Elastic Network
Modeling of Proteins

Jerome Nilmeier
BBSI 2004
Mentor: Ivet Bahar
The Challenge with Proteins...

- Proteins exist in *Dynamic Equilibrium*
- The paradigm:
  - Sequence → Structure → **Dynamics** → Function
- Dynamics can be inferred from an ensemble of structures
- It is generally regarded as a difficult computational problem

Computational Methods of Exploring Protein Motions

- **Monte Carlo**
  - Faster, Guaranteed to produce thermodynamic ensembles
  - Fundamental Method in Molecular Modeling
  - Disadvantages: Difficult to find appropriate move sets
  - Scales roughly as $N(N-1)/2 n$, ($N = \# \text{of atoms}; n=\# \text{ of samples}$)

- **Molecular Dynamics**
  - Advantages: Can give continuous trajectories
  - Disadvantages: Much more time intensive
  - Incomplete sampling / More susceptible to kinetic trapping
  - Scales as $(N(N-1)/2) T$, (T= # of uncorrelated timesteps)

**FOR MORE INFO...**

*Molecular Modelling, Principles and Applications, A.R. Leach*
*Molecular Modeling and Simulation, an Interdisciplinary Guide*  T. Schlick
Methods of Overcoming Energetic Barriers

- Detailed Balance Methods
  - Simulated Annealing
  - Replica Exchange

- Restraint Based Methods
  - Umbrella Sampling
  - Thermodynamic Integration
  - Steered Molecular Dynamics
Principal Component Methods

- Relies on the fact that directionality of fluctuations in a dataset can be identified through diagonalization of a matrix.

- Diagonalization of a matrix will yield a new set of dimensions that will represent the natural coordinates of your dynamical system.
Recall the ellipse equation from lecture:

\[ r^2 = 2(x^2 + y^2 - xy) \]

In matrix notation, this becomes:

\[ r^2 = (x \ y) \cdot \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} \]

The Eigenvalues and eigenvectors of this system are given as:

\[ \Lambda = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \]

\[ T = \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \]

The eigenvectors point in the direction of the principal axes, and the inverse eigenvalues give the magnitude of the axis.
Principal Axes of a Gaussian Function

Consider the 2D Gaussian DF:

\[ f(x, y) = \exp \left[ -2 \left( \frac{x^2 + y^2 - xy}{\sigma^2} \right) \right] \]

We can decompose this function in the same way to locate principal axes of the fluctuations.
A Gaussian Results from a Harmonic Potential

- From Prof. Meirovitch’s Lecture, we know that a system at Temperature T will generate the following distribution:

\[ p(x) = N \cdot \exp\left(-\frac{U(x)}{kT}\right) \]

- If the potential is harmonic, it will have the same form as a hyperellipsoid, which will result in a probability distribution that is Gaussian.
The Gaussian Network Model

- Assumes that a protein is in a near native state
- Residues interact with each other through a network of (harmonic) springs attached at \( \alpha \) carbons
- Advantages: Large scale fluctuations are easily accessible and accurate using a coarse-grained model

FOR MORE INFO...


Graphics courtesy of Prof. I. Bahar
Force Balance on a Network of Beads

Force Balance:

\[ f_{i,i} = \sum_{j=1}^{N} f_{i,j} = -\sum_{j=1}^{N} \gamma \Delta R_{i,j} \]

In Matrix Notation:

\[ \gamma \Gamma \Delta R = 0 \]

Where the Kirchhoff Matrix can be defined from the force balance:

\[ \Gamma = \begin{pmatrix}
  2 & -1 & 0 & -1 & 0 & 0 \\
  -1 & 2 & -1 & 0 & 0 & 0 \\
  0 & -1 & 2 & -1 & 0 & 0 \\
  -1 & 0 & -1 & 3 & -1 & 0 \\
  0 & 0 & 0 & -1 & 2 & -1 \\
  0 & 0 & 0 & 0 & -1 & 1 \\
\end{pmatrix} \]

\[ \Gamma_{i,j} = -1 \quad \text{If } r<r_c \]

\[ \Gamma_{i,i} = -\sum_{j=1}^{N} \Gamma_{i,j} \]
Potentials and Fluctuations

The potential is now easily defined:

\[ V = \Delta R^T \Gamma \cdot \Delta R = \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix} \cdot \begin{pmatrix} \Gamma \cdot \Delta R \end{pmatrix} \]

