Improving Protein Function Prediction with Molecular Dynamics Simulations

Dariya Glazer
Russ Altman
Motivation
Sometimes the 3D structure doesn’t score well for a known function.
The experimental structure may have an altered configuration of atoms.

Possible Solution
Treat molecules as dynamic entities, sample other conformations and test for function
Potential Energy function:

\[
U = \sum_{\text{bonds}} K_b (b_i - b_o)^2 \\
+ \sum_{\text{angles}} K_\theta (\theta_i - \theta_o)^2 \\
+ \sum_{\text{dihedrals}} K_\phi [1 - \cos (n \phi_i + \delta)] \\
+ \sum_{\text{pairs}} \varepsilon \left[ \left( \frac{r_o}{r_{ij}} \right)^{12} - 2 \left( \frac{r_o}{r_{ij}} \right)^6 \right] \\
+ \sum_{\text{charg es}} \frac{q_i q_j}{r_{ij}}
\]
Molecular Dynamics

- Given the equations for energy, and the time scale, perform a large numerical simulation to see how the molecule moves.

- Time step = 0.001 ps
- Starting point = experimentally defined structure (already in good local minima)
- Include water molecules around protein
- Goal: study dynamic motion of protein
Numerical Solution for M.D.

\[ x(t + \Delta t) = x(t) + v(t)\Delta t + \left[ 4a(t) - a(t - \Delta t) \right]\Delta t^2 / 6 \]

\[ v(t + \Delta t) = v(t) + \left[ 2a(t + \Delta t) + 5a(t) - a(t - \Delta t) \right]\Delta t^2 / 6 \]

\[ U_{\text{kinetic}} = \frac{1}{2} \sum m_i v_i(t)^2 \]

\[ U_{\text{potential}} = \ldots \]

\[ U_{\text{total}} = U_{\text{kinetic}} + U_{\text{potential}} \]

\[ F = \frac{dU}{dx} \]

\[ a = F / m \]
Molecules Come Alive

Static Structure: $n = 1$

Simulation Trajectory

Structural Diversity: $n >>>> 1$
Starting Static Structures: PDB

5 x HOLO

with bound ligand

5 x APO

without the ligand

Simulations:

GROMACS, duration 1ns, explicit solvation

Scoring Function:

FEATURE, resolution 1Å
Threshold score of this model is 50

<table>
<thead>
<tr>
<th>MODEL FEATURES</th>
<th>SHELL</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATOM-NAME-IS-ANY</td>
<td>0</td>
</tr>
<tr>
<td>ATOM-NAME-IS-C</td>
<td>1</td>
</tr>
<tr>
<td>ATOM-NAME-IS-N</td>
<td>2</td>
</tr>
<tr>
<td>ATOM-NAME-IS-O</td>
<td>3</td>
</tr>
<tr>
<td>ATOM-NAME-IS-S</td>
<td>4</td>
</tr>
<tr>
<td>ATOM-NAME-IS-OTHER</td>
<td>5</td>
</tr>
<tr>
<td>HYDROXYL</td>
<td></td>
</tr>
<tr>
<td>AMIDE</td>
<td></td>
</tr>
<tr>
<td>AMINE</td>
<td></td>
</tr>
<tr>
<td>CARBONYL</td>
<td></td>
</tr>
<tr>
<td>RING-SYSTEM</td>
<td></td>
</tr>
<tr>
<td>PEPTIDE</td>
<td></td>
</tr>
<tr>
<td>VDW-VOLUME</td>
<td></td>
</tr>
<tr>
<td>CHARGE</td>
<td></td>
</tr>
<tr>
<td>NEG-CHARGE</td>
<td></td>
</tr>
<tr>
<td>POS-CHARGE</td>
<td></td>
</tr>
<tr>
<td>CHARGE-WITH-HIS</td>
<td></td>
</tr>
<tr>
<td>HYDROPHOBICITY</td>
<td></td>
</tr>
<tr>
<td>MOBILITY</td>
<td></td>
</tr>
<tr>
<td>SOLVENT-ACCESSIBILITY</td>
<td></td>
</tr>
<tr>
<td>RESIDUE_NAME_IS_GLU</td>
<td></td>
</tr>
<tr>
<td>RESIDUE_NAME_IS_GLY</td>
<td></td>
</tr>
<tr>
<td>RESIDUE_NAME_IS_LEU</td>
<td></td>
</tr>
<tr>
<td>RESIDUE_NAME_IS_LYS</td>
<td></td>
</tr>
<tr>
<td>RESIDUE_NAME_IS_PRO</td>
<td></td>
</tr>
<tr>
<td>RESIDUE_NAME_IS_VAL</td>
<td></td>
</tr>
<tr>
<td>RESIDUE_NAME_IS_HOH</td>
<td></td>
</tr>
<tr>
<td>RESIDUE_CLASS1_IS_HYDROPHOBIC</td>
<td></td>
</tr>
<tr>
<td>RESIDUE_CLASS1_IS_CHARGED</td>
<td></td>
</tr>
<tr>
<td>RESIDUE_CLASS1_IS_POLAR</td>
<td></td>
</tr>
<tr>
<td>RESIDUE_CLASS1_IS_UNKNOWN</td>
<td></td>
</tr>
<tr>
<td>RESIDUE_CLASS2_IS_NONPOLAR</td>
<td></td>
</tr>
<tr>
<td>RESIDUE_CLASS2_IS_POLAR</td>
<td></td>
</tr>
<tr>
<td>RESIDUE_CLASS2_IS_ACIDIC</td>
<td></td>
</tr>
<tr>
<td>RESIDUE_CLASS2_IS_UNKNOWN</td>
<td></td>
</tr>
<tr>
<td>SECONDARY_STRUCTURE1_IS_3HELIX</td>
<td></td>
</tr>
<tr>
<td>SECONDARY_STRUCTURE1_IS_4HELIX</td>
<td></td>
</tr>
<tr>
<td>SECONDARY_STRUCTURE1_IS_BRIDGE</td>
<td></td>
</tr>
<tr>
<td>SECONDARY_STRUCTURE1_IS_STRAND</td>
<td></td>
</tr>
<tr>
<td>SECONDARY_STRUCTURE1_IS_TURN</td>
<td></td>
</tr>
<tr>
<td>SECONDARY_STRUCTURE1_IS_BEND</td>
<td></td>
</tr>
<tr>
<td>SECONDARY_STRUCTURE1_IS_COIL</td>
<td></td>
</tr>
<tr>
<td>SECONDARY_STRUCTURE1_IS_HET</td>
<td></td>
</tr>
<tr>
<td>SECONDARY_STRUCTURE2_IS_HELIX</td>
<td></td>
</tr>
<tr>
<td>SECONDARY_STRUCTURE2_IS_BETA</td>
<td></td>
</tr>
<tr>
<td>SECONDARY_STRUCTURE2_IS_COIL</td>
<td></td>
</tr>
</tbody>
</table>
**Simulation Analysis**

