Correlation between local structural dynamics of proteins inferred from NMR ensembles and evolutionary dynamics of homologues of known structure.

Swapnil Mahajan[1, 2, 3], Alexandre G. de Brevern[3, 4, 5, 6], Bernard Offmann[7, 8] and Narayanaswamy Srinivasan[9,*]

[1]Université de La Réunion, Faculté des Sciences et Technologies, F-97715 Saint Denis Messag Cedex 09, La Réunion, France.

[2]INSERM, UMR-S S665, DSIMB, F-97715 Saint Denis Messag Cedex 09, La Réunion, France.

[3]Laboratoire d'Excellence GR-Ex, F-75739 Paris, France.

[4]INSERM, U665, DSIMB, F-75739 Paris, France;

[5]Univ Paris Diderot, Sorbonne Paris Cité, UMR-S665, F-75739 Paris, France;

[6]Institut National de la Transfusion Sanguine, F-75739 Paris, France;

[7]Université de Nantes, F-44322 Nantes Cedex 03, France.

[8]UFIP CNRS FRE 3478, F-44322 Nantes Cedex 03, France.

[9]Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India.

*Corresponding author: E-mail: ns@mbu.iisc.ernet.in

Running head: Inherent and evolutionary structural dynamics

Abstract

Conformational changes in proteins are extremely important for their biochemical functions. Correlation between inherent conformational variations in a protein and conformational differences in its homologues of known structure is still unclear. In this study we have used a structural alphabet called Protein Blocks (PBs). PBs is used to perform abstraction of protein 3-D structures into a 1-D strings of 16 alphabets (*a* to *p*) based on dihedral angles of overlapping pentapeptides. We have analyzed the variations in local conformations in terms of PBs represented in the ensembles of 801 protein structures determined using NMR spectroscopy. In the analysis of concatenated data over all the residues in all the NMR ensembles, we observe that the overall nature of inherent local structural variations in NMR ensembles is similar to nature of local structural differences in homologous proteins with a high correlation coefficient of 0.94. High correlation at the alignment positions corresponding to helical and β-sheet regions is only expected. However, the correlation coefficient by considering only the loop regions is also quite high (0.91). Surprisingly, segregated position-wise analysis shows that this high correlation does not hold true to loop regions at the structurally equivalent positions in NMR ensembles and their homologues of known structure. This suggests that the general nature of local structural changes is unique; however most of the local structural variations in loop regions of NMR ensembles do not correlate to their local structural differences at structurally equivalent positions in homologues.

Supplementary Material

Supplementary material for this paper is available online at http://www.bo-protscience.fr/Supplementary_table_S1.pdf

Key words: Protein blocks, protein structures, conformational changes, structurally equivalent positions, loop regions.

Introduction

In recent years protein structural dynamics is gaining importance as the fourth dimension in structure-function paradigm of structural biology (Henzler-Wildman & Kern, 2007; Teilum, Olsen, & Kragelund, 2009; Zhuravlev, Materese, & Papoian, 2009). Recent studies show the importance of conformational dynamics in protein function, mainly in enzyme catalysis (Bhabha et al., 2011; Doshi, McGowan, Ladani, & Hamelberg, 2012) and molecular recognition (Lange et al., 2008). Due to dynamic nature of proteins, a native state of protein is best represented by ensemble of conformations than a single structure (DePristo, De Bakker, & Blundell, 2004; Friedland, Lakomek, Griesinger, Meiler, & Kortemme, 2009; Furnham, Blundell, DePristo, & Terwilliger, 2006; Larson, England, Desjarlais, & Pande, 2002). NMR has been extensively used to study dynamics in proteins (Ishima & Torchia, 2000; Sze & Lai, 2011; Tzeng & Kalodimos, 2011) and it also provides users with an ensemble of conformations for a single protein. NMR spectroscopy can provide us with information of conformational variation in a protein structure on pico- to milliseconds timescale (Salmon, Bouvignies, Markwick, & Blackledge, 2011).

The correspondence between the native state conformational ensembles and the conformational heterogeneity observed in crystal structures have been studied extensively (Best, Lindorff-Larsen, DePristo, & Vendruscolo, 2006; Friedland et al., 2009) and used in modeling of side-chain conformational variations, structural differences on mutation and protein design (Friedland, Linares, Smith, & Kortemme, 2008; Keedy et al., 2012; Smith & Kortemme, 2008). Recently few studies also have shown that sequence diversity in a protein family is related to molecular dynamics observed in a single protein from that protein family (Friedland et al., 2009; Liu & Bahar, 2012). Friedland *et al* (Friedland et al., 2009) have shown that the regions of high conformational variations in the sampled conformations of ubiquitin by Backrub motions (Davis, Arendall, Richardson, & Richardson, 2006) are similar to the regions of high conformational differences observed in homologues of ubiquitin. But this study by Friedland *et al* does not address the question of the correlation between the conformational variations observed in Backrub ensemble of ubiquitin and the corresponding conformational differences in the homologues of ubiquitin.

