



Bioinformatics Tutorial 2013



Use free public tools to predict protein structure via comparative modeling

For use in the Scripps Graduate Structural Biology Course
taught by [Adam Godzik](#), Peter Lee and Jeong Hyun Lee

This website was initially developed by Mallika Veeramalai, Graham Johnson and [Adam Godzik](#) in 2009

I. Create an account on the homology modeling server

- A. Visit <http://swissmodel.expasy.org/> and use the website's navigation tools to complete the following:
 1. To register, please click on "myWorkspace"
 2. Collect the assigned password from your email
 3. Use this password to login to Swiss-Model
 4. Download the Swiss-PdbViewer
- B. Before, starting your exercises click the link below to review the steps in comparative modeling
 1. <http://bioinformatics.burnham.org/SSBC/review.htm>
 2. Skim this webpage to comprehend the number of steps involved in Modeling and Analysis
- C. OK, you're ready to build a model

II. Your first homology model– an "easy target"

In this exercise, we will model the structure of a protein's 1D sequence using one of its homologs as a template. We will then compare the generated homology model to the actual structure of the protein to gain insight.

- A. We will begin slowly by stepping through an "easy target" protein. It should be easy to find usable templates
 1. Click link below to view the Ligand Binding Domain of the Vitamin D Nuclear Receptor
 - a. <http://bioinformatics.burnham.org/SSBC/easytargets/1.fasta>
 2. Copy the FASTA sequence
 - a. Copy all of the text on the linked page
 - b. In this case, the FASTA sequence contains single letter residue codes for each amino acid in the protein
 1. The first line is a comment >containing header information about the file
 2. You can visit http://en.wikipedia.org/wiki/FASTA_format for more information

III. Comparative modeling– General methods

- A. Generate your first attempt at a homology model fast and easy using the **fully automated Swiss-Modeler** as a first approach
1. In your workspace (logged in to <http://swissmodel.expasy.org/>) click *Modeling>Automated Mode*
 2. Enter your email address
 3. Enter a sensible title in the "Project Title:" box
 4. Paste your FASTA sequence into the large box that says "Provide a protein sequ..."
 5. Click on the "Submit Modeling Request" button
 - a. It may take several minutes to return a result as you're relying on the server to do a lot of work for you.
 - b. A test run on 9/17/09 took 7 minutes.
 - c. In the mean time, open a new browser tab and start working on the next step "Using ProtMod Server"
 - d. Save your result in a browser window for use in the **Model Analysis** and **Advanced Model Analysis** sections of this tutorial. You may also access the result at any time from the link in an email that Swiss model will send when finished.
- B. Generate a second version of this homology model using the **ProtMod Server**
1. Visit <http://ffas.burnham.org/protmod-cgi/login.pl?qryType=sequence>
 2. Login or register if a first time user
 3. In left panel under *Target By*, click *Sequence*
 4. Enter a sensible title in the *Job Name:* box
 5. Paste your FASTA sequence into the *Target Sequence:* box
 6. Select *FFAS* option for Database Searching Methods
 7. Select *FFAS* option for Sequence Alignment Methods
 8. Click on the *Submit* button
 9. Once you have received the results in your email, then go to the results page and down-load the top model
 - a. Save your result in a browser window for use in the **Model Analysis** and **Advanced Model Analysis** sections of this tutorial. You may also access the result at any time from the link in an email that ProtMod will send when finished.
- C. Generate a third version of this homology model using a **TemplateIdentification Tool**. As part of the tutorial, try both options
1. Begin by searching for appropriate templates
 - a. **Option 1 (possibly slow): Use the Template Identification Tool in Swiss-Model**
 1. While logged into your Swiss-Model workspace click *Tools>Template Identification*
 - a. Here, you can step through sequentially and modify many of the processes that the Fully Automated algorithm ran.
 2. Enter your email address
 3. Enter a sensible title in the *Project Title:* box
 4. Paste your FASTA sequence into the large box that says *Provide a protein sequ...*
 5. If searching for a *difficult target* protein, check to turn on the *Iterative Profile Blast:* option, otherwise leave it unchecked.
 6. Click on the *Submit Modeling Request* button
 - a. A test run for VitaminD with IterativeProfileBlast turned on took 3 minutes to get to an interproScan/GappedBlast page and 30 minutes to finish with a completed list of results.
 - b. The same FASTA sequence with IterativeProfileBlast turned off took 26 minutes to finish.
 - c. Please move on to other sections while you wait for your results to return.

