Molecular Dynamics Flexible Fitting

Ryan McGreevy
Research Programmer

University of Illinois at Urbana-Champaign
NIH Resource for Macromolecular Modeling and Bioinformatics
Molecular Dynamics Flexible Fitting (Ribosome-bound YidC)

Electron Microscope

EM density map

Match through MD

Supercomputer

APS Synchrotron

Crystallographic structure
Two terms are added to the MD potential

\[ U_{total} = U_{MD} + U_{EM} + U_{SS} \]

An external potential derived from the EM map is defined on a grid as

\[ U_{EM}(\mathbf{R}) = \sum_j w_j V_{EM}(\mathbf{r}_j) \]

\[ V_{EM}(\mathbf{r}) = \begin{cases} \xi \left(1 - \frac{\Phi(\mathbf{r}) - \Phi_{thr}}{\Phi_{max} - \Phi_{thr}}\right) & \text{if } \Phi(\mathbf{r}) \geq \Phi_{thr}, \\ \xi & \text{if } \Phi(\mathbf{r}) < \Phi_{thr}. \end{cases} \]

A mass-weighted force is then applied to each atom

\[ \mathbf{f}_{i}^{EM} = -\nabla U_{EM}(\mathbf{R}) = -w_i \frac{\partial V_{EM}(\mathbf{r}_i)}{\partial r_i} \]

Harmonic restraints are applied to preserve secondary structure of proteins and nucleic acids, avoiding “overfitting.”

\[ U_{SS} = \sum_{restraints} k_\mu (\mu - \mu_0)^2 \]

For proteins, $\phi$ and $\psi$ dihedral angles of residues within helices or beta strands are restrained. For nucleic acids, distance and dihedral restraints are applied to a selected set of base pairs.
Additional Restraints

Cis-peptide and Chirality
Eduard Schreiner, et al. BMC Bioinformatics, 12, 190, 2011

Domain-wise

Symmetry

Current structure
Average positions of C-alpha atoms
Perfectly symmetric structure

Harmonic restraints
(strength increasing over simulation for convergence)

Acetyl – CoA Synthase

B. pumilus cyanide dihydratase

Symmetry restrained MDFF - Test Case 1

Improve quality of fit for low-resolution data

Blue: without symmetry restraints
Red: with symmetry restraints

低分辨率案例（8Å）
更好的结构（更低的RMSD）

高分辨率案例（4.3Å）
没有效果

Archaeal group II chaperonin from *M. maripaludis* (Mm-cpn)
8-fold rotational + 2 fold reflection symmetry
homology model (based on PDB 3LOS) fitted into EM map (EMDB 5140)

Finite-size Simulation (9 dimers)

helical symmetry

Fitted models of J1 nitrilase from *R. rhodochrous*

homology model and EM map (EMD 1313) from collaborator T. Sewell, U. of Cape Town

Without Symmetry Restraints
With Symmetry Restraints

Domain restrained MDFF

Use Targeted MD (TMD) feature of NAMD to restrain non-overlapping groups of atoms to maintain rigid domains.
MDFF can be run in different environments:

1. **Vacuum**
   - No water molecules
   - Fastest but potentially inaccurate

2. **Explicit Solvent**
   - Explicit atomic detail water molecules
   - Computationally slow and introduces effects of viscous drag

3. **Implicit Solvent**
   - Generalized Born approximation of electrostatics
   - Compromise between speed and accuracy

MDFF Software Suite

- NAMD and VMD used together to run MDFF
- Every NAMD and VMD feature is available in MDFF

Fitting time is dependent on:
- system size
- map and structure quality
- YidC: 6 months to model; 6 hours MDFF (20ns); workstation

Input: MDFF only requires a PDB, PSF, and density map
Output: produces simulation trajectory from which an ensemble of structures can be extracted

http://www.ks.uiuc.edu/Research/mdff/
New MDFF GUI (VMD 1.9.2) makes setting up, running, and analyzing fitting simulations even easier.

