MODELLER - II - Chimera GUI interface

Jean-Yves Sgro

April 11, 2017

Contents

1 Learning Objectives

- Upload a sequence from UniProt to Chimera
- Perform a Blast search against the PDB
- Select a template
- Build a model
- Edit, evaluate and refine the model

2 Introduction

The MODELLER software[1] is used for homology or comparative modeling of protein three-dimensional structures (Webb and Sali 2016, Marti-Renom et al. (2000)) based on an alignment of a sequence to be modeled with known related structure(s). `MODELLER` automatically calculates one or multiple models containing all non-hydrogen atoms based on a comparative protein structure modeling algorithm by satisfaction of **spatial restraints** (Sali and Blundell 1993, Fiser, Do, and Sali (2000)).

`MODELLER` is a set of python scripts and tables that can perform many additional tasks, including *de novo* modeling of loops in protein structures, optimization of various models of protein structure [...].

Therefore the default method to `MODELLER` is via python scripts and line command.

There are various commercial and non commercial packages that exist as "wrappers" around `MODELLER` to make use of a Graphical User Interface (GUI) and this workshop is dedicated to the GUI offered by the molecular graphics software UCSF Chimera[2] (Pettersen et al. 2004).

3 MODELLER process

Regardless of the method used, the steps from sequence to model are the same. However, it should be noted that the quality of the alignment between the target sequence and the 3D template(s) will determine the quality of the final model.

The main steps of the procedure can be visually summarized:

Figure 1.

In more details, homology modeling is a multi-step process:

1. Template recognition and initial alignment
2. Alignment correction
3. Backbone generation
4. Loop modeling
5. Side chain modeling
6. Model optimisation
7. Model validation
8. Iteration to correct mistakes of flaws (if any)

4 Prerequisites

Attendees should have some basic knowledge of using Chimera and MODELLER from personal experience or from attending previous workshops on this subject.