Does this look like our ellipse problem?

\[ \Delta R_i \Delta R_j = \frac{1}{Z} \int \exp\left( \frac{-V}{k \cdot T} \right) d\Delta R = \frac{3 \cdot kT}{\gamma} \left( \Gamma^{-1} \right)_{i,j} \]

This has terms you may not be familiar with..don’t worry…

When expressed in terms of the eigenvalue decomposition

\[ \Delta R_i \Delta R_j = \sum_{k=1}^{N} \frac{3 \cdot kT}{\gamma} \cdot (\lambda_k)^{-1} \cdot \left( u_k \cdot u_k^T \right)_{i,j} \]

Each summand is called the kth mode
Comparison with X-Ray Temperature Factors

The variance is computed from:

\[(\Delta R)^2 = \frac{3 k \cdot T}{\gamma} (\Gamma^{-1})_{i,i}\]

The DeBye-Waller temperature factor is given by:

\[B_i = \frac{8\pi^2}{3} (\Delta R_i)^2\]

Validation of the model against structural data shows that a coarse-grained can be quite accurate.

FOR MORE INFO...


Graphics courtesy of Prof. I. Bahar
Domains are distinguished by different mobilities

- Structural domains undergo ‘en bloc’ motions
- Hinge sites correspond to minima in global mode shapes
- Substrate recognition sites usually form maxima in global modes
- Core contacts correlate with peaks in local modes

FOR MORE INFO...


Graphics courtesy of Prof. I. Bahar
iGNM

ignm.ccbb.pitt.edu

- A database of Structures that have been analyzed using the Gaussian Network Model
- Provides a color-coded picture of the protein, identifying fluctuating residues
- Source code is freely available for use
Influenza virus hemagglutinin dynamics – HCG
Examples of further coarse-graining

Isin, Doruker & Bahar (2002) Biophysical J.

Computational efficiency: Computing time reduced by \(m^2\), where \(m\) is the rescaling factor

Graphics courtesy of Prof. I. Bahar
An extension: The Anisotropic Network Model

- Derive the B matrix using force balance, as before, but now we will keep track of each component:

\[
\begin{pmatrix}
  f_{i,i} \\
  f_{i,i} \\
  f_{i,i}
\end{pmatrix}
\begin{pmatrix}
  \cos(\alpha_x) \\
  \cos(\alpha_y) \\
  \cos(\alpha_z)
\end{pmatrix}
= -\gamma \sum_{j=1}^{N}
\begin{pmatrix}
  \Delta X_{i,j} \\
  \Delta Y_{i,j} \\
  \Delta Z_{i,j}
\end{pmatrix}
\]

- In matrix notation, this becomes:

\[
B_{3N\times M} \cdot f_{N\times 1} = 0
\]

- The solution to this equations allows us to look at dominant modes, but we now have a vector associated with it

FOR MORE INFO...

Bahar, et al. (2001) Biophys. J. 80, 505
The Anisotropic Network Model:
Motions of the ribosome

Graphics courtesy of Prof. I. Bahar
HK97 Bacteriophage capsid

Capsid chainmail- catenated circle topology

FOR MORE INFO...


Graphics courtesy of Prof. I. Bahar
Acknowledgements

- Dr. Ivet Bahar and the Bahar Group
- BBSI Faculty and Students
- Dr. Rajan Munshi and the BBSI Program
- U. Pitt., Duquesne, and CMU
- NIH, NSF
- UCSF, the Jacobson and Dill Groups