- system sizes up to 100 million atoms (viruses, chromatophore)
- maps from 3 to 15 Å
- runs on laptops to petascale computing resources (Blue Waters, Titan)

http://www.ks.uiuc.edu/Research/mdff/
Molecular Dynamics Flexible Fitting - Example

Cryo-EM map of the *E. coli* ribosome at 6.7-Å resolution
Obtaining Initial Structures

1. X-ray crystallography or NMR structures

2. Refine structures from low-res X-ray data with xMDFF

3. Homology or ab initio modeling with Modeller, Rosetta, MUFOOLD (Ci-VSP, YidC, Holotranslocon)

Rosetta structure prediction to fill **missing pieces** and MDFF to filter, refine and validate candidate structures
Rosetta/MDFF interactive Modeling

Combining structure prediction with the user’s expertise to interpret densities

**incomplete structural model deposited in the PDB**

- **de novo** structure prediction
- energy ranking
- model filtering
- interactive MDFF of cryo-EM data

**complete structural model that fits cryo-EM data**

**Rosetta**

Leaver-Fay *et al.* Methods Enzymol. 2011

**VMD/NAMD**

Humphrey *et al.* J. Mol. Graph. 1996
Interactive Modeling with MDFF GUI

- Apply forces to manually manipulate structure into the density
- Useful for difficult to fit structures with large conformational changes

New MDFF GUI in VMD 1.9.2

Set up and run interactive (or traditional) MDFF/xMDFF simulations

Analyze interactive simulations in real-time
xMDFF: MDFF for low-resolution x-ray crystallography

Diffraction data → Amplitude from X-ray diffraction intensity → Phase from search model → X-ray electron density using PHENIX → Model-Phased Maps

xMDFF iteration

Convergence test

R-value > tolerance → R-value < tolerance

Final model

Final Model

Initial Target Final

xMDFF Improves Structures Posted at the Protein Data Bank

<table>
<thead>
<tr>
<th>PDB ID</th>
<th>Molprobity</th>
<th>R-work</th>
<th>R-free</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>initial</td>
<td>final</td>
<td>initial</td>
</tr>
<tr>
<td>1AV1</td>
<td>3.72</td>
<td>1.94</td>
<td>0.38</td>
</tr>
<tr>
<td>1YE1</td>
<td>2.68</td>
<td>1.89</td>
<td>0.25</td>
</tr>
<tr>
<td>1JL4</td>
<td>3.24</td>
<td>1.47</td>
<td>0.36</td>
</tr>
<tr>
<td>1AOS</td>
<td>3.40</td>
<td>2.45</td>
<td>0.21</td>
</tr>
<tr>
<td>1XDV</td>
<td>2.87</td>
<td>2.01</td>
<td>0.39</td>
</tr>
<tr>
<td>1YI5</td>
<td>3.08</td>
<td>1.73</td>
<td>0.27</td>
</tr>
</tbody>
</table>

- Better R-work and R-free values than published before.
- Close R-work and R-free implies less over-fitting.
- Improved geometry implied by low Molprobity score.

Refinement statistics

xMDFF Solves Voltage Sensor Protein Structure at 4 Å Resolution

Collaboration with E. Perozo (U. Chicago)

• xMDFF reproduces helix position and arginine alignment.
• Refined model confirms electrophysiological measurements.

Search model preparation
Largest xMDFF structure has 2252 amino acids
Search model used from MUFOLD structure prediction software
(Dong Xu U. Missouri)

Refinement statistics

<table>
<thead>
<tr>
<th></th>
<th>initial</th>
<th>final</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>0.45</td>
<td>0.26</td>
</tr>
<tr>
<td>R-free</td>
<td>0.47</td>
<td>0.28</td>
</tr>
<tr>
<td>score</td>
<td>3.07</td>
<td>2.19</td>
</tr>
<tr>
<td>helix RMSD</td>
<td>4.65</td>
<td>1.34</td>
</tr>
</tbody>
</table>

Qufei Li, et. al. Nature Structural & Molecular Biology, 21:244-252, 2014
xMDFF for Abiological Materials
Cyanostar (2Å)

Importance of Checking Initial Structure

<0.05% non-proline bonds found in the cis conformation natively, however:

- The frequency of non-proline cis-peptide bond errors has been increasing for low-resolution

- These errors can hide issues in other parts of the structure

Wrong chirality, cis-peptide bonds, and torsion angle outliers may arise during modeling.

VMD provides tools to check, visualize, and correct these errors.

These tools, together with MD force fields, produce models with good structural geometry.