In this study we have compared the local conformational variations observed in the NMR ensemble of a single protein and local conformational differences observed in different homologues of known structure of that protein. We have not used the classical root-mean-squared-deviation (RMSD) to identify the changes in local backbone conformations because RMSD could be misleading when comparing local backbone conformations. Low RMSD value between two local backbone structures does not always suggest that these two conformations are similar and *vice versa* (Agarwal, Mahajan, Srinivasan, & De Brevern, 2011). For this reason we have used a structural alphabet, Protein Blocks (PBs), to characterize and compare local backbone conformations in protein structures. PBs is a classification of overlapping pentapeptide

fragments into 16 different classes, named *a* to *p*, based on backbone dihedral angles (φ and ψ) (De Brevern, Etchebest, & Hazout, 2000). PBs have been used previously for many different applications including protein structural alignment and protein structure prediction (Etchebest, Benros, Hazout, & De Brevern, 2005; Joseph et al., 2010; Tyagi, Gowri, Srinivasan, de Brevern, & Offmann, 2006).

Materials and Methods

Dataset

A dataset of 2180 NMR structures was derived from PDB (Protein Data Bank) (Berman et al., 2000). All these NMR structures are largely non-homologous (30% sequence similarity cutoff), have more than one model and are at least 50 amino acids long. These 2180 structures have only one chain and are not bound to ligands.

Restraints information is a widely used measure for the quality of protein structures solved using NMR spectroscopy (Clore, Robien, & Gronenborn, 1993; Laskowski, Rullmannn, MacArthur, Kaptein, & Thornton, 1996). Therefore we filtered these 2180 NMR entries according to 'restraints per residue' and 'restraints per atom' to ensure the good quality of NMR structures. Restraints data for NMR structures was obtained from BMRB (Biological Magnetic Resonance Data Bank) (Ulrich et al., 2007). Out of 2180 entries, 801 NMR structures with at least 15 restraints per residue and with at least 75% of the residues having ≥ 0.5 restraints per atom were selected. These 801 NMR structures were used to construct global PB variation matrix.

Protein Blocks

Protein blocks (PBs) (De Brevern et al., 2000; Joseph et al., 2010) are a structural alphabet, a library of 16 local backbone structural prototypes which can be used for abstraction of three-dimensional protein structures into one-dimensional sequences of alphabets *a* to *p* (Offmann, Tyagi, & De Brevern, 2007). PBs are assigned on the basis of φ and ψ dihedral angles of overlapping windows of 5 amino acids. PBs *d* and *m* correspond to central β-strand and α-helix conformations, respectively. PBs *a* to *c* correspond to N-terminus of β-strand and PBs *e* and *f* correspond to C-terminus of β-strand. Similarly, PBs *k* and *l* correspond to N-terminus of α-helix and PBs *n* to *p* correspond to C-terminus of α-helix; whereas PBs *g* to *j* are specific to loop regions. PBs were assigned using an in-house Perl script.

Concatenated data analysis of local conformational changes

PBs were assigned to each model of the dataset of 801 NMR structures. PB changes were counted at each position in all the models of NMR ensembles. We considered each position as structurally equivalent for concatenated data analysis because each position corresponds to the same amino acid residue in different models of an ensemble. These raw counts of PB changes were normalized by using the following formula adapted from Johnson and Overington (Johnson & Overington, 1993)

$$S_{x,y} = \log \left[\frac{N_{x,y} * \sum\limits_{x=1}^{M}\sum\limits_{y=1}^{M} N_{x,y}}{\sum\limits_{y=1}^{M} N_{x,y} * \sum\limits_{y=1}^{M} N_{x,y}} \right] \tag{1}$$

where, $N_{x,y}$ is the count of changes from PB x to PB y at the same position in NMR ensembles; M is total number of different PBs (*i.e.* 16) and $S_{x,y}$ is normalized score for PB variation between PB x to PB y. A 16x16 matrix was generated using normalized PB variation

scores ($S_{x,y}$). This normalized global PB variation matrix was compared to PB substitution matrix (Joseph, Srinivasan, & De Brevern, 2011) derived from structurally equivalent positions (Cα atoms within 3 Å distance) in homologous structures. PB substitution matrix was first generated and used by Tyagi *et al* for protein structural alignments (Tyagi et al., 2006). PB substitution matrix was improved by Joseph *et al* (Joseph et al., 2011) by using 1922 domain families from PALI version 2.8a database (Balaji, Sujatha, Kumar, & Srinivasan, 2001). Counting of PB changes and normalization of matrix is exactly same for global PB variation and PB substitution matrices.

Segregated position-wise analysis of local conformational changes

Out of the 801 protein structures, a refined dataset of 65 protein structures was constructed on the basis of defined SCOP (Structural Classification Of Proteins) (Murzin, Brenner, Hubbard, & Chothia, 1995) domains. NMR ensembles with at-least 7 models and at-least 7 distinct but homologous domains, , in the corresponding SCOP families were included for segregated position-wise data analysis.

