in solution state. During the past decades, the NMR of proteins in solution demonstrated their potential for studies regarding both structure and function. Also NMR-based techniques may be used to observe the surface characteristics of a protein without the possible artifacts caused by molecular packing in the crystal [15] and also can inform about hydrodynamic properties of protein molecules in solution [16]. Thus, we focus our study on the differences between the surface roughness of proteins based on solution NMR structures in comparison to crystal ones.

NMR spectroscopy and X-ray crystallography are complementary methods that are routinely used to study protein structures and dynamics. The two methods share some similarities: both determine time and spatial averages of the protein atoms coordinates and the averages values are obtained from measurements performed on an ensemble of molecules [17]. Beyond these similarities, there are also significant differences between crystal and solution structures. A crystallographic structure is directly determined from the experimental data and, on the contrary, a NMR structure is determined indirectly by combining NMR experimental data with traditionally geometrical potential energy functions. It

means that X-ray data represent a single optimal structure but NMR can be represented by an ensemble of structures. Also, there are differences in terms of spatial distribution of the molecule in the samples and the time scales accessible to each method [18].

From the similarity point of view between the solution and crystal structures, there might be three situations [19]: (i) the NMR and X-ray structures are coincident, (ii) they are somewhat different and (iii) they are totally distinct. The case of somewhat different solution and crystal structures means that they exhibit the same fold, but local functional differences may be attributed to dynamics. Some examples illustrating this case are the calmodulin, in which the EF hands motifs of the N-terminal domain are considerably less open in the solution structure than in crystallographic one [20], and the maltodextrin binding protein (MBD) that shows in the solution structure a domain closure by comparison to the crystal structure [21]. The case of totally distinct crystal and NMR structures has as example the same protein, calmodulin, having four calcium ions bound, the crystallographic structure (PDB entry code 1CLL) suggests a "dumb-bell" structure with an extended alpha helix and the NMR structure (1CDM) indicates an unstructured and dynamic central region. A flexible interdomain linker has also been demonstrated by NMR structure of xylanase in comparison to its X-ray structure [22] and for matrix metalloproteinase 12, those domains in solution show conformational freedom in respect to each other by comparison to the crystal structure revealing a well-defined relationship between domains [23]. Comparison analysis of the crystal and solution structures of the complexes made by calmodulin with target peptides reveals that the significant differences between the crystal and solution structures may come from the loss of some hydrogen bonds in solution by comparison to crystal structure [19].

Although differences may appear between crystal and solution structures, NMR spectroscopy has a number of advantages in studying protein dynamics because it may reflect the different degrees of flexibility of distinct parts of the protein molecule, provides local, structural and conformational variability information about macromolecules and regarding the protein folding, it allows to follow the time dependence of structure formation [18].

The aim of the study is to analyze the differences between the surface fractal dimensions of proteins obtained using crystal and solution structural data. We take into consideration 15 different proteins having solved the structures both by X-ray crystallography and NMR spectroscopy and we compare the global surface fractal dimensions for each case.

2. METHOD

For this study we have chosen 15 different proteins having solved their 3D structures by both X-ray crystallography and NMR spectroscopy methods. The structural data necessary for the calculations have been retrieved from the Protein Data Bank [24] and the PDB codes entry are presented in Table 1 for both X-ray and NMR files.

For each protein we have calculated the global surface fractal dimension (D_s) and we have appreciated the degree of compactness of crystallographic structures in comparison to NMR ones. Conventional approach to study fractal dimension of protein surface is largely described in literature, as for example in [10], here we only make a short presentation of the equations used in our calculations. The method uses the scaling law between the surface area (SA) of the protein and the radius of the rolling probe molecule (R)

$$\mathrm{SA} \sim R^{2-D_s} \tag{1}$$

and the fractal dimension is determined from the slope of the double logarithmical plot of SA *versus* R. The surface area of the protein has been calculated using the GETAREA on-line free software (URL: http://curie.utmb.edu/ getarea.html) where probe radii of 1, 1.2, 1.4, 1.6, 1.8 and 2 Å respectively were used [25].

