Collective Motions in Protein Structures: Application of Elastic Network Models Built from Electron Density Distributions

Leherte, Laurence; Vercauteren, Daniel

Publication date:
2007

Document Version
Peer reviewed version

Collective Motions in Protein Structures: Applications of Elastic Network Models Built from Electron Density Distributions

Laurence Leherte, Daniel P. Vercauteren
Laboratoire de Physico-Chimie Informatique (PCI), University of Namur (FUNDP)

Abstract

Computational simulations of protein dynamics play an important role in deciphering protein functions, and usually require the knowledge of atomic coordinates. However, for a number of cases, one can only obtain fuzzy images of the molecules by means of experiments. Therefore, a question is whether one can describe the motion of a protein, at least the principal features, based on such images.

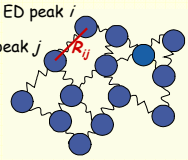
It has recently been shown that it is feasible to extract information about protein motions, at a reasonable degree of accuracy, without knowing the precise amino acid sequence. The models that are used, such as the Gaussian Network Model (GNM) and the Anisotropic Network Model (ANM), operate under the fundamental assumption that a folded protein can be viewed as an elastic network [1-2]. Numerous Web servers are now available to easily and rapidly evaluate the slow and large-magnitude dynamics of protein structures [3-9].

The present work consists in studying the dynamics of protein structures using topological and structural informations contained in low-resolution promolecular electron density distributions. Dynamical information are obtained from two approaches. The first one consists in building networks from ED maxima calculated at various smoothing levels [10]. The second approach also considers ED networks, with edges weighted by ED overlap integral values.

1. C_α- and ED-Based Protein Networks

C_α or ED peak *i*

C_α or ED peak *j*



Promolecular ED distribution

$$\rho_{M,t}(\mathbf{r}) = \sum_{a}^{\text{No. of atoms}} \rho_{a,t}(\mathbf{r}-\mathbf{R}_a)$$

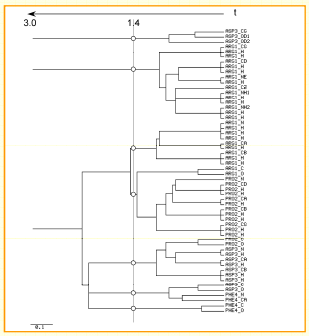
$$\rho_{a,t}(\mathbf{r}-\mathbf{R}_a) = \sum_{i=1}^{3} \alpha_{a,i} e^{-\beta_{a,i}|\mathbf{r}-\mathbf{R}_a|^2}$$

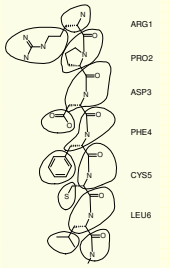
t = smoothing factor

A hierarchical merging algorithm is used to lacate maxima (peaks) in ED distribution, and to define corresponding molecular fragments.

At *t* = 1.4 bohr², fragments are located on the backbone and residue side chains.

Illustration is given for the first amino acids of structure *5pti*.





2. Kirchhoff Matrix Γ of a Protein Network

Laplacian Matrix

$$\Gamma_{ij} = \begin{cases} -1, & \text{if } i \neq j \text{ and } R_{ij} \leq r_c \\ 0, & \text{if } i \neq j \text{ and } R_{ij} > r_c \\ \#neighbors & \text{if } i = j \end{cases}$$

i and *j* are either C_α or ED peaks at *t* = 1.4 bohr²

ED Overlap Integral Matrix

$$\Gamma_{ij,t} = -\int d\mathbf{r}\, \rho_{i,t}(\mathbf{r})\, \rho_{j,i}(\mathbf{r})$$

calculated from atom content of the protein fragments that are associated with the ED peaks at *t* = 1.4 bohr²

3. Gaussian Network Model

$\mathbf{\Gamma}^{-1} = \mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}^T$ where $\mathbf{\Lambda}$ is a diagonal matrix of eigenvalues of $\mathbf{\Gamma}$ (frequencies²), and $\mathbf{U}$ is the matrix of eigenvectors.

Fluctuation of node *i* is given by: $\langle \Delta\mathbf{R}_i . \Delta\mathbf{R}_i \rangle = \frac{k_b T}{\gamma} \left[\mathbf{\Gamma}^{-1}\right]_i$ $\quad B_i = 8\pi^2 \langle \Delta\mathbf{R}_i . \Delta\mathbf{R}_i \rangle$

4. Anisotropic Network Model

$$\mathcal{H} = \begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} & \dots & \mathbf{H}_{1n} \\ \mathbf{H}_{21} & \dots & \dots & \mathbf{H}_{2n} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \mathbf{H}_{nn} \end{pmatrix}$$

where:

$$\mathbf{H}_{ij} = \frac{\gamma \Gamma_{ij}}{r_{ij}^2} \begin{pmatrix} (X_i - X_j)(X_i - X_j) & (X_i - X_j)(Y_i - Y_j) & (X_i - X_j)(Z_i - Z_j) \\ \dots & (Y_i - Y_j)(Y_i - Y_j) & (Y_i - Y_j)(Z_i - Z_j) \\ \dots & \dots & (Z_i - Z_j)(Z_i - Z_j) \end{pmatrix}$$