- d. Once you receive your template search results hold them in a browser window until we learn how to select a good template
- b. **Option 2 (speedy results): Use the NCBI's Protein BLAST server for Template Identification**
1. Visit <http://blast.ncbi.nlm.nih.gov>
 2. Enter a sensible title in the *Job Title:* box
 3. Paste your FASTA sequence into the large box that says *Enter accession numb...*
 4. Under *Program Selection*>*Algorithm* select the *PSI-BLAST* (Position_Specific_Iterated_BLAST) option
 5. Click on the *BLAST* button
 - a. A test run for VitaminD with IterativeProfileBlast turned on took only **2 minutes** to return a completed list of results.
 - b. Once you receive your template search results hold them in a browser window until we learn how to select a good template in the next section

2. Select a template

- a. Examine your Template Identification results from Option 1 & 2
 1. There are several potential template structures
 2. How should one choose a template?
 - a. Consider several factors
 1. Which structure has the highest percentage identity with your sequence?
 2. Which structure has the lowest percent identity with your sequence?
 3. How long is the alignment between your protein and the first potential template shown?
 4. How long is the alignment between your protein and the last potential template shown?
 5. Which alignment has the highest E-value?
 6. Which alignment has the lowest E-value?
 7. Which E-value is the best?
 8. If your testing the accuracy of comparative modeling, what do you notice about the highest identity template?
 - b. Fill in this form with the 5 digit PDB+1 ID (e.g., 1abcA) for your selected template as you model each "difficult target" from the list in the *How Accurate Is Comparative Modeling* section
 - 1.

Your Protein Sequence	Your Template PDB
MG3368B_protein1	
GE1527P_protein2	
RK10653A_protein3	
MH7542A_protein4	
PX15502A_protein5	

2. Show the list to an instructor once you've completed two or three.
3. Align your target to your template (**Advanced level approach**)
 - a. Obtain the FASTA format sequence for your **template** PDB.
 1. Visit [the Protein Data Bank](http://www.rcsb.org/) (PDB database <http://www.rcsb.org/>)
 2. Search box> type your 4 letter-PDB code, then click search
 3. You'll see a page with full structural information and different menus on the top.
 4. Select the *Sequence* menu
 1. Under *Chain Display*, click *fasta*
 2. Save the fasta to your computer

3. You have collected your template sequence and you have your target sequence from the tutorial page.
- b. Generate an alignment
 1. Visit ClustalW at <http://www.ebi.ac.uk/Tools/msa/clustalo/>
 1. Paste your FASTA formatted "Target Sequence" from the tutorial page and the FASTA formatted "Template sequence" you downloaded from the [PDB](#) into the sequences box. **Please note: use only one-word without any spaces in the description line of the FASTA sequences.** See [example](#) for ClustalW input sequences. The output ([example](#)) from clustalW for sequences in this format can be used directly in SWISS Modeling Server. Otherwise you need to change the description line in the clustalW output in order to use it as input for SWISS Modeling server.
 2. Fill in your email (optional)
 3. Under OUTPUT FORMAT > Choose *aln w/numbers*
 4. Under OUTPUT ORDER > Choose *input*
 2. Click the *Submit* button
 3. The ClustalW results will display in a new page
 1. SKIP this first step in class, but COMPLETE it at home well before the exam.
 - a. In the first section you will see list of result files and a button for "Start JalView"
 - b. If you want to examine or edit your alignment then click the button.
 - c. You'll see an editor with your target and template sequence alignment
 1. The color code format shows amino acid properties
 2. Conservation, Quality and Consensus details w.r.t to your alignment column are at the bottom
 3. Study and analyses the alignment closely
 4. Read details about how to use JalView from the JalView help page <http://www.jalview.org>.
 2. In the result page check the "Alignment file" with a link to your alignment file
 3. Save that file into your directory and use it in the next section to model your target to your template
4. Model your target to your template (Simple approach)
 - a. Model to a specific template you've hand picked using the **semi automated** in [Swiss Modeler](#) server.
 1. In your workspace (logged in to <http://swissmodel.expasy.org/>) click *Modeling>Automated Mode*
 2. Enter your email address
 3. Enter a sensible title in the *Project Title:* box
 4. Paste your FASTA sequence into the large box that says *Provide a protein sequ...*
 5. In the option **Use a specific template:** type your selected template pdb code (a pdb code is 4 characters long and a ChainID is usually a single character in the homology modeling world)
 6. Click on the *Submit Modeling Request* button
 - a. Save your result in a browser window for use in the **Model Analysis** and **Advanced Model Analysis** sections of this tutorial. You may also access the result at any time from the link in an email that Swiss model will send when finished.
5. Model your target to your template using alignment (**Advanced approach**)
 - a. Model to a specific template you've hand picked using the **Alignment Mode** in [Swiss Modeler](#) server.