This dataset of 65 NMR structures and their corresponding 1762 homologues of known structure was used for segregated position-wise analysis of local backbone changes; these homologues are taken as member of the same SCOP family. Most of these structures of homologues are solved by X-ray crystallography. The first model from each of these 65 NMR ensembles were structurally aligned with their homologues using MUSTANG software (Konagurthu, Whisstock, Stuckey, & Lesk, 2006) to find out the structurally equivalent positions (Cα atoms within 3 Å distance) between homologues and NMR ensembles. Structure-based multiple sequence alignments were converted to PB sequence alignments. Total of 1281 structurally equivalent positions excluding regular secondary structures were identified in these

structural alignments; positions with gaps were also not analyzed. At each of these structurally equivalent positions in loop regions, the frequencies of PB changes were calculated for NMR ensembles and their homologues separately. Pearson's correlation coefficients were calculated for the frequencies of PB changes at each of the 1281 structurally equivalent positions between NMR ensemble and their homologues. For each of these 1281 positions, average of all the pairwise PB variation scores in the NMR ensemble, namely 'average PB variation score', was calculated using global PB substitution matrix. Similarly, average of all the pairwise PB substitution scores at structurally equivalent positions of homologues, namely 'average PB substitution score', was also calculated using global substitution matrix for each of the 1281 positions.

PB variation and PB substitution matrices for loop regions were constructed by combining the frequencies of PB changes at these 1281 structurally equivalent positions from NMR ensembles and their homologues separately. Correlation coefficient between these two matrices for structurally equivalent positions in only loop regions was calculated. Similarly, PB variation and PB substitution matrices were generated for all 4348 structurally equivalent positions, including regular secondary structures, from NMR ensembles and their homologues separately. Correlation coefficient between these matrices corresponding to all the structurally equivalent positions was calculated. Same normalization (equation (1)) was used to construct these matrices. It is important to analyze all structurally equivalent positions and, separately, only those structurally equivalent positions present in loop regions. Indeed, regular secondary structures are known to be conserved better than loop regions. So correlation coefficients could be biased if the alignment positions corresponding to regular secondary structures are included..

Results and discussion

The main objective of this work is to address the correlation between inherent local structural dynamics inferred from NMR ensembles and evolutionary local structural dynamics of their homologues. Here, we have considered local structural changes by means of PB changes. Figure 1 shows different models from NMR structure of phospholipase A2 (Yuan, Byeon, Li, & Tsai, 1999) enzyme (PDB: 1BVM). This example underlines local backbone structural variations between different models of a NMR ensemble can be easily characterized and distinguished from regions with no local backbone structural variations thanks to PB assignments for these regions. Local structural variation analysis on large dataset of NMR structures can provide the general rules for allowed and disallowed conformational variations inherent to proteins irrespective of their family, superfamily or fold. To analyze such overall local structural variations in NMR ensembles we generated a global PB variation matrix (see Materials and Methods) in a similar way in some of the previous works (Joseph et al., 2011; Tyagi et al., 2006). Surprisingly, global PB variation matrix showed similar patterns for local structural changes as global PB substitution matrix (fig. 2) generated by Joseph *et al* [31] even if the datasets used for these matrices are completely different. Global PB variation matrix is based on non-homologous protein structures from NMR ensembles and global PB substitution matrix is based on only structurally equivalent positions from homologous proteins. Pearson's correlation coefficient between global PB variation and PB substitution matrices equals to 0.94. This high correlation suggests that the nature of PB variations in NMR ensembles and nature of PB substitutions in homologous proteins are overall similar.

We decided to further investigate the correlation between PB variations and PB substitutions at structurally equivalent positions in NMR ensembles and their corresponding homologues, respectively, using a refined dataset of 65 NMR structures with their homologues (see Materials and Methods). Structurally equivalent positions in a NMR ensemble and its corresponding homologues were identified by structural alignment between the first model from each of the NMR ensemble to all its homologues. PB variation and PB substitution matrices were calculated using the structurally equivalent positions. Pearson's correlation coefficient of 0.93 between these matrices confirms that the high correlation between PB variation and PB substitution matrices is independent of datasets used. Similarly, we constructed PB variation and PB substitution matrices only for structurally equivalent positions in the loop regions and still we got a high correlation coefficient of 0.91. This suggests that even in loop regions the general rules for local structural changes are followed as observed in global PB variation or substitution matrices.

We also compared each structurally equivalent position of the loop regions separately in NMR ensembles and their homologues. Surprisingly this position-wise analysis showed that 793 out of 1281 (61.90%) of these positions, PB variations and PB substitutions are not correlated *i.e.* correlation coefficient is less than 0.5 (fig. 3); *e.g.* residue positions 44 to 48 in second Kunitz domain of human tissue factor pathway inhibitor protein (Burgering et al., 1997) (PDB: 1ADZ) have no correlation between PB variations in this NMR ensemble and PB substitutions at structurally equivalent positions in its homologues (fig. 4). PBs at these positions in NMR ensemble and at the corresponding structurally equivalent positions in homologues are given in supplementary table S1. Similar example is shown in figure 5a, at residues 30 and 31 (shown in gray color) from NMR structure of phospholipase A2 (Yuan et al., 1999) (PDB: 1BVM), there is

a weak correlation between PB variations in the NMR ensemble and PB substitutions at structurally equivalent positions in its homologues of phospholipase A2.

