

Homology Modeling and Threading Tutorial

Survey Course in Bio Crystallography and Bioinformatics
Lima, Peru, March 17-29, 2009

Homology modeling and threading tutorial

Very often when you are studying a protein, the structure is unknown. Homology modeling is the cheapest alternative to solving a crystal structure of your protein.

Outline of document:

1. Question
2. Properties of proteins needed for building a model
3. Auto modeling server to thread and simple minimize
4. Check model
5. Improve model?
6. Compare *M. tub* to *M. Ulc*

1. Question

- a. How does the structure of *M. ulcerans* FABG compare to the FABG from *M. tuberculosis*. Let us model and compare the similarities and differences between these two proteins.

2. Properties of proteins

Without spending a lot of time reinventing the wheel, understanding how a protein is built is the number one way to make a good homology model and validating it. Proteins are made up of amino acids that are covalently attached to one another through a path of alternating nitrogen and carbon atoms. This **backbone** structure is described though the angles that are produced as one runs along the atoms that make up **Phi** (CNCaC) and **Psi** (NCaCN) angles. The angles will take up in specific conformations that can be plotted and used to see if something is wrong with the structure. The nitrogen has a slight positive charge to it and the C has a slight negative charge to it. These charged properties of the backbone is important when **secondary structure** is built up. Specific hydrogen bonding patterns develop when forming an **alpha helix** or a **beta sheet** (see tutorial on visualization of a structure.).

- a. For more information see:

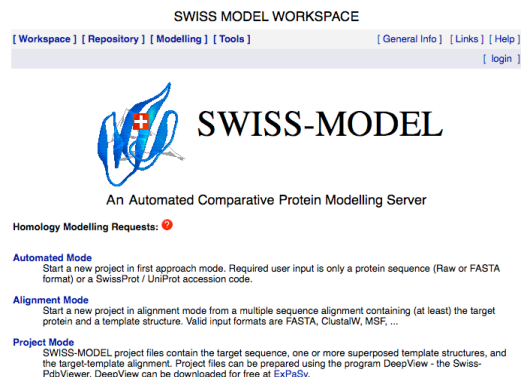
<http://ekhidna.biocenter.helsinki.fi/downloads/teaching/spring2006/proteinianalyysi/Proteinianalyysi-06-VI-AppendixA.htm>

- b. Google; "protein modeling".

3. Auto Modeling servers

There are several reasons why we are going to build a homology using automatic methods. 1) Time 2) ease of use and 3) convenience.

- a. Go to http://swissmodel.expasy.org/workspace/index.php?func=modelling_overview



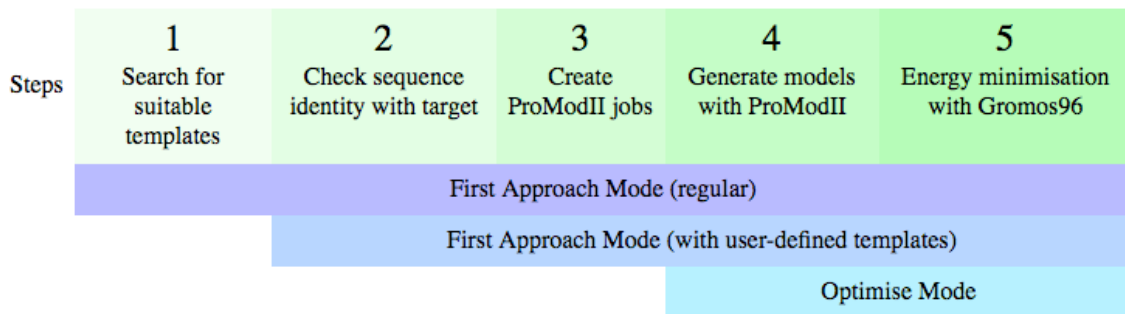
- b. There are 3 ways to build a model with this server. We are going to use the Automated mode. We are going to do it this way because the sequence identity between the

Homology Modeling and Threading Tutorial

tuberculosis FABG and ulcerans FABG is 90%. There are only a few residues that are different between the two sequencing and so there “should be” very minimal structural differences.

- i. Alignment mode is good if your sequence shows ~50 identity to a known structure and manual intervention is needed on the alignment
- ii. The Project mode is the manual mode. Human intervention at each step.
- c. What does the automatic mode do?
 - i. Search for suitable templates
 - ii. Check sequence identity with target
 - iii. Generate models with PROMODII
 - iv. Energy minimize.

Graphic below shows each of the steps for modeling building and where each of the “modes” start.



*<http://swissmodel.expasy.org/SWISS-MODEL.html>

- d. Click on automated mode. You should see the web form below.

SwissModel Automatic Modelling Mode ?

Email:
Project Title:

Provide a protein sequence or a UniProt AC Code: ?

Options: ?

Use a specific template: ?