TorsionPlot Plugin new in VMD 1.9.3
Cascade MDFF for **high-resolution** cryo-EM:
Successively higher resolution maps

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Global Cross-correlation</th>
<th>RMSD (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>0.732</td>
<td>-</td>
</tr>
<tr>
<td>Direct</td>
<td>0.699</td>
<td>12.41</td>
</tr>
<tr>
<td>Cascade</td>
<td>0.724</td>
<td>2.30</td>
</tr>
</tbody>
</table>

Analyzing MDFF Model Quality 0: Known Structures

MDFF has been validated against a wide-ranging set of known high-resolution structures.
Analyzing MDFF Model Quality 1: Structure Checking

Eduard Schreiner, et al. BMC Bioinformatics, 12, 190, 2011

- Wrong chirality, cis-peptide bonds, and torsion angle outliers may arise during modeling

- VMD provides tools to check, visualize, and correct these errors

- These tools, together with MD force fields, produce models with good structural geometry

<table>
<thead>
<tr>
<th>PDB ID</th>
<th>Molprobity initial (published)</th>
<th>Molprobity final</th>
</tr>
</thead>
<tbody>
<tr>
<td>1AV1</td>
<td>3.72</td>
<td>1.94</td>
</tr>
<tr>
<td>1YE1</td>
<td>2.68</td>
<td>1.89</td>
</tr>
<tr>
<td>1JL4</td>
<td>3.24</td>
<td>1.47</td>
</tr>
<tr>
<td>1AOS</td>
<td>3.40</td>
<td>2.45</td>
</tr>
<tr>
<td>1XDV</td>
<td>2.87</td>
<td>2.01</td>
</tr>
<tr>
<td>1YI5</td>
<td>3.08</td>
<td>1.73</td>
</tr>
</tbody>
</table>

Global CC is not always a good indicator of fit

CC = 0.728
RMSD(reference) = 6.23 Å

CC = 0.723
RMSD(reference) = 2.30 Å

Analyzing MDFF Model Quality 2: Local Cross Correlation

- Local cross correlation indicates quality of fit of specific regions across the entire structure
- New parallel CPU and GPU algorithms provide significant speed up (25-50x speedup over Chimera), allowing for fast computation along fitting trajectories

Structure is colored by cross correlation, along with Timeline analysis of the trajectory

Good Fit

Intermediate Fit

Bad Fit

Analyzing MDFF Model Quality 3: Local Resolution Analysis

Local resolution of the experimental density from ResMap for error analysis and simulation parameterization

Root Mean Square Fluctuation (RMSF) correlates highly with local resolution

Local resolution of the experimental density from ResMap for error analysis and simulation parameterization

Root Mean Square Fluctuation (RMSF) correlates highly with local resolution

RMSF During Equilibration

Cascade and **direct** fitting structure to one half map and calculating the cross correlation to the other

<table>
<thead>
<tr>
<th>Fit to</th>
<th>CC w.r.t. Halfmap I</th>
<th>CC w.r.t. Halfmap II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Halfmap I</td>
<td>0.715 (0.686)</td>
<td>0.714 (0.685)</td>
</tr>
<tr>
<td>Halfmap II</td>
<td>0.716 (0.688)</td>
<td>0.716 (0.688)</td>
</tr>
</tbody>
</table>

CC of reference structure w.r.t. each half map was 0.719

Analyzing MDFF Model Quality 5: MD post-processing
Stability of structure during equilibration

Deviation from fitted structure after equilibration is within map resolution (~3Å)

Cascade MDFF (starting from structure 7Å RMSD) structure matches with reference structure (2.32 Å)

Equilibration of cascade MDFF structure

Equilibration

B-galactosidase (3.2 Å)

Ribosome-bound structure predicted by MDFF from cryo-EM map ~ 7.5 Å

Crystal Structure (3WVF) 3.2 Å

Beckmann, Schulten et al.
eLife; 3:e03035 (2014)

Nascent chain confirmed also by chemical cross-linking, gel filtration chromatography and mass spectroscopy.
MDFF Has a Wide Range of Applications

Over 60 reported MDFF applications:

- **By intramural Researchers:**

- **By extramural Researchers:**

**MDFF/xMDFF Methodological Articles:**
Acknowledgements and Further Information

Find out more about MDFF including:

- software downloads
- publications
- documentation
- tutorials

http://www.ks.uiuc.edu/Research/mdff/

Abhi Singharoy

Ivan Teo