Weak correlations at majority of the loop region structurally equivalent positions in NMR ensembles and their homologues inform us about the different local structures adopted at these positions but it does not inform us on how different these local structures are. To quantify the differences in adopted local structures, we calculated 'average PB variation scores' and 'average PB substitution scores' respectively from NMR ensembles and their homologues at structurally equivalent positions in loop regions (see Materials and Methods). We compared the distributions of 'average PB variation scores' and 'average PB substitution scores' (fig. 6) for the structurally equivalent positions in loop regions. The significance of difference between these distributions was assessed by paired t-test. 'Average PB substitution scores' were significantly lower than 'average PB variation scores' (t=38.36, df=1280, p-value=2.2e-16), suggesting that observed local structural variations in NMR ensembles are quantitatively different than the local structural differences observed in their homologues at loop region structurally equivalent positions. Lack of negative scores in the score distributions in figure 6 also confirm that inherent structural dynamics as well as evolutionary structural dynamics follow the rules defined for local structural changes in global PB variation and substitution matrices.

Out of 488 loop region structurally equivalent positions with high (at least 0.5) correlations (shaded region in figure 3) in PB variations and substitutions, 262 (53.69%) positions have more than 2 score difference in 'average PB variation scores' and 'average PB substitution scores'; *e.g.* at structurally equivalent position corresponding to residue 19 from NMR structure of human macrophage inflammatory protein (PDB: 2JYO) (Chan, Hunter, Tack, & Vogel, 2008), the correlation coefficient between PB variations in NMR ensemble and PB

substitutions in its homologues is 0.59 but 'average PB substitution score' is higher than 'average PB variation score' (figure 5b).

Conclusions

Most of the local structural variations in the loop regions in NMR ensembles do not correlate to their local structural differences at structurally equivalent positions in homologous proteins of known structure. However, the overall nature of local structural changes remains highly correlated, *i.e.* the pattern followed for these local structural changes is unique. These rules for local backbone structural changes can be used to guide molecular dynamics and protein design approaches.

This study also underlines the caution we must exercise while arriving at general conclusions based on analysis of concatenated data. High or low correlations for concatenated data may not always apply to segregated data.

Acknowledgements

(France) and Conseil Régional Pays de La Loire (France) to BO and Department of Biotechnology (India) to NS.

References

Agarwal, G., Mahajan, S., Srinivasan, N., & De Brevern, A. G. (2011). Identification of Local Conformational Similarity in Structurally Variable Regions of Homologous Proteins Using Protein Blocks. *PLoS ONE*, *6*(3), e17826.

Balaji, S., Sujatha, S., Kumar, S. S. C., & Srinivasan, N. (2001). PALI—a database of Phylogeny and ALIgnment of homologous protein structures. *Nucleic Acids Research*, *29*(1), 61–65.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., et al. (2000). The Protein Data Bank. *Nucleic Acids Research*, *28*(1), 235–242.

Best, R. B., Lindorff-Larsen, K., DePristo, M. A., & Vendruscolo, M. (2006). Relation between native ensembles and experimental structures of proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(29), 10901–6.

Bhabha, G., Lee, J., Ekiert, D. C., Gam, J., Wilson, I. A., Dyson, H. J., Benkovic, S. J., et al. (2011). A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science (New York, N.Y.)*, *332*(6026), 234–8.

Burgering, M. J., Orbons, L. P., Van der Doelen, A., Mulders, J., Theunissen, H. J., Grootenhuis, P. D., Bode, W., et al. (1997). The second Kunitz domain of human tissue factor pathway inhibitor: cloning, structure determination and interaction with factor Xa. *Journal of molecular biology*, *269*(3), 395–407.

Chan, D. I., Hunter, H. N., Tack, B. F., & Vogel, H. J. (2008). Human macrophage inflammatory protein 3alpha: protein and peptide nuclear magnetic resonance solution structures, dimerization, dynamics, and anti-infective properties. *Antimicrobial agents and chemotherapy*, *52*(3), 883–94.

Clore, G. M., Robien, M. A., & Gronenborn, A. M. (1993). Exploring the limits of precision and accuracy of protein structures determined by nuclear magnetic resonance spectroscopy. *Journal of molecular biology*, *231*(1), 82–102.

Davis, I. W., Arendall, W. B., Richardson, D. C., & Richardson, J. S. (2006). The backrub motion: how protein backbone shrugs when a sidechain dances. *Structure (London, England : 1993)*, *14*(2), 265–74.

De Brevern, A. G., Etchebest, C., & Hazout, S. (2000). Bayesian probabilistic approach for predicting backbone structures in terms of protein blocks. *Proteins*, *41*(3), 271–287.

DePristo, M. A., De Bakker, P. I. W., & Blundell, T. L. (2004). Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography. *Structure (London, England : 1993)*, *12*(5), 831–8.

Doshi, U., McGowan, L. C., Ladani, S. T., & Hamelberg, D. (2012). Resolving the complex role of enzyme conformational dynamics in catalytic function. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(15), 5699–704.

Etchebest, C., Benros, C., Hazout, S., & De Brevern, A. G. (2005). A structural alphabet for local protein structures: improved prediction methods. *Proteins*, *59*(4), 810–827.

