Improving Elastic Network Models of Protein Fluctuations with Microsecond-Scale Molecular Dynamics

Nicholas Leioatts, Tod D. Romo, Alan Grossfield
University of Rochester Medical School, Rochester, NY, USA

Abstract

Elastic Network Models (ENMs) are simple harmonic models that describe the collective motions of a biomolecule in its native structure. Despite their simplicity, there are a number of ways to improve their predictive power. Here, we systematically optimize and test several spring functions by comparing their predictions to long timescale molecular dynamics. Our results show that combining a sophisticated distance-dependent spring function with parametrization against μs-scale all-atom molecular dynamics can dramatically improve the value of ENM calculations. Moreover, increasing their spatial resolution by including sidechains also produces a statistically significant improvement while allowing better comparison to experiment.

Elastic Network Models

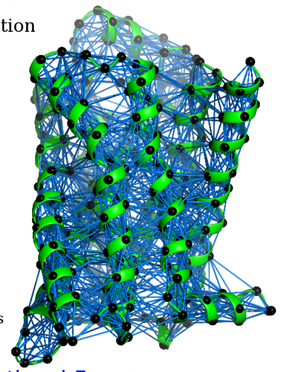
- Coarse-grained model, Cα resolution
- "Beads on springs"
- Single harmonic potential:

$$U_{ij} = k(r_{ij})\left(|r_{ij}| - |r^{\circ}_{ij}|\right)^2$$

$$k(r_{ij}) = \begin{cases} 1 & : r_{ij} < r_c \\ 0 & : r_{ij} \geq r_c \end{cases}$$

-k is a uniform spring constant
-r°_{ij} minimum energy - starting structure

- Diagonalize Hessian Matrix
-Yields eigenpairs
·Eigenvalues describe frequency
·Low frequencies → collective dynamics
·Eigenvectors describe direction



Alternative Functional Forms

Distance-dependence models → tighter coupling between nearby beads

Name	Description	Equation
Standard	Heavyside function	$k(r_{ij}) = \begin{cases} 1 & : r_{ij} < r_c \\ 0 & : r_{ij} \geq r_c \end{cases}$
Exponential	Constant decays exponentially	$k(r_{ij}) = ar_{ij}^{-b}$
Distance*	Function distance dependence	$k(r_{ij}) = \begin{cases} ar_{ij} + b & : r_{ij} < r_c \\ cr_{ij}^{-d} & : r_{ij} \geq r_c \end{cases}$
Bonded	Explicit connectivity	$k(r_{ij}) = \begin{cases} ar_{ij} + b & : \text{Bonded} \\ cr_{ij}^{-d} & : \text{Non} - \text{bonded} \end{cases}$

*Hinsen et al, Chem Phys (2000), 261: 25-37

Accessing Collective Motions

- How well do ENMs reproduce dominant fluctuations of MD?
- Covariance Overlap:

$$\Omega_{A,B} = 1 - \left[\frac{\sum_i^N(\lambda_i^A + \lambda_i^B) - 2\sum_i^N\sum_j^N\sqrt{\lambda_i^A\lambda_j^B}(\vec{v}_i^A \cdot \vec{v}_j^B)^2}{\sum_i^N(\lambda_i^A + \lambda_i^B)}\right]^{\frac{1}{2}}$$

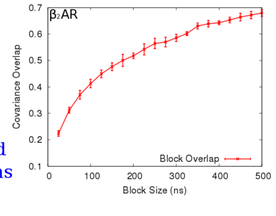
- Compare ENM to MD
-Eigenvalue weighted projection of eigenvectors
-Considers magnitude of motions as well as direction
-Quantifies the difference between collective motions
-Use inverse eigenvalue from ENM

-Quantify similarity of modes
-Scales [0:1]
-1 is complete overlap
-0 is completely orthogonal

Hess, Phys Rev E (2000), 62, 8438–48

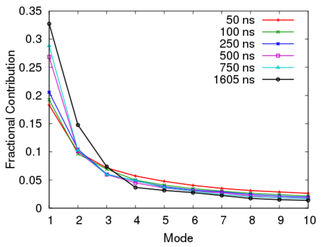
Block Covariance Overlap

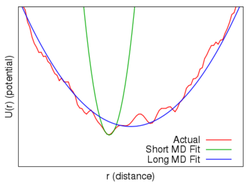
- Compare collective motions
-Whole trajectory vs. blocks
-Covariance vs. block length
-Overlap increases with sampling time
- ENMs vs. MD
-MD highly detailed, statistical errors
-ENM very simple, no statistical error
- How much simulation time needed to reproduce dominant fluctuations with better accuracy than ENM?



