

Normal Mode Analysis of Thermophilic Cellulase FnCel5A Using Elastic Network Models

Shah Faisal Mohammad^{1*}, Fawad Ali¹, Mueed Ur Rahman¹ and Asim Muhammad²

¹State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China

²Department of Agronomy, The University of Agriculture, Peshawar, 25130, Khyber Pakhtunkhwa, Pakistan

Abstract

Normal mode analysis (NMA) based on elastic network models is an efficient, cost-effective and powerful, computational approach for characterizing protein flexibility and the resulting dynamics. In this study, we have analyzed a single protein structure, a hyperthermophilic enzyme (FnCel5A) with a given PDB ID: 3rjy, in order to calculate deformation energies, Eigenvalues, atomic fluctuations and displacements, and overlap and correlation matrices that display correlations among all of the C-alpha-atom motions in the FnCel5A structure. The WEBnm@ server was used to provide a quick automated computational low-frequency normal protein structure mode analysis. Single mode analysis using NMA has been applied on the web server, which can provide recently improved functionality for the single protein structures. This includes new visualization of protein motions, inter-amino acids correlations, and conformational transition analysis applying the overlap analysis. Furthermore, we have studied the structural and single mode FnCel5A (PDB: 3rjy) analysis in order to calculate the lowest and normal frequency protein modes. This study provides enough information about the loop flexibility and highest B-factors regions of FnCel5A which play an important role in rational and semi-rational designing to engineer this enzyme for improved activity.

Keywords: Single mode analysis; Deformation energy; Overlap; Atomic displacement

Introduction

Proteins are the basic building block of nearly all biological functions. All the protein consists of primary amino acid sequence that translates the tertiary structure of the protein, and ultimately describes the overall protein's function. According to the central protein dogma, the amino acid sequence determines the protein structure and functional relationship. Protein structure and function is commonly correlated but usually, proteins are dynamic in nature, and their inner motions among the molecule facilitate their function. In fact, the protein's dynamics is the intrinsic interface between its structure and function [1]. However, only the local dynamical property of an enzyme active sites are not involved in the functional and catalytic competent state essential to perform the proteins biological functions. Indeed, the overall protein's dynamics is integral for its whole biological functions. Numerous studies and investigations affirm the impacts of protein dynamics on enzyme catalysis, substrate binding, the proliferation of allosteric signal, and the interaction between the proteins [2-8]. Further, it has been investigated that the dynamics of amino acids that are not directly engaged in catalysis are also significant for the other functions of enzyme [9]. Protein dynamism is the result of time-dependent change and evolution in the structure of a protein which involves fluctuations of equilibrium controlling the biological function [10]. By deforming the inner structure of the protein can successfully determine these time-dependent fluctuations. The mechanistic features of the protein determine the structural flexibility at all levels, from the static fluctuations of residues chains to the reorientation of huge rigid bodies like protein domains. Formerly numerous studies have been focused on the qualitative descriptions of very less number of static structures, but there is wide evidence that the structure of a protein is correlated to the dynamics and it may be evolutionarily and structurally conserved. Thus, it's a need for analyzing protein dynamics in a systematic, efficient and organize quantitative manner. Despite the fact that the number of dynamical experimental data is overgrowing day by day e.g. from the investigation of the time dimension in X-ray

and NMR studies, the knowledge gained from them is still incomplete. A number of computational approaches have been adapted to counterpart the experimental structural biology data and furnish dynamical models of the biomolecules. Molecular Dynamics (MD) simulations are extensively used to simulate the protein dynamics on time scale of up to the microsecond. Computationally it's laborious and cost-effective to analyze the data by MD simulation and the time scales essential to large sample motions such as domain reorientation is not easily accomplished. Comparatively Normal Mode Analysis (NMA) using Elastic Network Models (ENM) is an economical and efficient method for characterizing and predicting slow motions of proteins. The productivity and performance of ENM methodology also makes them more suitable for automated comparative analysis among multiple protein structures. NMA has been extensively used to analyze the latent protein flexibility [11-14]. NMA modeling the atoms motions in a protein as a pair harmonic oscillator and combine each mode of motion to an oscillation frequency. Lower frequencies of different modes correspond to overall or domain motions and have been investigated to coordinate well to functionally related protein dynamics [15-18]. In addition it has been shown to obtain results in precise harmony with molecular dynamics simulations when characterizing the cumulative proteins motions [14,19-23]. In Elastic Network Models, the protein is shown as a spring's network, grabbing the density of particles and the interaction strength between them. Inter-atomic interactions illustrated by a simple harmonic potential