1. In your workspace (logged in to <http://swissmodel.expasy.org/>) click *Modeling* > *Alignment Mode*
 - a. Enter your email address
 - b. Enter a sensible title in the *Project Title*: box
 - c. In the Alignment Input Format: > Choose "CLUSTALW"
 - d. Upload your clustalW alignment file ".aln" from your directory using the *Browse* button .
 - e. Click *submit alignment*
 - f. A page will show details about "Target Sequence" and "Template Sequence"
 1. Enter your template name
 2. PDB-code: (4 letter PDB code)
 3. Chain-ID: single-letter-code (remove any other characters except your Chain-ID "A" or "B")
 - g. Carefully check you have submitted the correct Target and Template sequences and your template PDBCode and the Chain-ID also correctly selected
 - h. Click "submit alignment" button
 - i. On the page that opens with your ClustalW alignment, ensure that the alignment has been interpreted correctly
 - j. Click "submit alignment" button
 - k. Save your result in a browser window for use in the **Model Analysis** and **Advanced Model Analysis** sections of this tutorial. You may also access the result at any time from the link in an email that Swiss model will send when finished.

IV. Basic Model Analysis

- A. Obtain a copy of the Template
 1. If you haven't already done so, download your template from www.rcsb.org using the first four digits of the PDB ID code in a search
- B. Quickly view and compare your model with a familiar tool
 1. Open PyMOL
 - a. File>Open>YourHomologyModel.pdb
 - b. PyMOL> load YourTemplateModel.pdb
 - c. PyMOL> align YourHomologyModel.pdb, YourTemplateModel.pdb
 1. Record the RMSD that prints and contemplate the other information output to the console
 - d. Color YourTemplateModel.pdb light blue
 - e. Color YourTargetModel.pdb red
 - f. Turn both models into cartoon representations A>Preset>Publication
 - g. Show sidechains as lines S>Sidechain>Show:Lines
 2. Write out a brief subjective observational analysis
 - a. Do your backbones appear to line up well?
 - b. Do the sidechains line up perfectly?
 - c. Are some patches more variable than others?
 - d. How do secondary structure elements relate to observable local RMSDs?
- C. Use Swiss-Pdb Viewer to perform more complicated analyses
 1. While Logged into your Swiss-Model Workspace
 - a. Visit the results page of the model you'd like to analyze
 - b. Click on the text *download model: as pdb - as Deepview project - as text*
 2. Start Swiss PDB Viewer
 - a. Open the DeepView.pdb file you just downloaded
 - b. Now, you should have two windows open: one with a toolbar and several pull-down menus and another that is a protein structure viewer.
 3. Moving Your Model in the Viewer
 - a. Ensure that your Toolbar is open and visible (Wind>Toolbar)