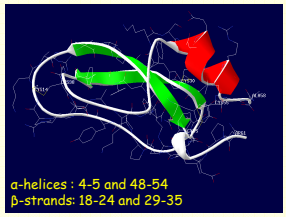
$\mathcal{H}^{-1} = \mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}^T$

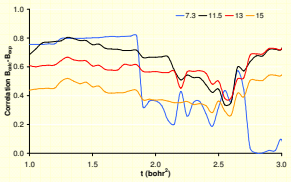
Total fluctuation of node *i* is given by: $< \Delta\mathbf{R}_i . \Delta\mathbf{R}_i > = \frac{3k_b T}{\gamma} tr[\mathbf{H}^{-1}]_{ii}$

Fluctuation of node *i* at mode *k* is given by: $(\Delta\mathbf{R}_i . \Delta\mathbf{R}_i)_k = \frac{3k_b T}{\gamma} tr[\lambda_k^{-1}\mathbf{U}_k\mathbf{U}_k^T]_{ii}$

5. Network Parameters for Pancreatic Trypsin Inhibitor (5pti)

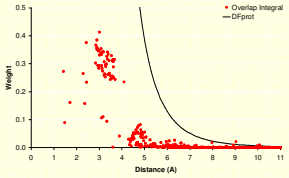
r_c = 7.3 Å (GNM) or 11.5 Å (ANM)





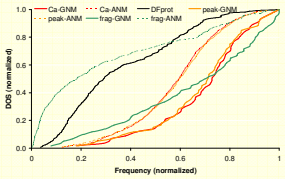
Correlation factors between B_{calc} (obtained using ANM applied to ED backbone peaks) and B_{exp}, as a function of *t*, at various cutoff values r_c.

Values of the overlap integral *vs.* peak-peak distance (red), and DFprot [8] force constant γ *vs.* C_α-C_α distance (black).



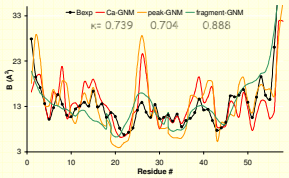
6. Density of States

Normalized DOS obtained for the C_α and ED-based networks of structure *5pti*.

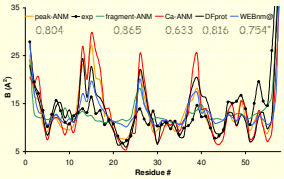


7. Residue Fluctuations

Isotropic



Anisotropic

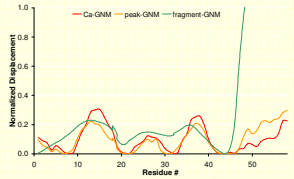


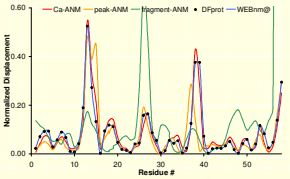
Better correlation value κ obtained with fragment-GNM approach (*e.g.*, reduced fluctuation of LYS26), but smoother profile.

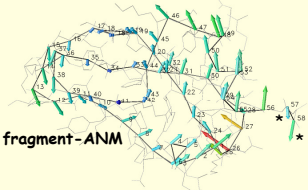
*calculation over first 6 modes

8. Residue Displacement Vectors (mode #1)

$\Delta\mathbf{r} = \mathbf{U}^T \mathbf{\Lambda}^{-1/2}$







Similar domains but different directions of fluctuations

*Amplitude reduced by 10

References

1- T. Haliloglu, I. Bahar, B. Erman, Phys. Rev. Lett. 79 (1997) 3090.
2- A. R. Atilgan, S. R. Durell, R. L. Jernigan *et al.*, Biophys. J. 80 (2001) 505.
3- K. Suhre, Y.-H. Sanejouand, Nucleic Acids Res. 32 (2004) W610 ; http://www.igs.cnrs-mrs.fr/elnemo/
4- Z. W. Cao, Y. Xue, L.-Y. Han *et al.*, Nucleic Acids Res. 32 (2004) W679 ; http://ang.cz3.nus.edu.sg/cgi-bin/prog/norm.pl
5- L.-W. Yang, X. Liu, Ch. J. Jursa *et al.*, Bioinformatics 21 (2005) 2978 ; http://ignm.ccbb.pitt.edu/
6- S. M. Hollup, G. Salensminde, N. Reuter, BMC Bioinformatics 6 (2005) 52 ; http://www.bioinfo.no/tools/normalmodes
7- E. Eyal, L.-W. Yang, I. Bahar, Bioinformatics 22 (2006) 2619 ;http://ignmtest.ccbb.pitt.edu/cgi-bin/anm/anm1.cgi
8- J. I. Garzón, J. Kovacs, R. Abagyan *et al.*, Bioinformatics 23 (2007) 901 ; http://sbg.cib.csis.es/Software/DFprot
9- Iowa State University, Plant Science Institute, Laurence H. Baker Center for Bioinformatics and Biological Statistics ; http://gor.bb.iastate.edu/anm/anm.htm
10- L. Leherte, Acta Cryst D 60 (2004) 1254.