***Corresponding authors:** Shah Faisal Mohammad, State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China, Tel: +8615121172526; E-mail: faisalbbt@yahoo.com

Received February 21, 2018; Accepted May 21, 2018; Published May 28, 2018

Citation: Mohammad SF, Ali F, Rahman MU, Muhammad A (2018) Normal Mode Analysis of Thermophilic Cellulase FnCel5A Using Elastic Network Models. J Proteomics Bioinform 11: 111-116. doi: [10.4172/jpb.1000475](https://doi.org/10.4172/jpb.1000475)

Copyright: © 2018 Mohammad SF, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

and the protein is commonly modeled with a small number of atoms, mostly single bead per residue residing at the C-alpha position. This granularity of the model is suitable to reduce the dimensionality, motivated by the approximate independence of overall-residue motion from the side-chain dynamics in globular proteins [24]. In this study, we apply the single mode analysis of NMA approach to our protein FnCel5A (PDB ID: 3rjy) to investigate the flexibility and motions of its structure (Figure 1). This analysis is helpful in providing information, how the structure of FnCel5A moves and how its flexibility is conserved over its structure.

Materials and Methods

Protein sequence and structure

An FnCel5A protein was selected from the online Protein Data Bank with PDB ID (3Rjy) [25] based upon structural data availability. From the previous study, we know that that the selected protein is a homopolymer, therefore, single polypeptide chain (chain A) was chosen to analyze the protein sequence and structure, as well as protein dynamics.

Pymol analysis

The FnCel5A structure was analyzed by using Pymol software to find out the highest B factor regions and was compared with the deformation energies and other information's of the Normal Mode Analysis.

Calculation of normal modes

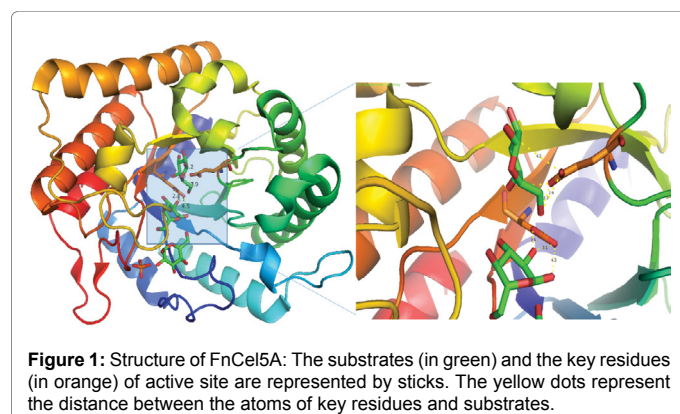
Hinsen et al., developed the WEBnm@ server using the Elastic Network Model (ENM) with the C-alpha force field [24], accessible in the Molecular Modeling Toolkit (MMTK) [26]. Each residue is indicated by a mass at the position of its C α atoms.

Results

The PDB file 3.rjy was submitted to the online WEBnm@ server [27,28] for single NMA analysis using elastic network model to analyze the FnCel5A structure to find out the highest B-factors and other flexible regions by calculating the deformation energies, eigenvalue atomic displacement and the fluctuation between C-alpha atoms.

Deformation energies

Deformation energies shown in Table 1 and eigenvalues shown in Figure 2 showing the energy associated with each mode and is inversely related to the amplitude of the motion described by the corresponding modes. The protein structure moves along all the normal modes at



once. A mode can only be evaluated in isolation if it is energetically well separated from other modes. The PDB file 3.rjy contains atoms that have no alpha-carbon, or not a standard amino acid, therefore those atoms were ignored. The eigenvectors and eigenvalues of the mass-weighted matrix of the second derivative of $U(r)$ correspond to normal modes and the frequency squares for each normal mode, respectively. In WEBnm@ uses a molecular modeling toolbox and uses the coarse granular ENM to compute normal low frequency modes as shown in Figure 3. As shown in Figure 2 the B-factors that corresponds to high deformation energy are the residues: 55-62, 88-105, 211-217, 236-275, 328-332, which comprise the surface loops of the FnCel5A protein. Both the normal mode analysis and Pymole study correlate with each other.