Fitting Requires Long Trajectory

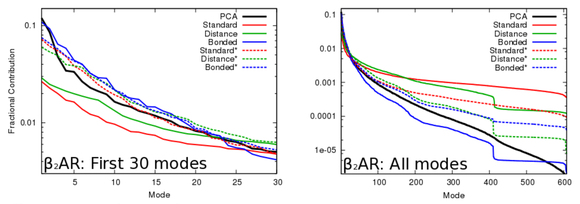
- Shorter trajectories sample less conformation space
-Statistical error
-Narrower potential well





- Low mode contributions increase with trajectory length
-Systematic error
-Example: Rhodopsin
- Has the system converged?

Low Modes Govern Power Spectra



- Power spectrum shows contribution of each mode to total motion
-Standard ENMs too stiff
-More sophisticated spring functions improve match
-Fitting improves power spectrum

- ENMs underestimate low frequency contributions
-Similar to short MD
- High frequency modes dramatically overestimated

Block Covariance Quantifies Value of ENMs

- Compare ENMs to μs MD
-Use covariance overlap
-Intersection is equivalent *accuracy*

- Test using 3 class A GPCRs
-β2 Adrenoreceptor (β2AR)
-1.02 μs simulation
Romo et al., Biophys J (2010), 98(1):76–84

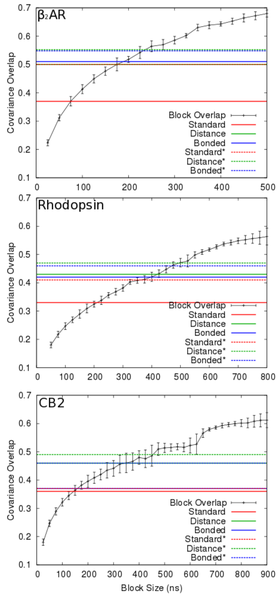
-Rhodopsin
-1.6 μs simulation
Grossfield et al., JMB (2008), 38(2):478–86

-Cannabinoid Receptor 2 (CB2)
-1.8 μs simulation
Hurst et al., JBC (2010), 285:17954-64

- Fit using β2AR & Rhodopsin
- Test with CB2
-Covariance overlap improves

- Network Model *value* increases
-ENMs equivalent to up to 450 ns of all-atom simulation

- For more information see:
Romo & Grossfield, Proteins (2011), 79:23-34

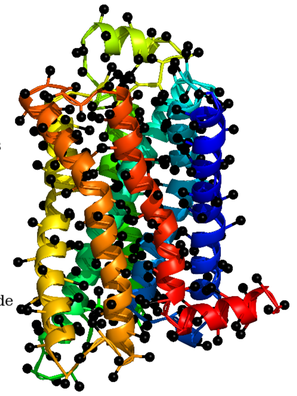


Increasing Spatial Resolution

- 2 beads per residue:
-Cα bead represents backbone
-C_{CoM} bead represents side chain
·Placed at side chain center of mass

- Approach
-Systematically fit many spring functions using molecular dynamics
-Bonded vs. non-bonded springs
-Distance-dependent spring function

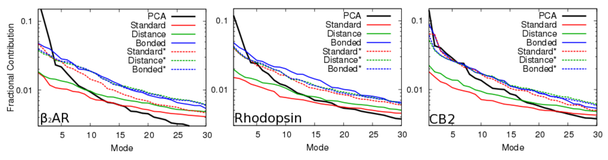
- Applications
-Study ligand binding via ENMs
·Most network models too coarse-grained
·Ligand usually represented by a single node
-Cysteine scanning analysis
·Manually place "disulfide" bonds



Low Mode Contributions

- Fit similar to 1 bead per residue network model
- Distance-dependent functions subtly better

