Flexibility-Rigidity Index for Protein Flexibility Analysis

Kristopher Opron[1], Kelin Xia[2] and Guo-Wei Wei[1,2,3]

[1]Department of Biochemistry and Molecular Biology, [2]Department of Mathematics and [3]Department of Electrical and Computer Engineering, Michigan State University

Introduction

This work introduces the multiscale theory of continuum elasticity with atomic rigidity (CEWAR) for macro- molecular analysis. The essential idea of CEWAR is to formulate bulk and shear moduli as continuous functions of atomic information. A flexibility-rigidity index (FRI) is introduced to analyze macromolecular flexibility and rigidity in atomic detail. A fundamental assumption is that interactions determine protein structures, while the structures of protein and its environment determine protein functions, such as stability and flexibility. As such, flexibility analysis can be carried out without resorting to interaction Hamiltonians. The FRI reflects the topological connectivity of macro- molecular atoms or residues and characterizes the geometric compactness of the macromolecule. The proposed methods are applied to the flexibility analysis of proteins.

Methods

Consider a macromolecule of N residues where r_j is the position of *j*th residue. Let us denote $|r_i - r_j|$ the Euclidean distance between *i*th and *j*th residues. We assume that the correlation between *i*th particle and the *j*th particle has the form

$C_{ij} = \Phi(|r_i - r_j|;\eta_{ij})$, where and $\Phi(r;\eta_{ij})$ is the correlation kernel which is a monotonically decreasing function such as generalized

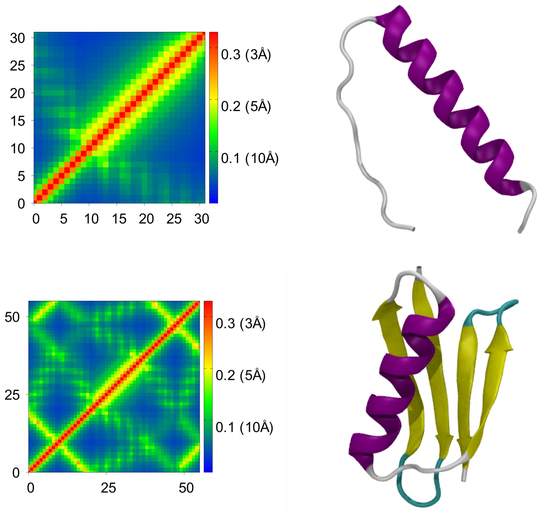
or generalized exponential function From the correlation values we calculate the atomic rigitity index for each residue:

$\mu_i =$, $\forall i = 1, 2...N$, where w_{ij} is a weigh The inverse of μ_i, i.e., $1/\mu_i$, is used to predict experimental B-factors and the interpolation of μ_i is used in CEWAR

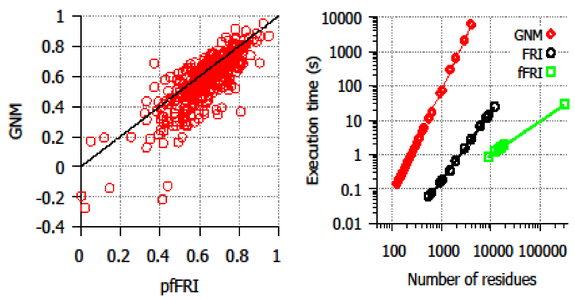
$\rho\ddot{\mathbf{w}} = \nabla(\lambda+ \mu) \nabla\cdot\mathbf{w} + \nabla\cdot\mu\nabla\mathbf{w} + \mathbf{f},$

where ρ is density, **w** the displacement, λ the Lame parameter, μ the shear modulus, and **f** the force.