Atomic Displacement

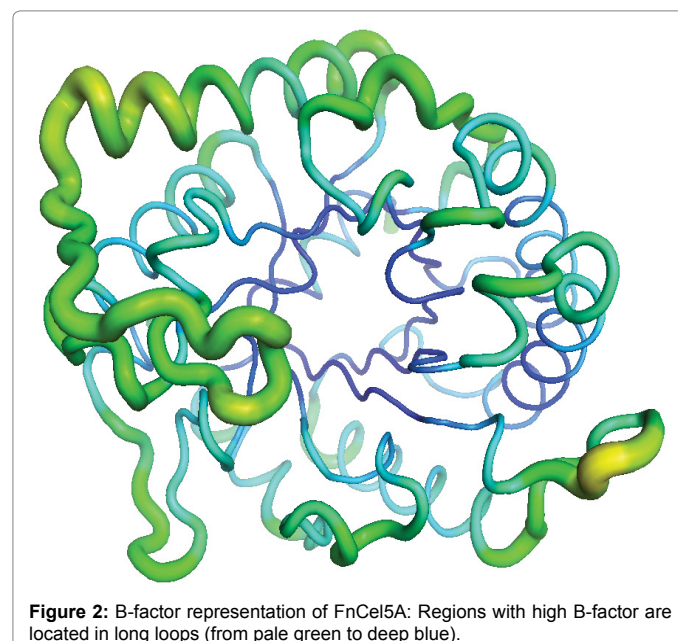
The Figure 4 showing the displacement square of each C-alpha atom (for modes 7 to 12) normalized so that the sum of all residues is equal to 100. The highest values correspond to the most displaced regions. Clusters of peaks on the plots identify significantly displaced regions, while isolated peaks may reflect local flexibility.

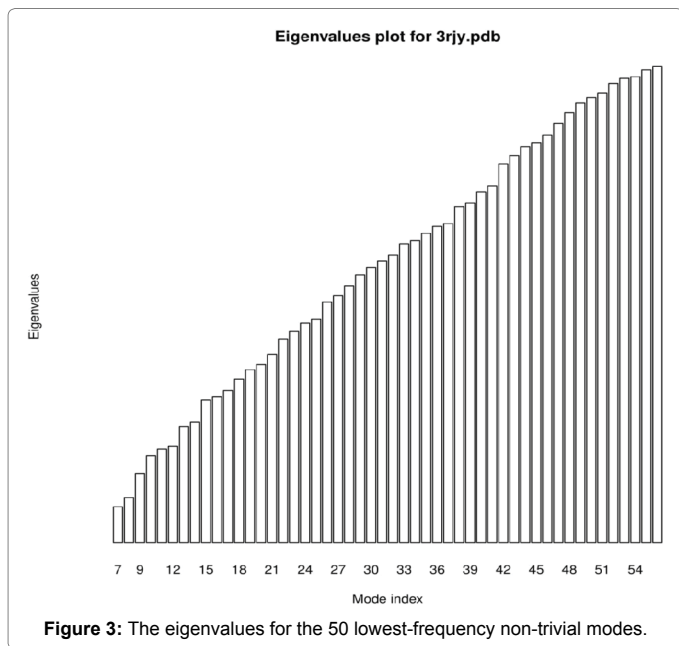
Fluctuations

Figure 5 describe the fluctuation flexibility of the C α atoms of the

| Mode Index | Deformation Energy | Mode Index | Deformation Energy |
|------------|--------------------|------------|--------------------|
| 7 | 1397.51 | 14 | 5857.17 |
| 8 | 1546.88 | 15 | 7643.34 |
| 9 | 3536.38 | 16 | 6121.13 |
| 10 | 3460.52 | 17 | 7863.84 |
| 11 | 4674.56 | 18 | 8391.23 |
| 12 | 5776.24 | 19 | 9609.12 |
| 13 | 6634.50 | 20 | 9245.02 |

Table 1: Values of the deformation energy for the lowest-frequency non-trivial modes (modes 7 to 20).