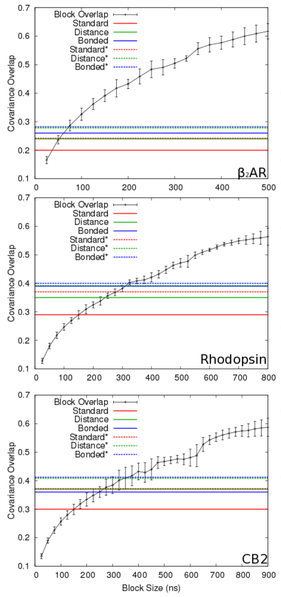
- Bonded spring function
-Good power spectrum
-Fitting doesn't improve



Comparison to Previous Results

- Overlap with MD for each protein
-CB2 not fit
-Fitting improves CB2
- 2 bead model has lower overlap

	1-bead			2-bead		
	β_2AR	Rhodopsin	CB2	β_2AR	Rhodopsin	CB2
Standard	0.37	0.33	0.36	0.20	0.29	0.30
Standard*	0.50	0.41	0.37	0.24	0.37	0.37
Distance	0.50	0.43	0.46	0.24	0.35	0.37
Distance*	0.55	0.47	0.49	0.28	0.39	0.41
Bonded	0.51	0.42	0.37	0.26	0.39	0.36
Bonded*	0.55	0.46	0.46	0.28	0.40	0.41



- Higher resolution ENM tested and fit with same trajectories

- β2AR and Rhodopsin optimized
- Test with CB2

- Block average of MD lower
-Longer blocks required for the same sampling quality
-Side chains make sampling harder

- Rhodopsin and CB2 show similar results as standard ENM
-β2AR covariance overlap lower

- Distance-dependent funtions & fitting improve results

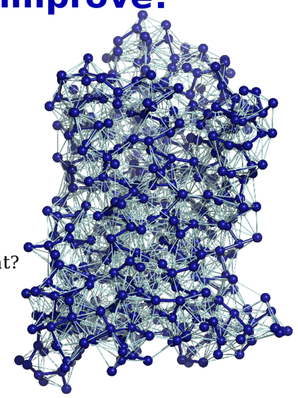
- ENM motions have equivalent accuracy to 100-350 ns simulation

Did we improve?

- Higher Dimensional Problem
-Contact matrix doubles

- Covariance overlap lower
-Is the problem harder?
·Density of beads
·Side chains may not fluctuate harmonically

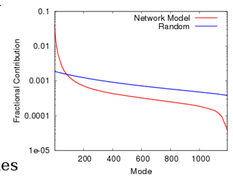
- How to measure improvement?
-Need a method that accounts for higher dimensionality



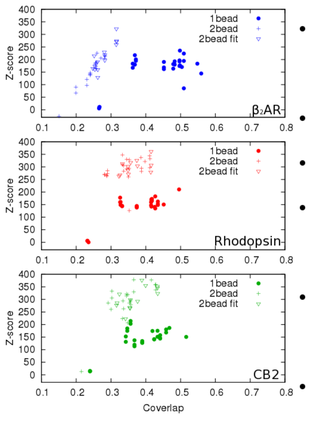
Statistical Significance of Covariance Overlaps

- Want to use a bootstrap-like approach

- Compare to random contact matrices
-Power spectrum qualitatively different
-Covariance overlap always very large
-Not a useful approach



- Scramble eigenpairs
-Randomly assign eigenvectors to eigenvalues



- Calculate Standard Score:

$$Z = \frac{x - \bar{x}}{\sigma}$$

- Plot as function of covariance overlap
- Many spring functions shown

- β2AR: 2-bead model has lower covariance overlap but same Z-score

- Rhodopsin and CB2: 2-bead has same covariance overlap, higher Z-score

- Higher resolution models statistically better

Continuing Progress

- Long MD required for correct power spectrum
- ENMs worth significant amount of MD sampling
- Can improve ENMs
-Distance-dependent spring function
-Fitting
- Higher resolution models
-Harder problem
-Statistically improves results

- Develop & inform experiment
-Spring functions for
·Cysteine scanning
·Ligand binding analysis
- Test with 10 μs simulation
- Test effect of order of magnitude changes on ENMs

Poster available online:
tinyurl.com/validating-enm

Work done in LOOS (Lightweight Object Oriented Structure analysis library), an open source C++ library designed and maintained by the Grossfield lab. LOOS provides a concise, adaptable framework for designing analysis tools that interfaces with native file formats of most simulation packages.

http://loos.sourceforge.net