- b. There are four buttons in the toolbar that are important for viewing your model
 1. Click the left most button to center and zoom in on the model to fill the screen.
 2. Select the hand tool and drag around in your view port. This is the translate tool.
 3. Drag with the third (zoom) tool to zoom in and out.
 4. Drag with the fourth (rotate) tool to spin the model.
4. Study your homology model with Swiss PDB Viewer
 - a. Menu>Wind>LayersInfo
 1. Each model or structure is displayed in it own layer.
 2. In the "Layers Infos" window, there should be two layers: one called "TARGET" and one named for the template structure you used to build your model.
 - b. Menu>Color>Layer
 1. Your Template should turn blue
 2. The wireframe display of your model is yellow
 3. A ribbon diagram of your model is green and perhaps some other colors as well.
 - c. Layers Info Window.
 1. Next to the name of each layer, there is a column labeled "vis".
 1. Click on the "v" next to your template in this "vis" column. Your blue template should disappear.
 2. The change may be subtle if your template does not have a cartoon showing- just a wireframe backbone will disappear.
 2. Click on "TARGET". Now the word "TARGET" should be red.
 - d. Menu>Wind>ControlPanel
 1. A new window should open.
 1. In the left-most column of this new window, all of the residues in your model are listed.
 2. To the right of this column is another column that is labeled "show".
 1. Hold "shift" while clicking on ones of the "v"'s in this second column from the left.
 2. Now only the ribbon diagram of your model should remain in the structure-viewing window.
5. Use **geometry-based tools** to gauge the quality of your model.
 - a. Menu>Select>Residues Making Clashes
 1. Press "enter" or return on a mac.
 2. Watch for a subtle change, yellow residue lines should appear.
 3. This will display amino acids that are in clashes with other amino acids.
 4. If no amino acids are displayed, there are no amino acids in clashes with other amino acids in your model (or you missed a step)
 - a. How many amino acids are clashing with others? None, some, or many?
 - b. Menu>Select>Residues Making Clashes with Backbone
 1. Press "enter".
 2. This will display amino acids that are in clashes with the backbone of your model.
 3. If no amino acids are displayed, there are no amino acids in clashes with the backbone in your model.
 - a. How many amino acids are clashing with the backbone? None, some, or many?
 - c. Menu>Select>Sidechains lacking Proper H-bonds".
 1. Press "enter".
 2. This will display sidechains that lack proper hydrogen bonding.
 - a. How many sidechains lack proper hydrogen bonding? None, some, or many

6. Use **empirical energy-based tools** to gauge model quality:

1. Menu>Win>Alignment

a. A new window will open that shows the sequence alignment of your target protein and your template.

1. Click on the word, "Target", so that it is red.
2. Click on the white arrow left of "Target" to turn up the graph view

a. This increases the height of the "Alignment" window.

b. Above the alignment, there is a line and a curve displayed.

3. Menu>SwissModel>Auto Color by Threading Energy

4. In the "Alignment" window, click on the word, "smooth".

a. In the popup window, set the "nb of aa to average" to 1.000 and click "ok"

b. In the "Alignment" window, click on the "E=" text, which is above "smooth"

1. This will recalculate the threading energy for your model.

2. The line, in the "Alignment" window, shows where the threading energy is zero.

3. The curve shows the threading energy per residue (actually the average energy of each residue plus the one residue towards the N-terminus and one residue towards the C-terminus) of your model.

a. A good model should have an average threading energy of 0.

b. A long stretch of residues with threading energy higher than zero and colored red or orange may indicate that this part of the model is not accurate.

D. Consider the function (activity, process) of your protein:

1. Does the template have the same or a similar function?
2. If so, how good is the alignment of the active site or residues involved in this function?
3. Are these residues conserved?

V. Advanced Model Analysis

A. Comparing the Model to the Actual Structure of the Protein if available

1. The proteins we are using in this exercise actually have known structures, so we can compare your model to the actual structures
2. Visit <http://fatcat.burnham.org/fatcat-cgi/cgi/fatcat.pl?-func=pairwise>
3. Enter your email address.
4. In the *Get the 1st stru...>Upload file (in PDB format):* box, upload your homology model and name it *modelProteinName*
5. In the *Get the 2nd stru...* fill in the box *Provide PDB code:* and *Chain:* appropriate to your homology model (1st structure) using these tables:

a. **Easy Target Models:**

Your Protein Model	PDB code	Chain
The Ligand Binding Domain Of The Vitamin D Nuclear Receptor	2hcd	A
Shikimate Kinase From Mycobacterium Tuberculosis	2dft	C
The Pleckstrin Homology Domain Of Akt1 In Cancer (Akt1-Ph_e17k)	1p6s	A