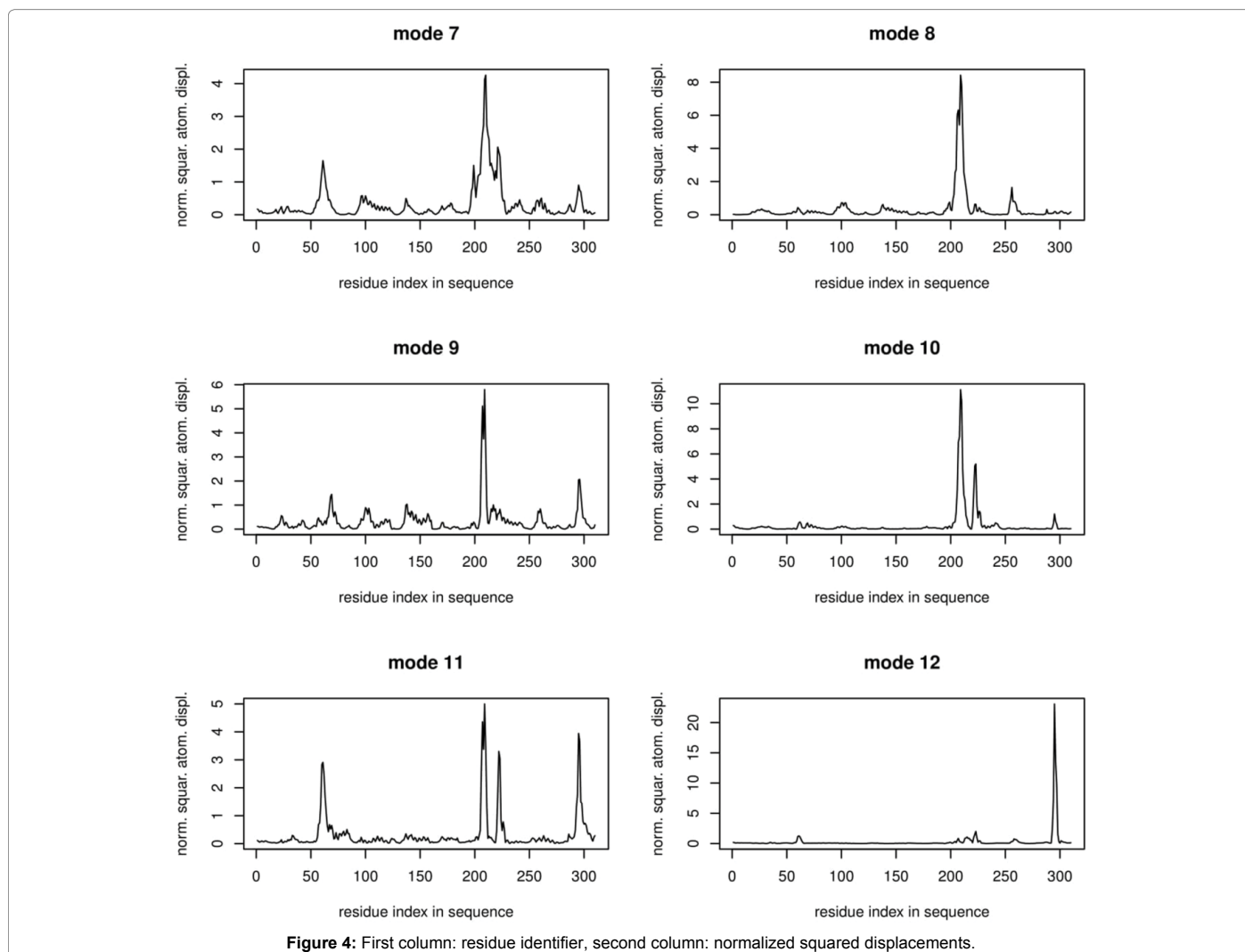
enzyme. The residue numbers are indicated on the x-axis of the graph and the amplitude of the fluctuations on the y-axis. The graph shows normalized squared fluctuations of each C α atom and is computed by using the default WEBnm@. The fluctuation square of each C-alpha atom (calculated from 200 lowest non-trivial modes), was normalized so that the sum of all residues become equal to 100.

