MDfit: flexible fitting of homology models and X-ray structures to cryo-EM maps on a desktop

Tutorial

MDFIT.LANL.GOV
smog@ctbp.ucsd.edu

Ratje et al., **Nature** 2010.

Sanbonmatsu Team

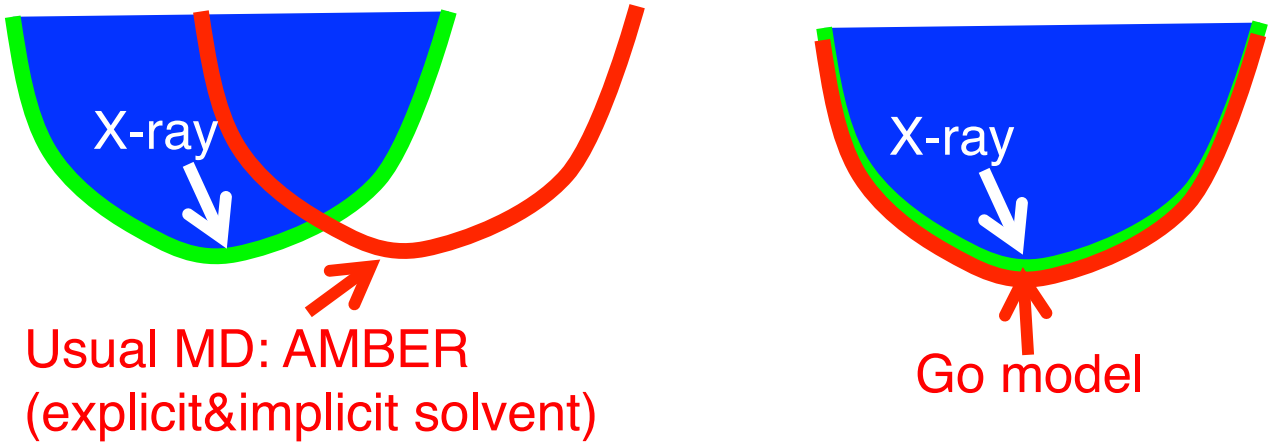
What is MDfit?

- Runs on desktop.
- Source code available.
- Input = pdb file (homology model or X-ray structure).
- Output = pdb fit to cryo-EM reconstruction that protects the stereochemistry and tertiary contacts in pdb file (any specific tertiary contact or dihedral or list of these can also be released if desired).
- Output pdb can also be used for stochastic MD simulations.
- Uses gromacs molecular dynamics simulation code (mdrun).
- Uses our structure-based (all-atom Go) forcefield (ucsd.smog.edu).

Ratje et al., **Nature** 2010.
Whitford et al., **PNAS** 2011.

Background: Structure-based MD (all-atom Go)

- Reduced force field
- Not the same as 'implicit solvent'
- System specific force field, defined by the pdb file
  (either homolog model or X-ray structure)
- Used to construct 'folding funnel'
  framework (Onuchic group, 1992)



Whitford, et al., **PNAS** 2011; Ratje et al. **Nature** 2010; Whitford et al. **RNA** 2010; Whitford et al. **Biophys. J.** 2009; Whitford et al. **Proteins** 2009.

Background: fitting potential

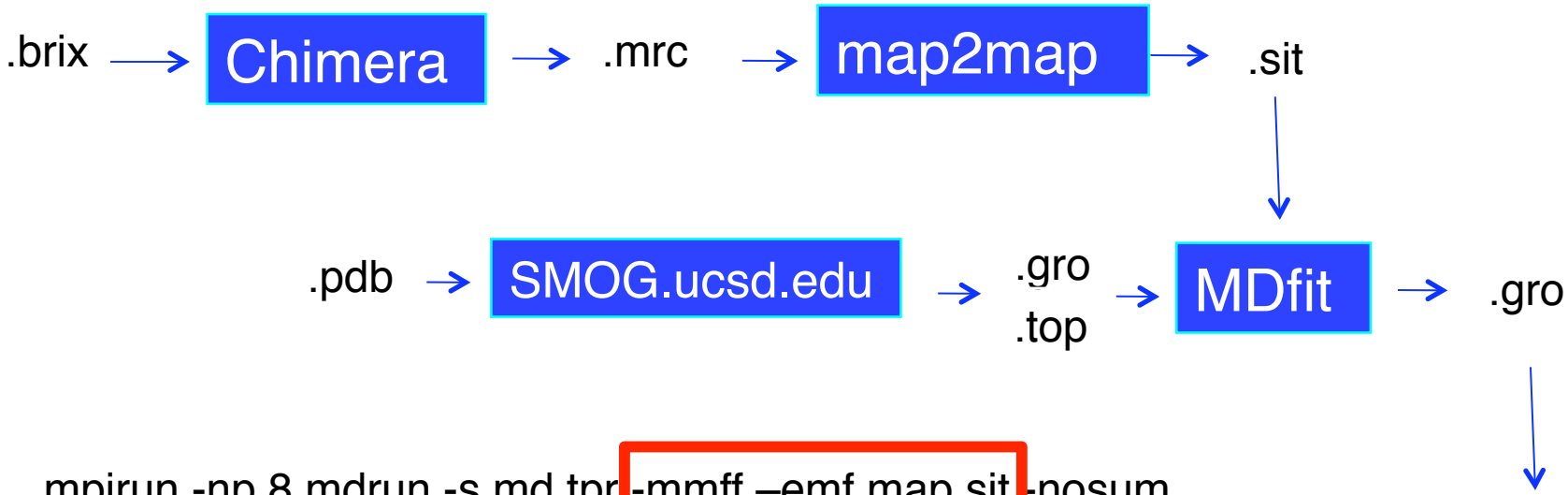
- Flexible Fitting Potential of Orzechowski & Tama, *Biophys J* (2008)
- Adds Go model term and energetic term based on correlation between cryoEM map and simulated map
- No constraints!