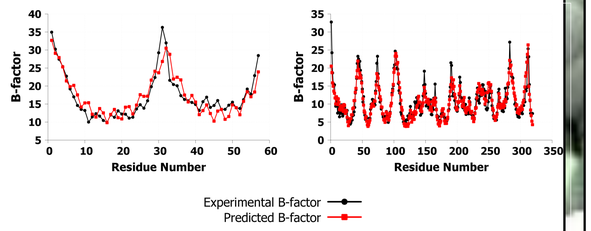
Results



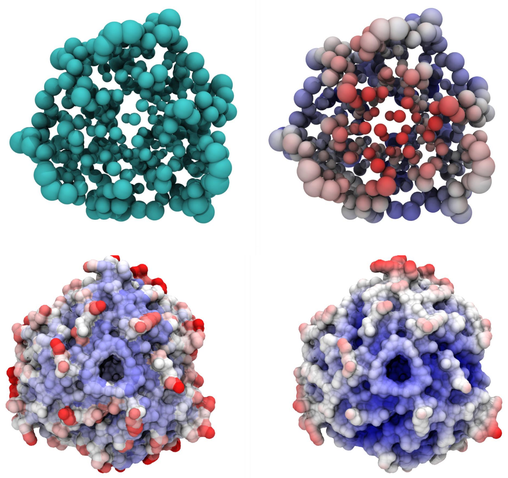
Correlation maps and secondary structure representations for proteins 1C26 and 1PGA. Correlation maps are generated using the power-law kernel with υ=2. Colors represent distance and correlation values for each pair of atoms. Red represents nearby atoms with high correlation values and blue represents distant atoms with low correlation values. The residue numbers for each Cα are listed along the x- and y-axes.



Left: Comparison of correlation coefficients from B-factor prediction using Gaussian network model (GNM) and parameter free FRI method for 365 proteins. The line *y=x* is included to aid in comparing scores.
Right: Efficiency comparison among GNM (red diamonds), FRI (black circles) and fast FRI (fFRI) (green squares). A set of 44 Cα only PDB files were used to evaluate the time complexity of GNM, FRI and fFRI.



Experimental B-factors (black) vs. predicted B-factors (red) using the exponential correlation kernel. The structures used are 1DF4 (left) and 2Y7L (right). For these comparisons, the optimal parameters were used for κ and η based on a parameter search.



Top: Cα atoms of Bacillus subtilis YabJ (PDB ID: 1QD9) in VDW representation scaled by flexibility index (left and right) and colored with electrostatics (right). Larger VDW radii represent more flexible atoms such as those near the surface of this soluble protein. Smaller VDW radii represent more rigid atoms such as those in the core of the protein

Bottom: The molecular surface of protein 1QD9 colored by B-factor (left) and continuous FRI representation (right). The flexibility index is calculated using the power-law kernel with υ=2. On the right, FRI is used to predict flexibility and the continuum representation is mapped to the protein surface. The continuum prediction matches the experimental flexibility pattern closely.

Conclusions

We propose flexibility-rigidity index (FRI) fast FRI (fFRI) methods for protein flexibility analysis. The fFRI is accurate and efficient for the prediction of flexibility and fluctuation of macromolecules compared to similar tools such as GNM. The average correlation score for B-factor prediction for 365 structures is 0.626 for fFRI vs. 0.565 for GNM. The fFRI scales with computational complexity as O(N), while others requiring matrix decomposition are approximately $O(N^3)$. FRI allows flexibility or rigidity to be visualized in either atomic discrete or atomic continuous representations of macromolecular structures. The continuous atomic rigidity from FRI is used in the multiscale modeling of continuum elasticity with atomic rigidity (CEWAR) and for visualization.

Acknowledgements

This work was supported in part by NSF grants IIS-1302285, CCF-0936830 and DMS-1160352, and NIH Grant R01GM-090208.

Reference

1) K. L. Xia, K. Opron, and G. W. Wei. Multiscale multiphysics and multi- domain models --- Flexibility and rigidity, Journal of Chemical Physics, 139, 194109, 2013.

2) Fast and anisotropic flexibility-rigidity index for protein flexibility and fluctuation analysis, Journal of Chemical Physics, 140, 234109, 2014.

Background image from gallery at: http://www.ks.uiuc.edu/Research/vmd/minitutorials/tachyonao/