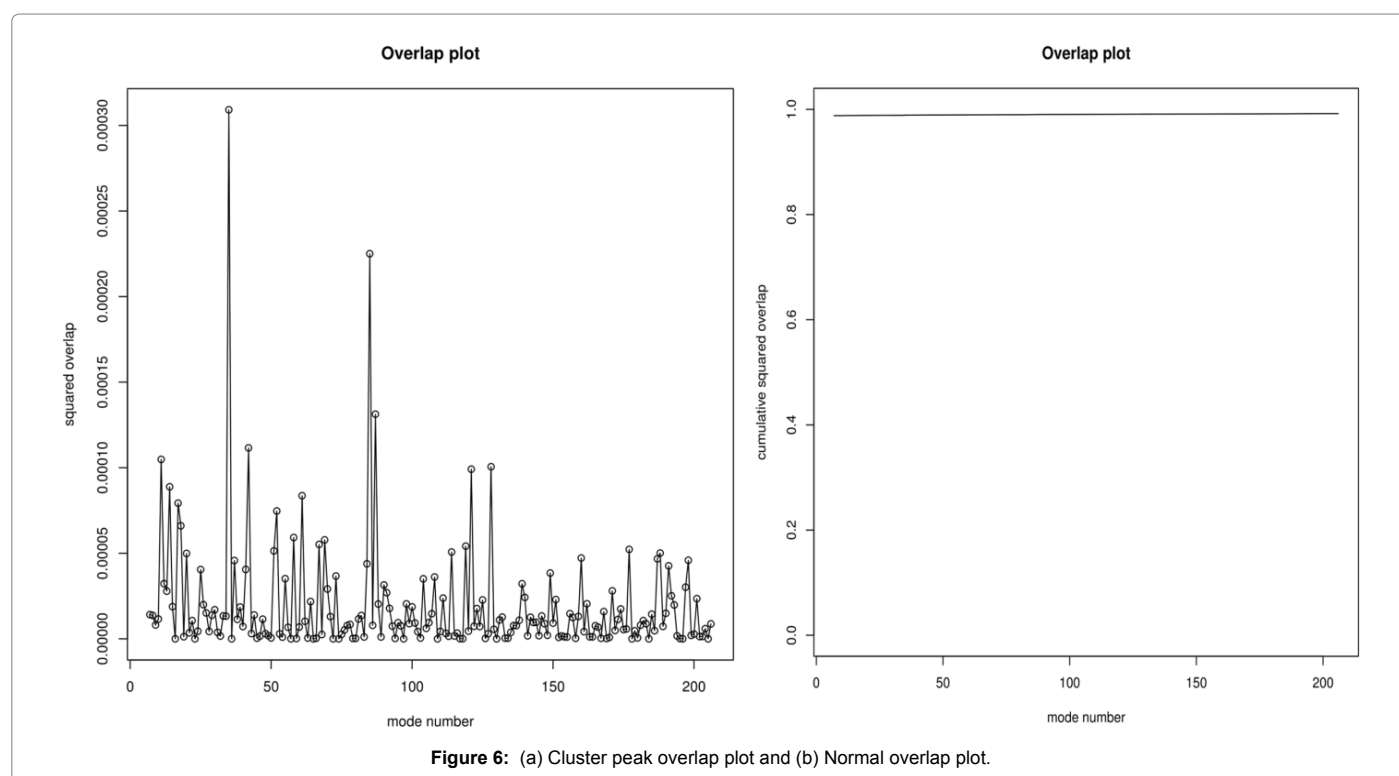
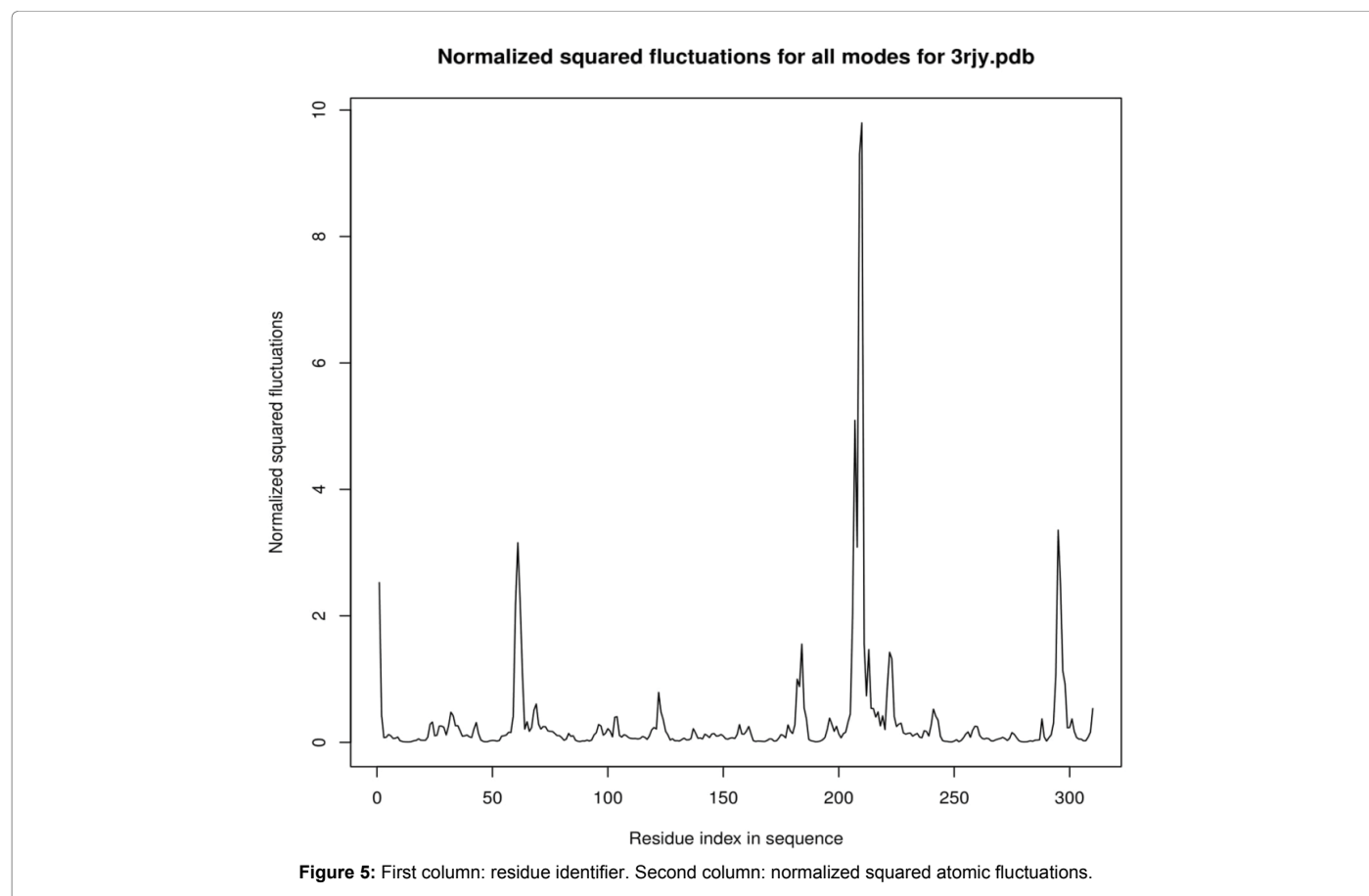
Overlap between normal modes and difference vector