$$V = V_{Go} + V_{map}$$

$$V_{map} = -C_{map} \sum_{i,j,k} \rho_{sim} \rho_{exp}$$

Tama term

File pipeline



mpirun -np 8 mdrun -s md.tpr -mmff –emf map.sit -nosum -noddcheck 2> md.out

Currently takes 1 day per run for 70S ribosome on 8-core MAC desktop (5/29/2012).

We do 3 runs for the entire fit: 4Å filtered, 2Å filtered, then original un-filtered map.

Workflow

1. Unpack and install mdfit (gmx_em-4.0.5-beta.tar.gz)
2. Prepare pdb
3. Prepare map
4. Generate topology files (smog)
5. Minimize
6. Translate coordinates and edit box size in .gro
7. Generate run file .tpr (grompp)
8. Run MDfit
9. Troubleshoot
10. Caviots

Unpack and install MDfit

http://smog.ucsd.edu/extension/compile_notes_MDfit

Notes on compiling MDfit in Gromacs Version 4.0.5: compiled with fftw v 3.2.2, openmpi 1.2.8, gcc, on Max OSX 10.6 with 64 bit Intel support
These commands assume that mpi is in your path.

```
tar xvfz gmx_em-4.0.5-beta.tar.gz

FFTW_SRC=/Users/kys2/src/fftw
GMX_SRC=/Users/kys2/src/gmx

FFTW_INSTALL=/Users/kys2/bin/fftw_tmp
GMX_INSTALL=/Users/kys2/bin/gmx-cryoem
export CFLAGS=-m64


cd $FFTW_SRC
./configure --enable-float --prefix=$FFTW_INSTALL
make
make install

cd $GMX_SRC
./configure --enable-mpi --prefix=$GMX_INSTALL  --enable-apple-64bit
make
make install
```

MDFIT.LANL.GOV

Prepare PDB

1. Put “END” line at end of pdb file.

2. Put “TER” lines between each chain and between fragments (if you have gaps in sequence, you must include a “TER” line).

3. Make sure you don’t have two different residues with the same chain and same residue numbers.

4. Make sure you don’t have messed up residues with wrong atoms in the residue or missing chemical groups.

5. For RNA, change 2*O to 2’O, etc.

Prepare map

1. Filter map in chimera:

Tools->Volume Data -> Volume filter

4.0

File->Save map as 70S_4A.mrc

Tools->Volume Data -> Volume filter

2.0

File->Save map as 70S_2A.mrc

2. Convert map to situs format (map2map):

./map2map 70S_4A.mrc 70S_4A.sit

map2map> Enter selection: 1

./map2map 70S_2A.mrc 70S_2A.sit

map2map> Enter selection: 1

./map2map 70S_0A.mrc 70S_0A.sit

map2map> Enter selection: 1

3. Edit map header (must have 0 origin because gromacs cannot handle negative coordinates)

vi 70S_4A.sit

Change:

1.260000 -226.800003 -226.800003 -225.539993 360 360 360

to:

1.260000 000.000000 000.000000 000.000000 360 360 360

Repeat for 70S_2A.sit and 70S_0A.sit.