In order to appreciate the degree of structural compactness, the scaling law between the radius of gyration and molecular weight has been used [10]

$$M \sim R_g^{\nu}. \tag{2}$$

The correlation coefficient reflecting the degree of structural compactness is obtained as the inverse of the slope of the linear fitting of logR_g *versus* log$M$ and it is a measure of the packing density of the protein. For every protein the molecular weight has been calculated using ProtParam on-line tool [26] and the radius of gyration for both X-ray and NMR structural files has been calculated using the on-line facilities provided by the Protein Dipole Moments Server [27] under Weizman Institute web page (URL: http://bioportal.weizmann.ac.il/ dipol/dipol2.html).

3. RESULTS AND DISCUSSIONS

To illustrate the surface fractal dimension determination, the log(SA) *versus* log$R$ plot for chicken lysozyme (PDB code entry 2ZQ4) is presented in Fig. 1. The linear fitting of the points presented in Fig. 1 gives the surface fractal dimension in this case, $D_s = 2.15 \pm 0.02$.

The PDB codes entry for the structural files and the values obtained for the investigated parameters for all proteins are presented in Table 1.

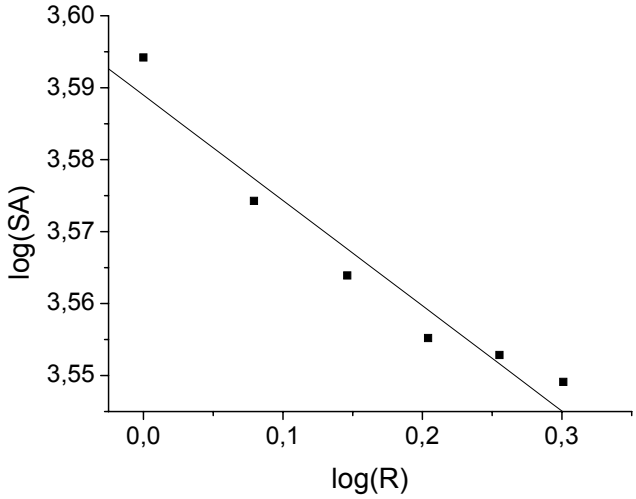


Fig. 1 – Determination of the surface fractal dimension for the X-ray structural file of chicken lysozyme (PDB code 2 ZQ4).

Table 1

The values of the investigated parameters for considered proteins

Nr crt	Protein	Molecular Weight (Da)	PDB code of the X-ray file	Surface fractal dim. for the X-ray file	Radius of gyration for the X-ray file ($Å^2$)	PDB code of the NMR file	Surface fractal dim. for the NMR-ray file	Radius of gyration for the NMR file ($Å^2$)
1	Human calmodulin	16706.3	1CLL	2.19±0.01	246.50	2K0E	2.28±0.03	219.61
2	Chicken troponin C	18245.2	4TNC	2.10±0.01	268.44	1TNW	2.33±0.05	304.11
3	Chicken parvalbumine	11726.2	3FS7	2.20±0.05	70.08	2KQY	2.36±0.08	80.79
4	Bovine recoverin	23202.2	1OMR	2.29±0.05	238.29	1LA3	2.45±0.04	244.22
5	Human parvalbumine	11889.7	1PVA	2.26±0.01	120.99	3PAT	2.34±0.04	120.09
6	Human interleukine 6	21112.1	1ALU	2,32±0.07	186.41	1IL6	2.57±0.02	212.63
7	Human interleukine 8	8385.7	3IL8	2.13±0.01	96.23	1IL8	2.52±0.06	190.77
8	Bacteriophage T4 glutaredoxine	10049.6	1AAZ	2.09±0.02	113.22	1DE1	2.26±0.03	125.77
9	Acathamoeba profiling	12952.5	1PRQ	2.14±0.02	131.10	2PRF	2.39±0.06	149.43
10	Sperm whale myoglobin	17199.9	1VXA	2.19±0.01	177.81	1MYF	2.24±0.02	190.00

Table 1 (continued)

11	E Coli maltodextrin binding protein	40707.3	4MBP	2.31±0.01	372.45	2KLF	2.59±0.02	441.43
12	Chicken lysozyme	14313.1	2ZQ4	2.15±0.02	146.84	1GXV	2.18±0.03	156.81
13	Mycobacterium tuberculosis adenylate kinase	22288.2	2CDN	2.22±0.01	200.77	1P4S	2.42±0.01	227.40
14	SH3 domain of chicken alpha-spectrin	7219.2	1U06	2.13±0.02	72.65	1M8M	2.19±0.02	84.75
15	Anabaena flavodoxine	18745.5	1FTG	2.31±0.03	177.30	2KQU	2.20±0.01	238.37

The values obtained for the surface fractal dimensions of the proteins, between 2.10 and 2.57, are in good agreement with other published data [6, 10–14, 16] and reveal that the concepts of fractal geometry are applicable to describe quantitatively the surface roughness of proteins.