The Figure 6A and B showing the squared inner product between the difference vector and the normal modes identifies which modes contribute most to the structural difference.

Correlation Matrix Result

The correlation matrix shows the correlated movement of the C-alphas atoms in the FnCel5A protein. Both axes denote the C-alphas atoms of the protein in sequential order, so that each cell in the plot shows the coupling of two residues in the protein on a range from -1 (anti-correlated) via 0 (uncorrelated) to 1 (correlated). The measure expresses the expected inner production of atomic displacement, so both correlated and anti-correlated motion could be interpreted as coupled motions. Selected, significant correlation matrix values are mapped onto the input structure for a 3-D interpretation of the





patterns seen on the heat map. The correlation matrix is searched in all four directions for an uninterrupted network of correlated residues that fulfill the following: First, a score threshold is set such that that it falls in a percentile between 95 and 99.9 (depending on the input structure size) of the distribution of the absolute scores in the correlation matrix. The positive correlation scores, thus, range from this threshold up to one, while the negative correlation scores range from the negative of the threshold down to -1. Then, pairs of correlated residues that have a distance below 0.8 nm are excluded. The resulting lists of residues form a group that is connected in sequence, and by regions of strong correlations that go beyond local contacts. These groups may consist of only a pair of residues many residues that are connected by sticks. There is a weak correlation in atomic fluctuation (red color regions in Figure 7), indicates that residues in this regions fluctuate simultaneously in a correlated fashion.

Discussion

The dynamic simulation is critical for understanding macromolecular function with biologically relevant sizes and time scales but this approach is cost-effective and time consuming as compared to Normal Mode Analysis. In this study, normal mode analysis (NMA) was used to analyze the complex dynamic function of a macromolecule FnCel5A as a set of simple harmonic oscillation vibrating around a fixed equilibrium conformation. This approach, was developed from classical mechanics, and was first used to find out the dynamical behavior of small biological systems about thirty years ago. During this era, abundant of evidence have been combined to validate NMA as a powerful tool for simulating macromolecular motions at expanding length scales. Recently, NMA, with coarse representations, has become an effective alternative to molecular dynamics simulations to study the slow and large-scale movements of macromolecular machines. Interesting insights into these systems can be obtained very quickly with NMA to characterize their flexibility, to predict the directions of their collective conformational changes, or to help in the evaluation of experimental structural data. In protein crystallography, Atomic displacement parameters (ADPs, also known as thermal or B-factors) have given increasing significance as they

contribute enough information on the flexibility of side-chain and main-chain atoms when anisotropically [29,30] or isotropically [31] maintained. The ADPs also play an important role to determine the accuracy of atomic position and its coordinates [32,33]. It can also contribute statistically by analyzing various structural results [34]. The best predictions of polypeptide chain flexibility were based on analyzes of ADPs in known structural kits [35-37]. In this study the eigenvalue and ADPs showing significant effect on FnCel5A structure. Unfortunately, ADP averages and variances are often significantly different for different protein structure determinations, since absolute ADP values do not depend only on various physical phenomena such as true oscillations for stable conformations or static, dynamic, and lattice distortions, but also depend on structural differences in refinement methods and levels [38].