- e. Fill in an email address. This will be used to access your files later.
- f. The project title is “Modeling M. ulcerans.”
- g. Either copy the amino acid sequence from M. ulcerans into the large text box or put the swiss prot id # into the text box. “A0PNU4”.
- h. Leave the “use a specific template” blank. We will let the program get the best model.
- i. Click “Submit Modeling Request”.
- j. Depending on the strain on the server the request might take a bit.

Homology Modeling and Threading Tutorial

- k. Download the model as a pdb. From here on we need to test the accuracy of the model. All we have is the assumption the auto server correctly built our model.

Model Details: Segment 1

Model info:
modelled residue range: 17 to 255
based on template **1uznA** (1.91 Å)
Sequence Identity [%]: 92.469
Evaluate: 5.02e-110

display model: as pdb - as DeepView project
download model: as pdb - as Deepview project - as text

- l.
m. Note the PDB that was used to model this protein. (1UNZ). We need to get this pdb so that we can compare it with our threaded model.

4. Check model

- a. Go to : <http://molprobity.biochem.duke.edu/>

Molprobity tests a number of characteristics of your structure (ramachandran angles, side chain orientation and side clash) against a database of the best structures in the PDB. It is a good metric to tell you the validity of your model.

PROBITY

Main page
Evaluate X-ray
Evaluate NMR
Fix up structure
Work with kins
View & download files
Lab notebook
Feedback & bugs
Site map

Main page

FILE UPLOAD/RETRIEVAL (MORE OPTIONS)

PDB/NDB code: type: PDB coords Fetch >
Browse... type: PDB coords Upload >

- b. Click browse. Find your model.pdb file and click upload.
c. Then click continue after your file has been uploaded.
d. Our model currently contains only the atoms Carbon, nitrogen, oxygen and sulfur. Organic molecules contain all of these atom plus hydrogen. We need to add the hydrogen atoms on the model.

PROBITY

Main page
Evaluate X-ray
Evaluate NMR
Fix up structure
Work with kins
View & download files
Lab notebook
Feedback & bugs
Site map

Main page

SUGGESTED TOOLS (ALL TOOLS)

Currently working on: Model_1.pdb

Add hydrogens
Make simple kinemages
Analyze geometry without all-atom contacts
Edit PDB file
Downgrade file to PDBv2.3 format (for download only)

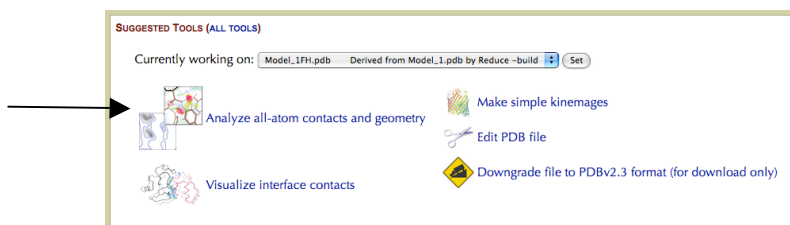
- e. Click “add hydrogen’s”.
f. The next page will ask you the method of adding the hydrogen’s. We will run in default mode. Click “Start adding H>”
g. Any problems with modeling the amino acids Glutamine, Asparagine or histidine residues in your structure will be flagged here.

Flip?	Chain	Res#	Res ID	Orig	Flip	Flip-Orig	Code	Explanation
<input checked="" type="checkbox"/>		41	GLN	-1.3	-0.028	1.272	FLIP	Some evidence recommending flip.
<input checked="" type="checkbox"/>		73	ASN	-5.7	-4.1	1.6	CLS-FL	Both orientations clash but flip was preferred.
<input checked="" type="checkbox"/>		96	ASN	-0.58	-0.0048	0.5752	FLIP	Some evidence recommending flip.
<input checked="" type="checkbox"/>		157	ASN	-3.1	-0.3	2.8	FLIP	Clear evidence for flip.
<input checked="" type="checkbox"/>		158	GLN	-2.5	2.1	4.6	FLIP	Clear evidence for flip.
<input checked="" type="checkbox"/>		160	ASN	-8.8	-1.7	7.1	CLS-FL	Both orientations clash but flip was preferred.

Regenerate H, applying only selected flips >

- h. Click “Regenerate H” which will fix the bad residues, and then click continue. It will ask you if you would like to download the new file. Click “no”.

Homology Modeling and Threading Tutorial



- i. A new web page will show up. This is where we will analyze the overall quality of the structure. Click on the link “Analyze all –atom contacts and geometry”.
- j. On the next page leave all the default checked items and click “run programs to perform these analysis”.
- k. A summary statistics window will show up next and give you the results of the analysis

Summary statistics

All-Atom Contacts	Clashscore, all atoms:	14.47	51 st percentile* (N=1784, 0Å - 9999Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å)		per 1000 atoms.
Protein Geometry	Rotamer outliers	0.56%	Goal: <1%
	Ramachandran outliers	0.00%	Goal: <0.2%
	Ramachandran favored	97.05%	Goal: >98%
	C β deviations >0.25Å	0	Goal: 0
	MolProbity score	1.83	84 th percentile* (N=27675, 0Å - 99Å)
	Residues with bad bonds:	0.00%	Goal: <1%
	Residues with bad angles:	0.00%	Goal: <0.5%

* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst.