Student and one-way ANOVA statistical tests performed under ORIGIN8.0 shows that, at the 0.05 level, the mean value for the surface fractal dimensions obtained using X-ray crystallographic files, 2.20 ± 0.01, significantly differs from the mean value of the surface fractal dimensions obtained using the NMR files, 2.36 ± 0.02 reflecting the higher roughness of NMR structures. As the surface roughness is mainly responsible for protein-protein and protein-ligand interactions due to its contribution to the specificity of these interactions [6], our result underlines that NMR structures are better candidates than X-ray ones for molecular modeling and drug design studies. This observation is also sustained by the fact that NMR structures allow to take into account other important factors regulating the protein interactions, such as the hydration level of the protein and its flexibility.

From Table 1 we also notice usually higher values of radii of gyrations obtained for NMR structures reflecting the distinct spatial distribution of monomers with a higher root mean square distance between them, and also a smaller degree of compactness of solution structures in comparison to X-ray crystals.

As the protein surface is a result of the three-dimensional folding, we may expect that surface enlargement due to the rise of amino acid number (and protein molecular weight) to be accompanied by the increase of surface irregularities and consequently to find some correlation between the surface fractal dimension and the protein radius of gyration and its molecular weight. We could notice only a weak dependence of the surface fractal dimension on radius of gyration (correlation coefficient R=0.60±0.06) and on molecular weight (correlation coefficient R=0.63±0.07), as it is presented in Fig. 2 for the calculations based on the X-ray structural files and Fig. 3, respectively. The result is similar for the

calculations based on NMR files, the correlation coefficients being R=0.64±0.06 for the dependence of the surface fractal dimension on radius of gyration. These results suggest that the number of amino acids in sequence has a weak influence on the protein folding process, the physical and chemical properties of amino acids being more important for this process than their number.

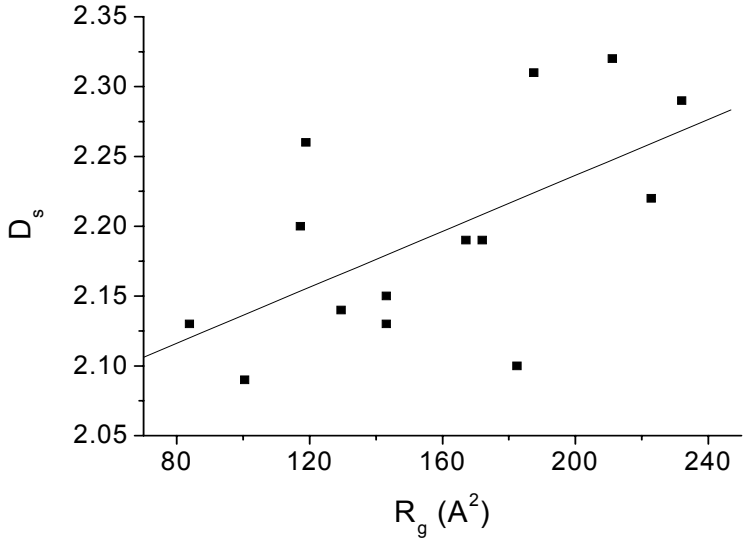


Fig. 2 – Surface fractal dimension *versus* radius of gyration for the X-ray structural files.