Conclusion

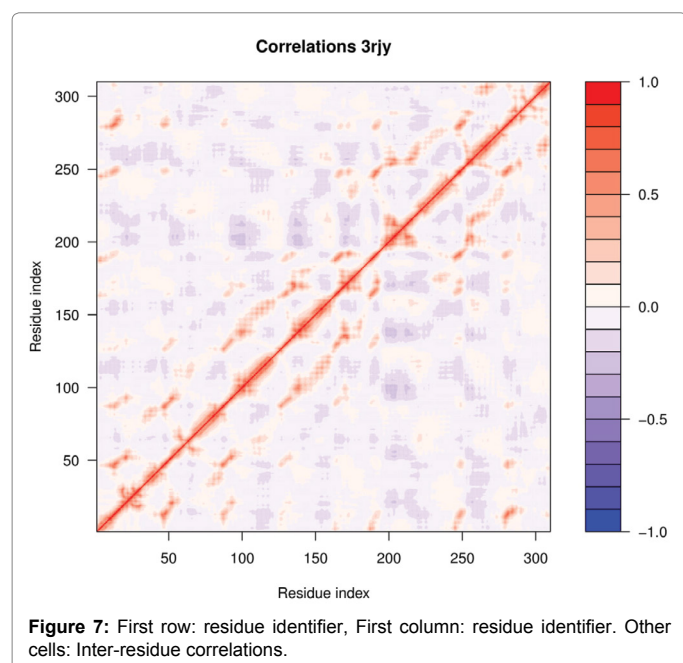
In recent years, great interest in NMA-inspired approaches has been observed as they provide a unique and biologically relevant analytical solution to the equilibrium dynamics of biomolecules. The single-mode analysis calculates the lowest frequency modes and normal proteins and offers different types of calculations to analyze these modes. The single-mode analysis of FnCel5A shows that the three-dimensional structures of this protein contain the information necessary to describe its functional movements. Most of the global or collective motion modes predicted by NMA are insensitive to the details of model and energy parameters, and rather depend on the topology of contacts between equilibrium residues; this justifies the widespread use of the more efficient particle size EN model described here. This information on FnCel5A can be used in conjunction with experimental studies to resolve the supramolecular dynamics and longtime scales of the FnCel5A structure, which may not be accessible with conventional simulation approaches.

Acknowledgement

The author enthusiastically acknowledges the Microbial Metabolism Laboratory, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University Shanghai P.R China, for the provision of facilities to carry out this research. The author has not got any special technical and financial support from any institute or funding agency to perform this research.

References

1. Hensen U, Meyer T, Haas J, Rex R, Vriend G, et al. (2012) Exploring protein dynamics space: The dynasome as the missing link between protein structure and function. *PLoS ONE* 7: e33931.
2. Agarwal PK, Billeter SR, Rajagopalan PR, Benkovic SJ, Hammes-Schiffer S, et al. (2002) Network of coupled promoting motions in enzyme catalysis. *Proc Natl Acad Sci USA* 99: 2794-2799.
3. Bhabha G, Lee J, Ekiert DC, Gam J, Wilson IA, et al. (2011) A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science* 332: 234-238.
4. Klinman JP, Kohen A (2013) Hydrogen Tunneling links protein dynamics to enzyme catalysis. *Annu Rev Biochem* 82: 4710-496.
5. Lisi GP, Loria JP (2016) Using NMR spectroscopy to elucidate the role of molecular motions in enzyme function. *Prog Nucl Magn Reson Spectrosc* 92-93: 1-17.
6. Liu X, Speckhard DC, Shepherd TR, Sun YJ, Hengel SR, et al. (2016) Distinct roles for conformational dynamics in protein-ligand interactions. *Structure* 24: 2053-2066.
7. Singh P, Francis K, Kohen A (2015) Network of remote and local protein dynamics in dihydrofolate reductase catalysis. *ACS Catal* 5: 3067-3073.
8. Martinez EV, Dörner C (2016) Conformational substrate selection contributes to the enzymatic catalytic reaction mechanism of pin1. *J Phys Chem B* 120: 12444-12453.