Friedland, G. D., Lakomek, N.-A., Griesinger, C., Meiler, J., & Kortemme, T. (2009). A correspondence between solution-state dynamics of an individual protein and the sequence and conformational diversity of its family. *PLoS computational biology*, *5*(5), e1000393

Friedland, G. D., Linares, A. J., Smith, C. A., & Kortemme, T. (2008). A simple model of backbone flexibility improves modeling of side-chain conformational variability. *Journal of molecular biology*, *380*(4), 757–74.

Furnham, N., Blundell, T. L., DePristo, M. A., & Terwilliger, T. C. (2006). Is one solution good enough? *Nature structural & molecular biology*, *13*(3), 184–5; discussion 185.

Henzler-Wildman, K., & Kern, D. (2007). Dynamic personalities of proteins. *Nature*, *450*(7172), 964–72.

Ishima, R., & Torchia, D. A. (2000). Protein dynamics from NMR. *Nature Structural Biology*, *7*(9), 740–743.

Johnson, M. S., & Overington, J. P. (1993). A Structural Basis for Sequence Comparisons: An Evaluation of Scoring Methodologies. *Journal of Molecular Biology*, *233*(4), 716–738.

Joseph, A. P., Agarwal, G., Mahajan, S., Gelly, J.-C., Swapna, L. S., Offmann, B., Cadet, F., et al. (2010). A short survey on protein blocks. *Biophysical Reviews*, *2*(3), 137–145.

Joseph, A. P., Srinivasan, N., & De Brevern, A. G. (2011). Improvement of protein structure comparison using a structural alphabet. *Biochimie*, *93*(9), 1434–1445.

Keedy, D. A., Georgiev, I., Triplett, E. B., Donald, B. R., Richardson, D. C., & Richardson, J. S. (2012). The role of local backrub motions in evolved and designed mutations. *PLoS computational biology*, *8*(8), e1002629.

Konagurthu, A. S., Whisstock, J. C., Stuckey, P. J., & Lesk, A. M. (2006). MUSTANG: a multiple structural alignment algorithm. *Proteins*, *64*(3), 559–574.

Lange, O. F., Lakomek, N.-A., Farès, C., Schröder, G. F., Walter, K. F. A., Becker, S., Meiler, J., et al. (2008). Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science (New York, N.Y.)*, *320*(5882), 1471–5.

Larson, S. M., England, J. L., Desjarlais, J. R., & Pande, V. S. (2002). Thoroughly sampling sequence space: large-scale protein design of structural ensembles. *Protein science : a publication of the Protein Society*, *11*(12), 2804–13.

Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *Journal of biomolecular NMR*, *8*(4), 477–86.

Liu, Y., & Bahar, I. (2012). Sequence Evolution Correlates with Structural Dynamics. *Molecular biology and evolution*, *29*(9), 2253–63.

Murzin, A. G., Brenner, S. E., Hubbard, T., & Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *Journal of molecular biology*, *247*(4), 536–540.

Offmann, B., Tyagi, M., & De Brevern, A. G. (2007). Local Protein Structures. *Current Bioinformatics*, *2*, 165–202.

Salmon, L., Bouvignies, G., Markwick, P., & Blackledge, M. (2011). Nuclear magnetic resonance provides a quantitative description of protein conformational flexibility on physiologically important time scales. *Biochemistry*, *50*(14), 2735–47.

Smith, C. A., & Kortemme, T. (2008). Backrub-like backbone simulation recapitulates natural protein conformational variability and improves mutant side-chain prediction. *Journal of molecular biology*, *380*(4), 742–56.

Sze, K. H., & Lai, P. M. (2011). Probing protein dynamics by nuclear magnetic resonance. *Protein and Peptide Letters*, *18*(4), 373–379.

Teilum, K., Olsen, J. G., & Kragelund, B. B. (2009). Functional aspects of protein flexibility. *Cellular and molecular life sciences : CMLS*, *66*(14), 2231–47.

Tyagi, M., Gowri, V. S., Srinivasan, N., De Brevern, A. G., & Offmann, B. (2006). A substitution matrix for structural alphabet based on structural alignment of homologous proteins and its applications. *Proteins*, *65*(1), 32–39.

Tzeng, S.-R., & Kalodimos, C. G. (2011). Protein dynamics and allostery: an NMR view. *Current Opinion in Structural Biology*, *21*(1), 62–67.

Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., Livny, M., et al. (2007). BioMagResBank. *Nucleic Acids Research*, *36*(Database), D402–D408.

Yuan, C., Byeon, I. J., Li, Y., & Tsai, M. D. (1999). Structural analysis of phospholipase A2 from functional perspective. 1. Functionally relevant solution structure and roles of the hydrogen-bonding network. *Biochemistry*, *38*(10), 2909–18.