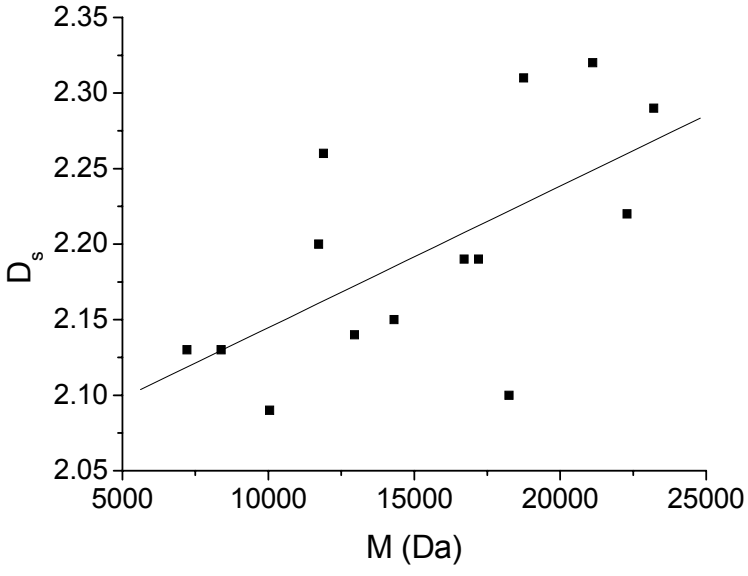
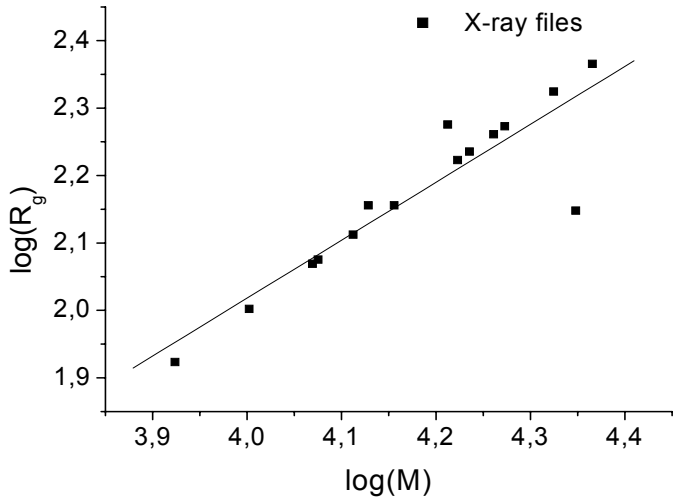
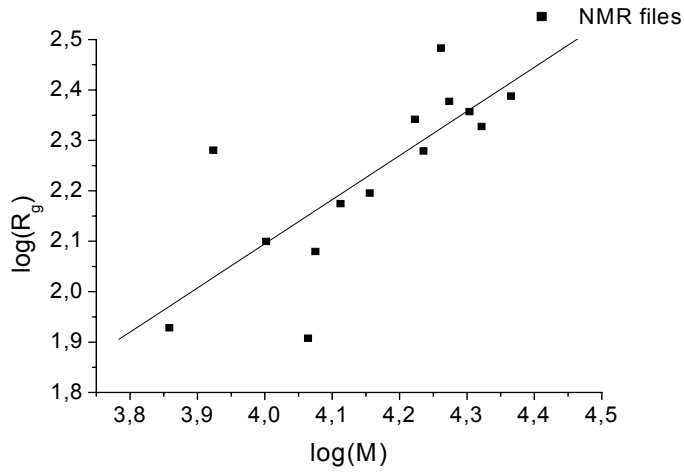


Fig. 3 – Surface fractal dimension *versus* molecular weight.

Our data show a linear dependence in the double logarithmical plot of the radius of gyration versus molecular weight (Figs. 4) and this dependence allows us to compare the packing density for the X-ray and NMR files.



a



b

Fig. 4 – Double logarithmical plot of radius of gyration *versus* molecular weight for the X-ray files (a) and NMR files (b).

The linear fitting of the points in Fig. 4 allows to obtain the correlation exponent between the radius of gyration and molecular weight and it is smaller for NMR structures (C=1.15) in comparison with X-ray crystals (C=1.28) reflecting a

higher packing density for X-ray structures in comparison to NMR ones. The lower packing density of NMR structures could be related to the spatial distribution of monomers and to the protein hydration, a property very important for many natural processes. From the other point of view, the higher packing density reflected by X-ray structures could also be due to the effect of crystal lattice packing being in this case only an artifact.