9. Roston D, Kohen A, Doron D, Major DT (2014) Simulations of remote mutants of dihydrofolate reductase reveal the nature of a network of residues coupled to hydride transfer. *J Comput Chem* 35: 1411-1417.
10. Wildman HK, Kern D (2007) Dynamic personalities of proteins. *Nature* 7172: 964-972.
11. Tama F (2003) Normal mode analysis with simplified models to investigate the global dynamics of biological systems. *Protein Pept Lett* 10: 119-132.
12. Ma JP (2005) Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure* 13: 373-380.
13. Bahar I, Rader AJ (2005) Coarse-grained normal mode analysis in structural biology. *Curr Opin Struct Biol* 15: 586-592.
14. Bahar I, Lezon TR, Bakan A, Shrivastava IH (2010) Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins. *Chem Rev* 110: 1463-1497.
15. Krebs WG, Alexandrov V, Wilson CA, Echols N, Yu H, et al. (2002) Normal mode analysis of macromolecular motions in a database framework: Developing mode concentration as a useful classifying statistic. *Proteins* 48: 682-695.
16. Petrone PP, Vijay S (2006) Can conformational change be described by only a few normal modes? *Biophys J* 90: 1583-1593.
17. Dobbins SE, Lesk VI, Sternberg MJE (2008) Insights into protein flexibility: The relationship between normal modes and conformational change upon protein-protein docking. *Proc Natl Acad Sci USA* 105: 10390-10395.
18. Wako H, Endo S (2011) Ligand-induced conformational change of a protein reproduced by a linear combination of displacement vectors obtained from normal mode analysis. *Biophys Chem* 159: 257-266.
19. Skjaerven L, Martinez A, Reuter N (2011) Principal component and normal mode analysis of proteins; a quantitative comparison using the GroEL subunit. *Proteins* 79: 232-243.
20. Rueda M, Chacón P, Orozco M (2007) Thorough validation of protein normal mode analysis: a comparative study with essential dynamics. *Structure* 15: 565-575.
21. Fuglebakk E, Reuter N, Hinsen K (2013) Evaluation of Protein Elastic Network Models Based on an Analysis of Collective Motions. *J Chem Theory Comput* 5618-5628.
22. Romo TD, Grossfield A (2011) Validating and improving elastic network models with molecular dynamics simulations. *Proteins* 79: 23-34.
23. Ahmed A, Villinger S, Gohlke H (2010) Large-scale comparison of protein essential dynamics from molecular dynamics simulations and coarse-grained normal mode analyses. *Proteins* 78: 3341-3352.
24. Hinsen K, Petrescu A, Dellerue S, Bellissent-FM, Kneller G, et al. (2000) Harmonicity in slow protein dynamics. *Chem Phys* 261: 25-37.
25. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. (2000) The protein data bank. *Nucleic Acids Res* 28: 235-242.
26. Hinsen K (2008) Structural flexibility in proteins: impact of the crystal environment. *Bioinformatics* 24: 521-528.
27. Hollup SM, Fuglebakk E, Taylor WR, Reuter N (2011) Exploring the factors determining the dynamics of different protein folds. *Protein Sci* 20: 197-209.
28. Tiwari SP, Fuglebakk E, Hollup SM, Skjærven L, Cragnolini T, et al. (2014) WEBnm @ v2.0: Web server and services for comparing protein flexibility. *BMC Bioinformatics* 15: 427.
29. Dauter Z, Lamzin VS, Wilson KS (1995) Proteins at atomic resolution. *Curr Opin Struct Biol* 5: 784-790.
30. Longhi S, Czjzek M, Lamzin V, Nicolas A, Cambillau C (1997) Atomic resolution (1.0 Å) crystal structure of fusarium solani cutinase: stereochemical analysis. *J Mol Biol* 268: 779-799.
31. Luedemann S, Carugo O, Wade R (1997) Substrate access to cytochrome P450cam: a comparison of a thermal motion pathway analysis with molecular dynamics simulation data. *J Mol Model* 3: 369-374.
32. Murshudov GN, Dodson EJ (1997) CCP4 Newslett. *Protein Crystallogr* 33: 31-39.
33. Cruickshank DW (1996) Proceedings of the CCP4 Study Weekend, edited by E. Dodson, M. Moore, A. Ralph & S. Bailey, Daresbury Laboratory, Warrington pp: 11-22.
34. Libeu C, Adman ET (1997) Displacement-Parameter Weighted Coordinate Comparison: I. Detection of Significant Structural Differences between Oxidation States. *Acta Cryst D* 53: 56-77.
35. Bhaskaran R, Ponnuswamy PK (1988) Positional flexibilities of amino acid residues in globular proteins. *Int. J Pept Protein Res* 32: 241-255.
36. Parthasarathy S, Murthy MRN (1997) Analysis of temperature factor distribution in high-resolution protein structures. *Proteins Struct Funct Genet* 6: 2561-2567.
37. Parthasarathy S, Murthy MRN (1998) Analysis of temperature factor distribution in high-resolution protein structures. *Proteins Struct Funct Genet* 7: 525.
38. Ringe D, Petsko GA (1986) Study of protein dynamics by X-ray diffraction. *Methods Enzymol* 131: 389-433.