Zhuravlev, P. I., Materese, C. K., & Papoian, G. A. (2009). Deconstructing the native state: energy landscapes, function, and dynamics of globular proteins. *The journal of physical chemistry. B*, *113*(26), 8800–12.
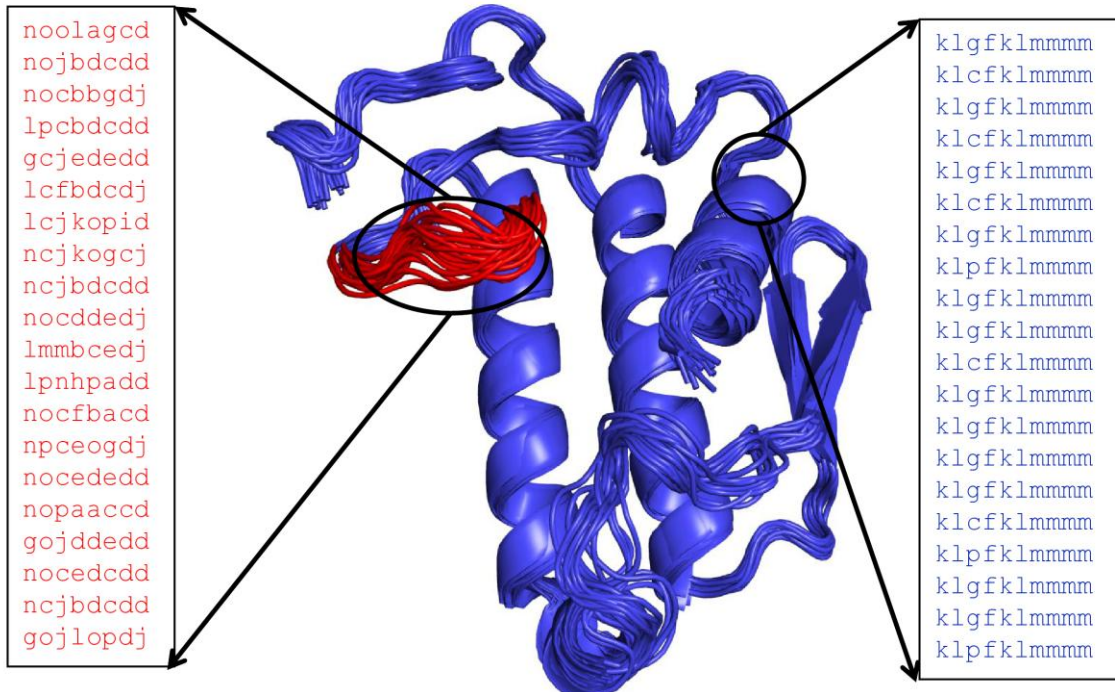
Figures

Figure 1. Characterization of local structural changes using PBs. Different models from NMR structure of phospholipase A2 (Yuan et al., 1999) enzyme (PDB: 1BVM) are shown in blue color. Regions with local backbone structural variations (encircled red residues) between different models of a NMR ensemble can be easily characterized and distinguished from regions

with no local backbone structural variations (encircled blue residues) by using PB assignments for these regions (shown in rectangular boxes).
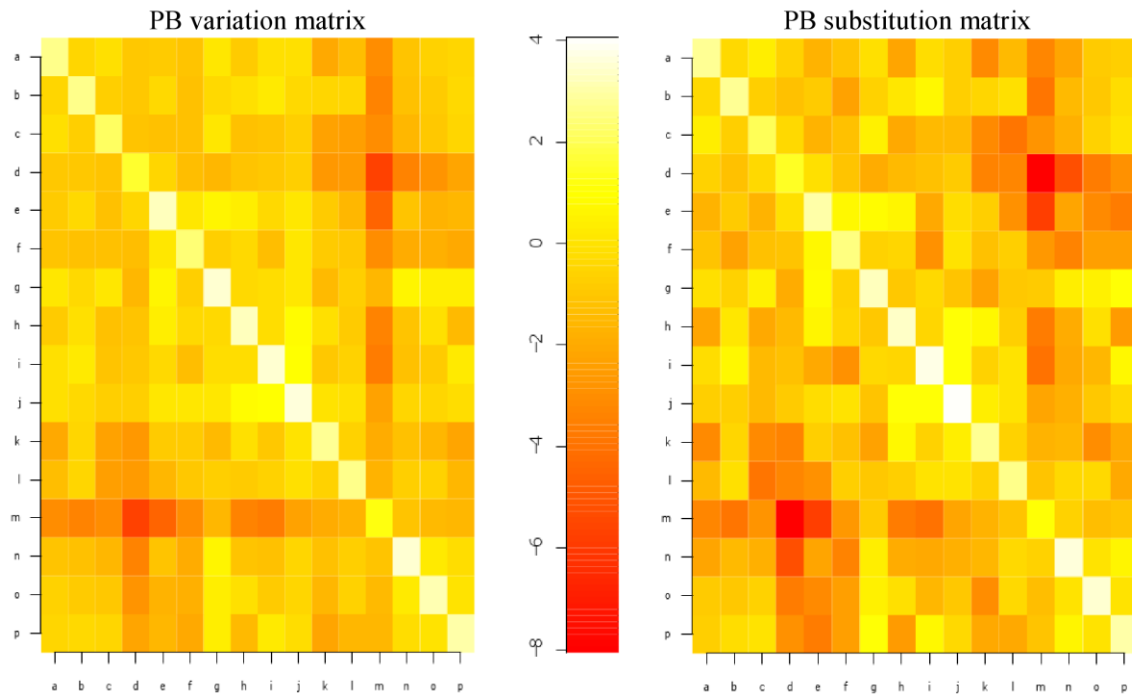