- l. You can see from the above plot there are some errors in our model. The clash score (a score that describes the radii of atom overlap) is low. As well the Rama angles on some residues could be improved.
- m. Your model may be better or worse then this. Either way ideally these should be all green.
- n. You need to save this model so we can upload it to another server.

5. Improve model?

- a. The best way to fix this model is to check each residue and manually edit it. Molprobity has an web editor built in called KiNG that can do that. If there is time, we will use it. For an faster and easier way of improving the clash score we will use a server called WHAT IF.
- b. Go To the WHAT IF Web Interface: <http://swift.cmbi.ru.nl/servers/html/index.html>



- c. Click the link on the left called “ Build/check/repair model”
- d. Then click the link to
 - i. “Remove bumps from a structure”
Atomic clashes (bumps) will be removed by rotating side chain torsion angles.
- f. Upload the saved file you downloaded from molprobity into this webform. And Click “send”.
- g. This will run for a bit and return a click to download a “fixed” file. Download this.
- h. Load this “fixed” file into a new session of Molprobity and run the “add hydrogens” and run “ analysis contacts” again.
- i. You should see a dramatic improvement in the clash score.

Homology Modeling and Threading Tutorial

Summary statistics

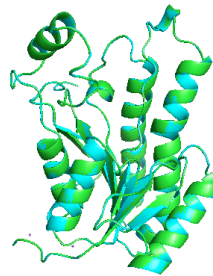
All-Atom Contacts	Clashscore, all atoms:	10.22	70 th percentile* (N=179, 0Å - 9999Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Protein Geometry	Rotamer outliers	0.56%	Goal: <1%
	Ramachandran outliers	0.00%	Goal: <0.2%
	Ramachandran favored	97.05%	Goal: >98%
	Cβ deviations >0.25Å	0	Goal: 0
	MolProbity score	1.70	89 th percentile* (N=27675, 0Å - 99Å)
	Residues with bad bonds:	0.00%	Goal: <1%
	Residues with bad angles:	0.00%	Goal: <0.5%

* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst.

- j.
k. Our “fixed” model has atoms that have been better placed according to molprobity.

6. Compare M. tuberculosis FABG to M. ulcerans FABG.

- The pdb ID for the FABG from M. tuberculosis is 1UZN.
- Using the techniques established in molecular visualization tutorial we can compare, in pymol, these 2 structures.
- Load our model in pymol “fixed”
- Load the model 1UZN.
 - They should both be centered on top of each other. If not you need to ask for help.
- Change the view to cartoon for both.



- Immediately, in terms of secondary structure, what do you notice?
- Why do you think there has been no change in the secondary structure?
- Here is the alignment between 1UZN FABG and M. ulc-FABG model

```

M_ulcer      MTDTATEQNTESAADGGKPAFVSRSLVTGGNRGIGLAIAQRLATDGHVAVTHRGSGAP
1UZN_A      -----AKPPFVSRSLVTGGNRGIGLAIAQRLAADGHKVAVTHRGSGAP
               . ** .*****;***;*****

M_ulcer      EGLFGVECDVTDNDVDAVDRAFKEVEEHQGPVEVLVSNAGLSADAFLIRMTEERFEKVIDAN
1UZN_A      KGLFGVEVDVTDSDAVDRAFTAVEEHQGPVEVLVSNAGLSADAFLMRMTEEFKVINAN
               :***** :***.*****.*****;*****;***;**

M_ulcer      LTGAFRVAQRASRSMQRKKFGRLIFIGSVSGSWGIGNQANYAASKAGVIGMARSIARELS
1UZN_A      LTGAFRVAQRASRSMQRNKFGRMIFIGSVSGLWIGNQANYAASKAGVIGMARSIARELS
               *****;***;***** *****

M_ulcer      KVNVTANVVAPPGYIDTDMTRALDERIQEGALQFIPAKRVGTAAEVAGVVSFLASEDASYI
1UZN_A      KVNVTANVVAPPGYIDTDMTRALDERIQQALQFIPAKRVGTAAEVAGVVSFLASEDASYI
               *.*****;*****.*****

M_ulcer      SGAVIPVDGGMGMGH
1UZN_A      SGAVIPVDGGMGMGH
               *****
    
```

- I have bolded 3 regions in the sequence alignment that are important for substrate binding to this protein. Overall 2 regions do not show any differences in sequence. For

Homology Modeling and Threading Tutorial

these two regions are any of the side chains in different conformations between the 2 structures? Observe closely the atoms of Gln, Asp and His residues.

- j. The middle region that do show a one-residue difference in sequence alignment.
Looking at the structure, would this difference probably cause a problem to substrate binding? Why or Why not?