4. CONCLUSIONS

Within this study, two different approaches to analyze protein surface fractality are considered: structural files obtained by X-ray crystallography and structural files obtained by NMR spectroscopy, respectively. Our study confirms that fractal analysis of protein surfaces allows to take into account the surface heterogeneity and to quantitatively characterize it using the fractal dimension. It also suggests that this quantity has a week dependence on protein molecular weight and radius of gyration but it strongly depends on the physical state of the protein being always higher for protein in solution in comparison to crystallized protein. This result is in good agreement with other studies reflecting that surface fractal dimension depends on the properties of the media, such as pH, electrolyte concentration and temperature [6, 28].

The higher fractality of protein surfaces reflected by NMR structures can reveal some important properties of protein surface regarding its interactions with small molecules (solvent, ligands) and/or other macromolecules. From this point of view, it is already known that protein surface topography dominates the physics and chemistry of water binding, a protein with higher surface roughness showing a higher number of bounded water molecules [29]. This interdependence of the protein surface shape and hydration level has general implications for modeling, prediction of protein surface shape, molecular recognition, local folding and solvent binding. It is also known that the fractality of the protein surface has significant influence on many biochemical and biophysical processes [30] such as protein aggregation, diffusion of small molecules along the protein surface with direct consequences on the selectivity of chemical reactions. Although limited in terms of size of the molecules that can be studied at present, NMR structural files offer some advantages over X-ray crystallography in studies of surface shape and roughness in correlation with the protein interactions and they may be successfully used for computational approaches regarding molecular modeling.

Tacking into account all these observations we may conclude that the use of NMR structural files for surface fractality analysis can improve the understanding of molecular biology problems and that, in addition to and X-ray crystallography, NMR spectroscopy is a complementary method for probing a wide range of structural properties of macromolecules.

REFERENCES

1. O. Cohavi, S. Corni, F. De Rienzo, R. Di Felice, K. E. Gottschalk, M. Hoefling, D. Kokh, E. Molinari, G. Schreiber, A. Vaskevich, R. C. Wade, *Protein–surface interactions: challenging experiments and computations*, Journal of Molecular Recognition, **23**, *3*, 259–262 (2010).
2. N. Niccolai, A. Ciutti, O. Spiga, M. Scarselli, A. Bernini, L. Bracci, D. Di Maro, C. Dalvit, H. Molinari, G. Esposito, P. A. Temussi, *NMR Studies of Protein Surface Accessibility*, The Journal of Biological Chemistry, **276**, *45*, 42455–42461 (2001).
3. L. Zhang, L. Wang, Ya-T. Kao, W. Qiu, Y. Yang, O. Okobiah, D. Zhong, *Mapping hydration dynamics around a protein surface*, PNAS, **104**, *47*, 18461–18466 (2007).
4. G. Otting, E. Liepinsh, K. Wuthrich, *Protein hydration in aqueous solution*, Science, **254**, 974–980 (1991).
5. M. Lewis, D.C. Rees, *Fractal surface of proteins*, Science, **230**, 1163-1165 (1985).
6. F.K. Pettit, J.U. Bowie, *Protein surface roughness and small molecular binding sites*, J Mol Biol., **285**, *4*, 1377–82 (1999).
7. P. Pfeifer, U. Weltz, H. Wipperman, *Fractal surface dimension of proteins: lysozyme*, Chem Phys Letters, **113**, 535–540 (1985).
8. T. Goetze, J. Brickmann, *Self similarity of protein surfaces*, Biophys. J., **61**, 109–118 (1992).
9. B.A. Feodorov, B.B. Feodorov, W. Schmidt, *An analysis of fractal properties of the surfaces of globular proteins*, J. Chem Phys., **99**, 4078–4083 (1993).
10. T.G Dewey, *Fractals in molecular biophysics*, Oxford University Press: Oxford, New York. 1997
11. E.W. Stawiski, A.E. Baucom, S.C. Lohr, L. M. Gregoret, *Predicting protein function from structure: Unique structural features of proteases*, PNAS, **97**, 3954–3958 (2000).
12. E.W. Stawiski, Y. Mandel-Goutfreund, A.C. Lowenthal, L.M. Gregoret, *Progress in predicting protein structure from sequence: unique features of O-glycosidases,* Pacific Simposium on Biocomputing, **7**, 637–648 (2002).
13. L Pitulice, A. Isvoran, A. Chiriac, *Structural features of proteins reflected by structural scaling laws,* Journal of Serbian Chemical Society, **73**, 805–813 (2008).
14. L. Pitulice, A. Isvoran, C.T. Craescu, A. Chiriac, *Scaling properties of radius of gyration and surface area for EF hand calcium binding proteins,* Chaos Solitons and Fractals, **40**, 684–690 (2009).
15. N. Niccolai, R. Spadaccini, M. Scarselli, A. Bernini, O. Crescenzi, O. Spiga, A. Ciutti, D. Di Maro, L. Bracci, C. Dalvit, A.T. Piero, *Probing the surface of a sweet protein: NMR study of MNEI with a paramagnetic probe,* Protein Sci., **10**, *8*,1498–1507 (2001).
16. D.A. Fushman, *Surface fractality of proteins from theory and NMR data*, J. Biomol. Struct. & Dynamics, **7** (1990).
17. D. Morikis, J.D. Lambris, *Physical methods for structure, dynamics and binding in immunological research*, Trends Immunol, **12**, 700–707 (2004).
18. A.T. Brunger, *X-ray crystallography and NMR reveal complementary views of structure and dynamics*, Nat Struct Biol, **4**, 862–865 (1997).
19. I. Bertini, P. Kursula, C. Luchinat, G. Parigi, J. Vahokoski, M. Wilmanns, J. Yuan, *Accurate solution structures of proteins from X-ray data and minimal set of NMR data: calmodulin-peptide complexes as examples*, J. Am. Chem. Soc, **131**, 5134–5144 (2009).
20. J.J. Chou, S. Li, C.B. Klee, A. Bax, *Solution structure of Ca2+-calmodulin reveals flexible hand-like properties of its domains,* Nat. Struct. Biol., **8,** 990–997 (2001).
21. N.R. Skrynnikov, N.K. Goto, D. Yang, W.Y. Choy, J.R. Tolman, G.A. Mueller, L.E. Kay, *Orienting domains in proteins using dipolar couplings measured by liquid state NMR: Differences in solution and crystal forms of maltodextrin binding protein loaded with β-cyclodextrin*, J.Mol.Biol., **295**, 1265–1273 (2000).