Figure 2. **Correlation between global PB variation and substitution matrix.** Global PB variation matrix (left) shows similar patterns for local structural changes as in global PB substitution matrix (right) generated by Joseph *et al* (Joseph et al., 2011), with high correlation coefficient of 0.94.
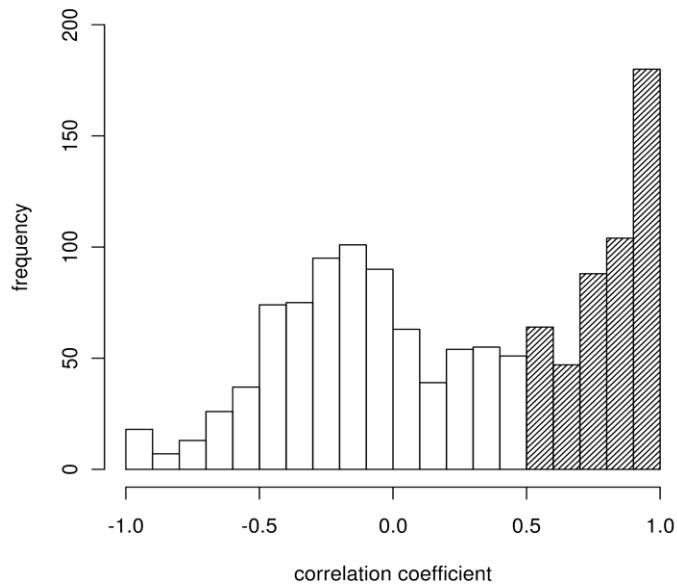
Figure 3. **Segregated position-wise analysis of PB changes.** Histogram of correlation coefficients between PB variations observed in NMR ensembles and PB substitutions in their homologues at 1281 structurally equivalent positions from loop region is shown here. For 61.90% of these positions correlation coefficient between PB variations and PB substitutions is less than 0.5 (area without shading).
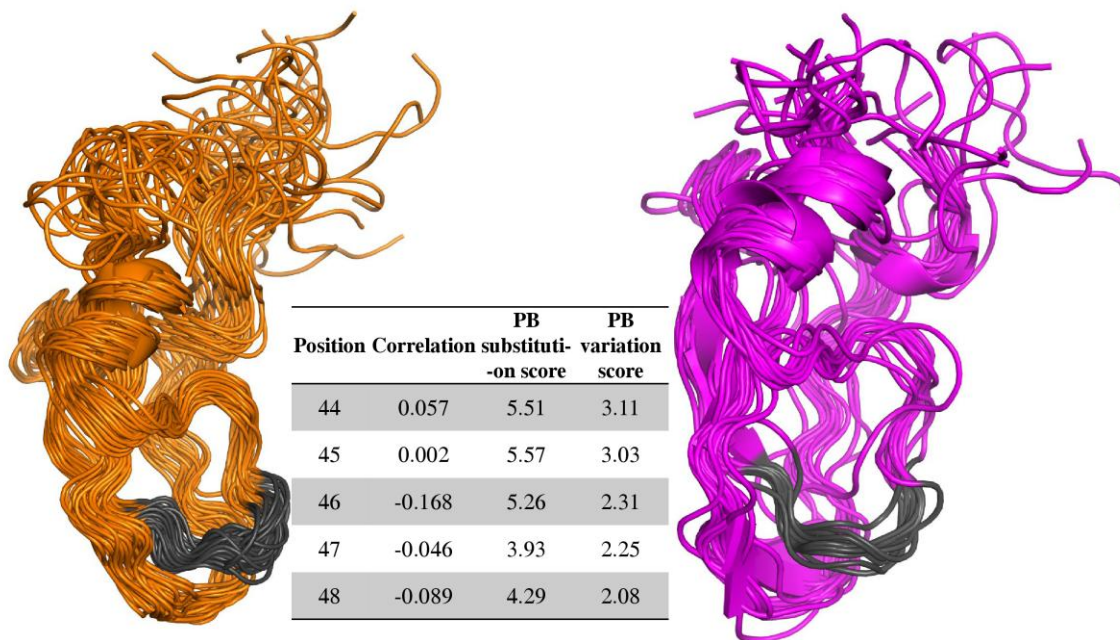
Position	Correlation	PB substituti-on score	PB variation score
44	0.057	5.51	3.11
45	0.002	5.57	3.03
46	-0.168	5.26	2.31
47	-0.046	3.93	2.25
48	-0.089	4.29	2.08