Generate topology files (smog)

1. Goto smog.ucsd.edu
2. Click “Prepare a simulation” on the right side bar/frame.
3. Upload pdb file by clicking “Browse” (step 1 on web page).
4. Change “shadow map” to “Cut-off” (step 2 on web page).
5. Name your system (step 9 on web page).
6. Generate topology (click “Submit”).
7. Troubleshoot pdb file by reading error messages and looking for line numbers in the error messages.
8. If you are really stuck, email paul.whitford@gmail.com
9. If smog is successful, download the files and unpack
10. Copy .gro and .top files into your working directory

```
cp l.6670.pdb.sb/l.6670.pdb.gro 70S.gro
cp l.6670.pdb.sb/l.6670.pdb.top 70S.gro
```

Minimize

1. Check input file

   vi min.mdp

2. Generate run file (.tpr)

mpirun -np 8 ~kys2/bin/gmx-cryoem/bin/grompp -f min.mdp –c 70S.gro –p 70S.top –o 70S.min.tpr

3. Run minimization

mpirun -np 8 ~kys2/bin/gmx-cryoem/bin/mdrun –s 70S.min.tpr

cp confout.gro 70S.min.gro

cp md.log 70S.min.log

Translate coordinates and edit box size (.gro)

We need to translate our coordinates twice. Once because smog translates the pdb. Then,

A second time because gromacs cannot handle negative coordinates. This requires us to translate the origin of the map , and , in turn, the coordinates.

1. Use the coordinates of the first atom in the pdb file.

1. Convert these coordinates from Å to nm
2. Look at the coordinates of the first atom in the .gro file (70S.gro)
3. Compute the difference between these coordinates

   (take pdb – gro)

   ans. = -16.4174 -15.4911 -15.5213

5. Translate:

   ~kys2/bin/gmx-cryoem/bin/trjconv –f 70S.min.gro –s 70S.min.gro -trans -16.4174 -15.4911 -15.5213 –o 70S.min.corrected.gro

6. Translate again because we have translated the map (above):

~kys2/bin/gmx-cryoem/bin/trjconv –f 70S.min.corrected.gro –s 70S.min.corrected.gro -trans +22.68 +22.68 +22.554 –o 70S.min.corrected2.gro

7. Edit box size

vi 70S.min.corrected2.gro

Change last line to:

47.27400 47.95940 47.76280

Run MDfit

1. Check input file (VERY IMPORTANT!!)

   emweight should be approximately the number of atoms in pdb (and in .gro):

   emweight = 127940

   emsigma should be 0.5 to start, but for the _2A and _0A maps, you can experiment with lower values.

   check nsteps

2. Do not ignore step 1!!
3. Generate run file (.tpr). This can take 3-4 hours for big systems.

   ~kys2/bin/gmx-cryoem/bin/grompp -f md_4A.mdp –c min.gro –p 70S.top -o md_4A.tpr&

4. Check coordinates:

   Load .sit and .gro files into vmd to make sure they superpose and are aligned.

5. Run MDfit

   mpirun -np 8 ~kys2/bin/gmx-cryoem/bin/mdrun -s md_4A.tpr -mmff –emf 70S_4A.sit -nosum -v -noddcheck 2> md_4A.out &

6. Monitor md_4A.out to make sure correlation begins with > ~0.5 and increases over ~100,000 steps
7. Repeat entire process for 70S_2A.sit and 70S_0A.sit
8. Translate back:

   ~kys2/bin/gmx-cryoem/bin/trjconv -f md_4A.gro -s md_4A.gro -trans -22.68 -22.68 -22.554 -o md_4A.uncorrected.gro

6. Convert to pdb (e.g., vmd). Clean your pdb (add TER, chain names)
7. Check stereochemistry.

Trouble shooting

1. Code won't compile
2. smog doesn't like pdb
3. Map and pdb are not aligned
4. MDfit
    1. Nothing happens to molecule, not being fit
    2. Molecule is crushed / stereochemistry ruined
    3. Correlation is quite low
    4. Molecule is broken in pieces, with some pieces wrapping around the period box

Caviots

1. MDfit results only as good as initial model.

Note: we have observed good results when map is partially empty.

Conclusions

- **MDfit code available**
- **Runs on desktop**
- **Preserves stereochemistry**

• Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid sites. Ratje AH, Loerke J, Mikolajka A, Brünner M, Hildebrand PW, Starosta AL, Dönhöfer A, Connell SR, Fucini P, Mielke T, Whitford PC, Onuchic JN, Yu Y, Sanbonmatsu KY, Hartmann RK, Penczek PA, Wilson DN, Spahn CM. Nature. 2010 Dec 2;468(7324):713-6.
• Excited states of ribosome translocation revealed through integrative molecular modeling. Whitford PC, Ahmed A, Yu Y, Hennelly SP, Tama F, Spahn CM, Onuchic JN, Sanbonmatsu KY. Proc Natl Acad Sci U S A. 2011 Nov 22;108(47):18943-8.
Tutorial by Karissa Sanbonmatsu

MDFIT.LANL.GOV