Avian Respiratory Complex Ii	1nek	B
Bacillus Circulans Strain 251 Cyclodextrin Glycosyltransferase	1cyg	Since this protein only has one chain, leave this field blank in FATCAT.

b. **Difficult Target Models:**

Your Protein Model	Download Actual Structure files	Chain
model_protein1	MG3368B_protein1.pdb	A
model_protein2	GE1527P_protein2.pdb	A
model_protein3	RK10653A_protein3.pdb	A
model_protein4	MH7542A_protein4.pdb	A
model_protein5	PX15502A_protein5.pdb	A

6. Click the button *Send Request*
7. Once you get your results, how similar is your model to the actual structure of your protein?
 1. What is the RMSD?
 2. Examine the superimposition of your model and the actual structure.
 3. Which parts of your model look the most similar to the actual structure?
 4. Which parts of your model are clearly different than the actual structure?
8. Discuss your results with your classmates.
 1. How similar were their targets and template in terms of percent sequence identity?
 2. What is the RMSD between their model and the actual structure?
 3. Which parts of their model was the most and least accurate?

VI. How accurate is comparative modeling?

- A. Repeat the exercise with one more easy target from this list of FASTA links.
 1. [The Ligand Binding Domain Of The Vitamin D Nuclear Receptor](#)
 2. [Shikimate Kinase From Mycobacterium Tuberculosis](#)
 3. [The Pleckstrin Homology Domain Of Akt1 In Cancer \(Akt1-Ph_e17k\)](#)
 4. [Avian Respiratory Complex Ii](#)
 5. [Bacillus Circulans Strain 251 Cyclodextrin Glycosyltransferase](#)
- B. The following list of proteins are "**difficult targets**", that is very difficult to find templates.
 1. Their sequence identity might be less than 30%.
 2. **Select at least one target protein from the list below, generate and analyze a model.**
 - a. [MG3368B_protein1.fasta](#)
 - b. [GE1527P_protein2.fasta](#)
 - c. [RK10653A_protein3.fasta](#)
 - d. [MH7542A_protein4.fasta](#)
 - e. [PX15502A_protein5.fasta](#)

VII. Predicting a Protein's Structure

- A. In this exercise, you will model the structure of a protein and gauge the quality of the model. This protein does NOT have a solved structure.
 1. Either select a target protein from the list below and copy its sequence or use the sequence of your favorite protein.
 - a. [Protein containing aminopeptidase domain](#)

- b. [CHICK Xenobiotic receptor](#)
2. Recognition
 - a. Use the *Template Identification* option to select a template for your FASTA sequence.
3. Alignment
 - a. Examine your alignment. If you think it is good, copy it. If not, use the suggestions in the challenge portion of the Alignment Section to refine the alignment of your protein to its template.
4. Modeling
 - a. Use the Alignment Mode in Swiss-Model to model your target protein.
5. Modeling Analysis
 - a. View your model
 - b. Superimpose your model to the template
 - c. Gauge the quality of your model using the following:
 1. Use Geometry-based tools
 2. Use empirical energy-based tools
 3. Use functional considerations

Useful Links

Programs and Webservers Used in These Exercises

[Swiss-Model](#)
[ProteinDataBank](#)
[Swiss Pdb Viewer \(aka DeepView\)](#)
[PyMol](#)
[Jmol](#)
[FATCAT](#)
[DALI](#)
[VAST](#)

There are many other programs and websites for modeling. You can also try advanced modeling tools available from [The Joint Center for Molecular Modeling \(JCMM\) - Server Links](#)

There are some more links in the [Review of the Steps in Comparative Modeling](#).

For more information about how to use the Swiss-Pdb Viewer, click on the [Swiss-Pdb Viewer Tutorial](#). It also includes a **tutorial on homology modeling**.

Contact us

Peter Lee [peterlee-at-scripps.edu]
 Jeong Hyun Lee [jhyunl-at-scripps.edu]
 Adam Godzik [adam-at-burnham.org]

[Lecture Notes 2009](#) | [Overview](#) | [Create Account](#) | [1st Model Structure](#) | [General Modeling Links](#) | [Advanced Analysis](#) | [How Accurate?](#) | [Predicting Contacts](#)