Figure 4. **Analysis of local structural changes in second Kunitz domain of human tissue factor pathway inhibitor protein (TFPI).** In NMR ensemble of second Kunitz domain of human TFPI protein (Burgering et al., 1997) (PDB: 1ADZ), at residue numbers 44 to 48 (shown in gray color) there is no correlation between PB variations in this NMR ensemble (shown in orange color) and PB substitutions at structurally equivalent positions in its homologues (shown in magenta color) of known structure. Residue number in NMR ensemble, correlation between PB variations and PB substitutions, 'average PB substitution score' and 'average PB variation score' are given in a table. PBs at these positions are given in supplementary table S1.
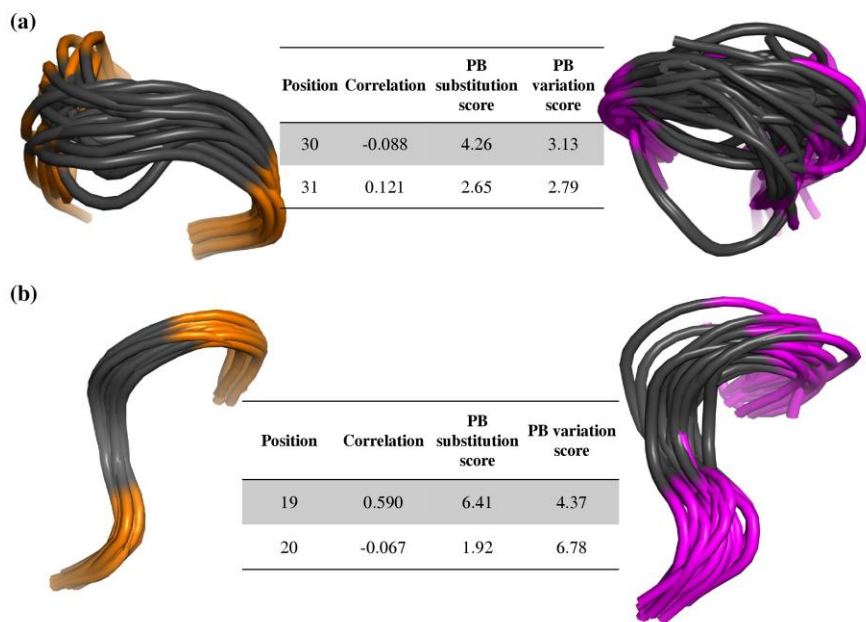
Figure 5. Case studies for the analysis of local structural changes. (a) At residues 30 and 31 (shown in gray color) from NMR structure of phospholipase A2 (Yuan et al., 1999) (PDB: 1BVM), there is no correlation between PB variations in the NMR ensemble (shown in orange color) and PB substitutions at structurally equivalent positions in its homologues (shown in magenta color). (b) At residue 19 (shown in gray color) from NMR structure of human macrophage inflammatory protein (PDB: 2JYO) (Chan et al., 2008), the correlation coefficient between PB variations in NMR ensemble (shown in orange color) and PB substitutions in its homologues (shown in magenta color) is 0.59 but 'average PB substitution score' is higher than 'average PB variation score'. Residue number in NMR ensembles, correlation between PB variations and PB substitutions, 'average PB substitution score' and 'average PB variation score' are given in a table.
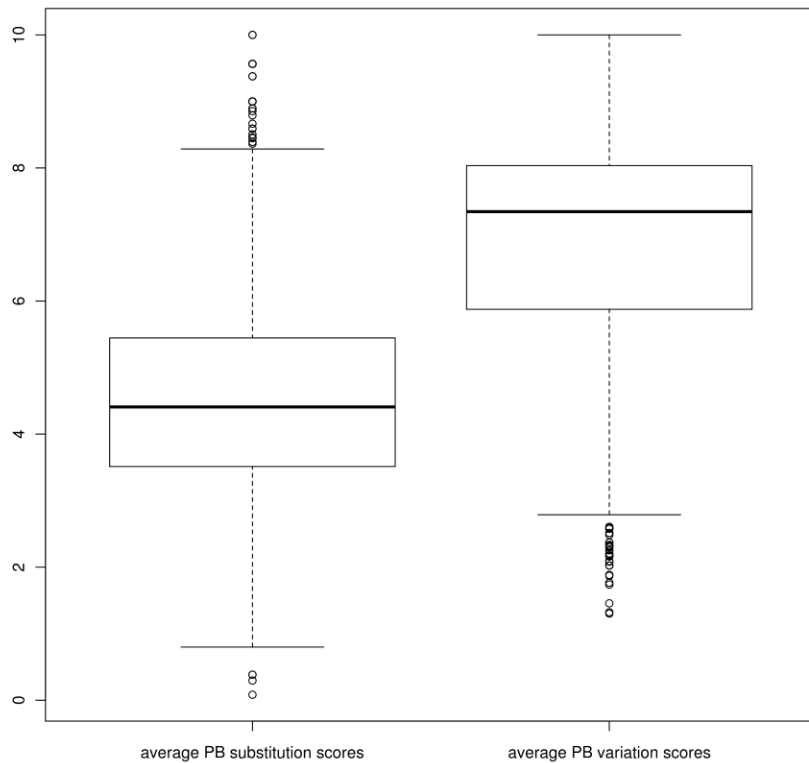
Figure 6. Segregated position-wise analysis of PB change scores. Box-plots of 'average PB variation scores' in NMR ensembles and 'average PB substitution scores' in their homologues at structurally equivalent positions from loop regions are shown here. 'Average PB substitution scores' are significantly lower than 'average PB variation scores' (paired t-test: t=38.36, df=1280, P-value=2.2e-16).