**Local Grid:** to observe dynamics of a known binding site
- Keep only the topmost score for each frame

**Global Grid:** to find all possible binding sites in a given structure
- Keep all scores over 50 for each frame
Local Grid: Non-Sites Score Low

1B9A LEU15N Ca Model

1K4 THR78CG2 Ca Model

Simulation  Original Score  Model Threshold
Results Agree With Experiments

1B9A CA109 Ca Model

1B8C CA109 Ca Model

Simulation  Original Score  Model Threshold
Results: 1DNK

Active Site of the molecule, Ca$^{2+}$ binds with the ligand
Exists in both HOLO and APO

MD helps FEATURE identify this site in both HOLO and APO
Results: 1MJW

Site opened up by mutation which breaks stable salt bridge
Exists only in APO

MD helps FEATURE identify this site in APO
Summary

TOTAL = 12 SITES

Static FEATURE
7 HOLO and 3 APO

MD + FEATURE
10 HOLO and 9 APO

1.4 – 3 Fold Improvement
Protein adopts functional conformation during simulation.
Simbios: National Center for Physics-Based Simulation of Biological Structure

- NIH-funded resource to promote biomedical computing (one of seven)
- Devoted to biological structural simulations at all scales

Two-fold mission:
- Perform high quality research
- Disseminate software and models to others

- Also publish quarterly magazine “Biomedical Computation Review”
Home

Physics-based Simulation of Biological Structures

About Simbios

Simbios Sites

SimK
Downloads

Simbiome

Simbiome

Biomedical Computation Review

Interested in Collaborating with Simbios?

Highlights of current and upcoming courses, symposiums and workshops

October 25th
Life in Motion Symposium
Clark Center, Stanford University

October 27th
BCATS Symposium
Clark Center, Stanford University

Past Events:

We released OpenSim on August 22nd, [view some of the pictures taken at the event](#).

We held a short course on SimVascular from August 27-31 -- the cardiovascular software tool kit.
Enabling groundbreaking biomedical research by providing open access to high-quality simulation tools, accurate models and the people behind them.

**About SimTK**

SimTK, the Simulation Toolkit, is part of the Simbios project funded by the National Institutes of Health. Learn more ...

**Related Sites**

- NIH Center for Physics-based Simulation
- Simbiome
- Biomedical Computation Review

**Where To Get Downloads**

**Applications and Models:** Free downloadable stand-alone simulation software and models

**Core Simulation Technology:** Free downloadable source code for the underlying algorithms and computational tools for simulations in a variety of biological application areas

**All Projects with Downloads:** All available free downloadable software and data

**Featured Project**

**Release of OpenSim 1.0:**

OpenSim is an open-source software system that lets users develop models of musculoskeletal structures and create dynamic simulations of movement.

Visit the OpenSim Project and download the new OpenSim 1.0 release.

Learn more about our latest OpenSim Workshop (8/22/07) on using OpenSim 1.0.
Covers of the four issues of *Biomedical Computation Review* from this past year.
Conclusions

- We can build statistical models of functional sites in proteins based on 3D structures.
- We can use these models to recognize function in new structures.
- We can extend the range of our models by adding physics-based simulation.

Challenges:
- Discovering and labeling new functions.
- Applying in high-throughput to all structures.
- Using for drug design/discovery.
Thanks!
russ.altman@stanford.edu

Mike Liang, Shirley Wu, Dariya Glazer,
Inbal Halperin, Jessica Ebert