22. D.K.Y. Poon, S.G. Withers, L.P. McIntosh, *Direct demonstration of the flexibility of the glycosylated proline-threonine linker in Cellulomonas fimi xylanase Cex through NMR spectroscopic analysis,* J. Biol. Chem., **282**, 2091–2100 (2007).
23. I. Bertini, V. Calderone, M. Fragai, R. Jaiswal, C. Luchinat, M. Melikian, E. Mylonas, D.I. Svergun, *Evidence of reciprocal reorientation of the catalytic and hemopexin-like domains of full-length MMP-12*, J Am Chem Soc., **130**, *22*, 7011–21 (2008).
24. H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, *The Protein Data Bank*, Nucl. Acids Res., **28,** 235–242 (2000).
25. R. Fraczkiewicz and W. Braun, *A new efficient algorithm for calculating solvent accessible surface areas of macromolecules, Journal of Computational Chemistry*, **19**, 319–333 (1998).
26. E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch, *Protein Identification and Analysis Tools on the ExPASy Server, The Proteomics Protocols Handbook*, Humana Press, 2005, pp. 571–607.
27. C. E. Felder, J. Prilusky, I. Silman, J. L. Sussman, *A server and database for dipole moments of proteins*, Nucleic Acids Research, **35**, Special Web Servers Issue, 2007.
28. R. Vreeker, L.L. Hoekstra, D.C. den Boer, W.G.M. Agterol, *Fractal aggregation of whey protein,* Food Hydrocolloids, **6**, 423–435 (1992).
29. L.A. Kuhn, M.A. Siani, M. E. Pique, C. L. Fisher, E.D. Getzoff, J.A. Tainer, *The interdependence of protein surface topography and bound water molecules revealed by surface accessibility and fractal density measures,* Journal of Molecular Biology, **228**, 13–22 (1992).
30. T. Lahiri, H. Mishra, S. Sarkar, K. Misra, *Surface characterization of proteins using multi-fractal property of heat-denatured aggregates*, Bioinformation, **2**, *9*, 379–383 (2008).