[1] https://salilab.org/modeller/
[2] http://www.cgl.ucsf.edu/chimera

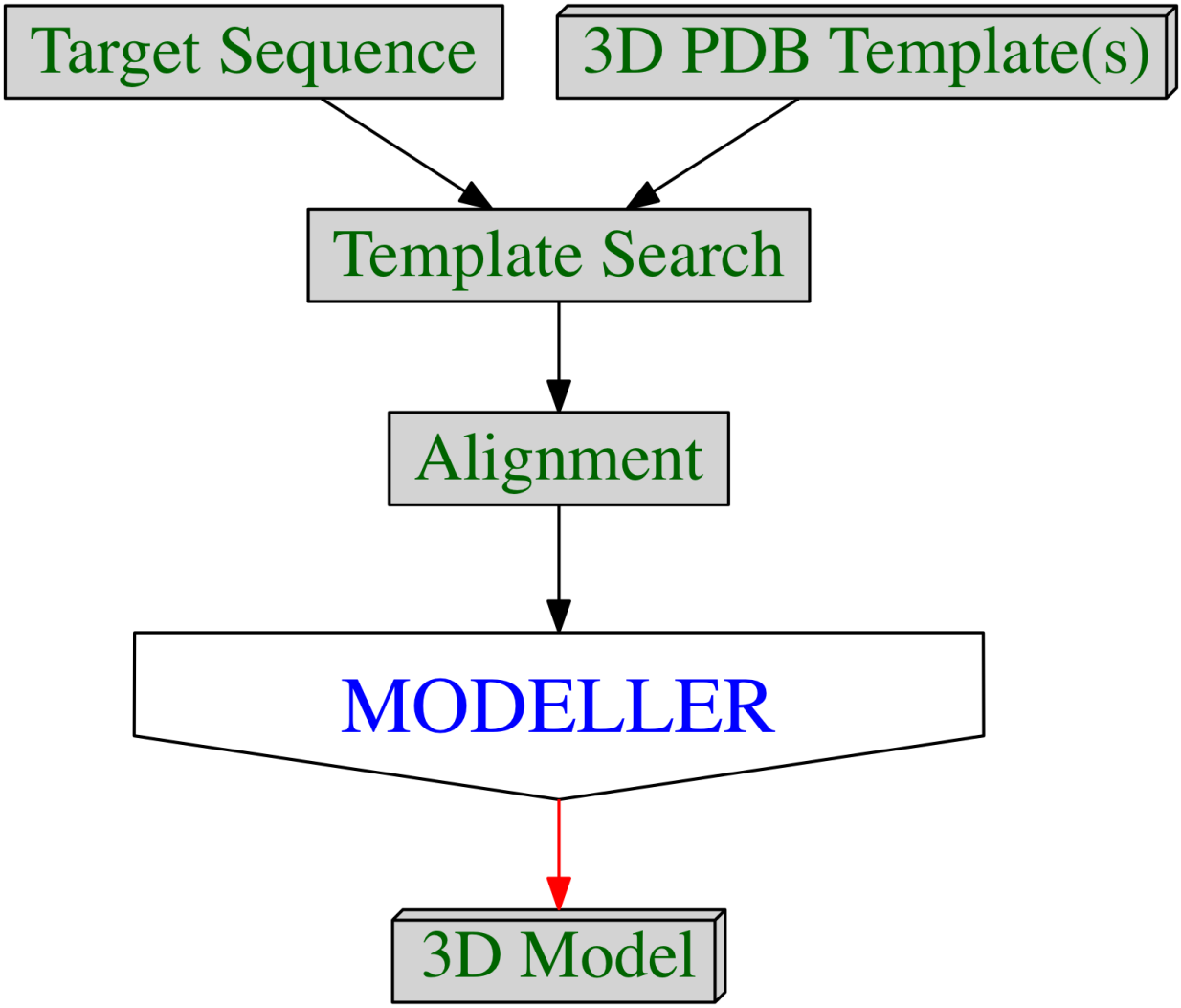

Figure 1: MODELLER process flow.

4.1 Register on MODELLER web site:

The current release of Modeller is `9.18`, which was released on Feb 22nd, 2017 and is intalled on all the iMacs. However, each user should register with the web site to obtain the install keyword at https://salilab.org/modeller/registration.html

5 Set-up

5.1 Target

In comparative (homology) modeling, theoretical models of a protein are built using at least one known related structure and a sequence alignment of the known and unknown structures. The protein to be modeled is the ***target****, and a related known structure used for modeling is a* ***template****.*

The target in this tutorial is the human GADD45 beta. This protein is involved in the regulation of growth and apoptosis. Its gene is a member of a group of genes whose transcript levels are increased following stressful growth arrest conditions and treatment with DNA-damaging agents. The genes in this group respond to environmental stresses by mediating activation of the p38/JNK pathway.

At the time of creating this tutorial (May 2016), no structure was available for GADD45 beta.

Its recommended name is *Growth arrest and DNA damage-inducible protein GADD45 beta* (Uniprot ID: O75293 (GA45B_HUMAN), gene name: GADD45B.)

5.2 Chimera

For this workshop we will use the iMacs onto which have been installed both `Chimera` and `MODELLER`. However, the process, method and graphical interface would be exactly the same on a Windows or Linux computer.

TASK

This button will invite you to act on **Open the `Chimera` software**.

- On a Macintosh `Chimera` is installed in `/Applications`.
- Double click on the Chimera icon.
- You can also find `Chimera` by using `Spotlight Search`: the "magnifying glass" icon on the top right of the Mac.

5.3 Tutorial caveats:

- *As database contents and web services are updated, the results of calculations are likely to differ from what is described here. However, this tutorial is meant to illustrate the general process rather than any specific result.*
- *This tutorial is not meant to indicate the optimal parameter settings for comparative modeling, as these will vary depending on the system of interest and the information available at the time.*
- *This tutorial describes only one of several equally valid approaches. For example, the sequence of the target could be read from a FASTA file instead of fetched from the UniProt database.*

A few of the commands will be shown as line commands. Therefore the line command option needs to be engaged, we can do it now for later:

TASK

This button will invite you to act on **Show `Chimera` line command**.

- `Favorites` > `Command Line`

This will open a `Command:` strip at the bottom of the graphical window where line commands can be typed or pasted in later.

6 Homology modeling of GADD45 beta.

6.1 Target sequence to model

TASK

This button will invite you to act on **Fetch the protein sequence from Uniprot with `Chimera`**.

Use the following menu cascade to fetch the protein sequence to model (the *target*.)

- `File` > **'Fetch by ID...**
- Click on `Uniprot` at the bottom of the new window and enter the sequence of the target, the human GADD45 beta UniProt ID: **GA45B_HUMAN**
- Click `Fetch` at the bottom

Note: clicking on `Web Page` would open the Uniprot page on a your default Internet Browser window (usually Safari on a Mac.) This is a useful way to verify that the sequence is the one you really wanted.

Two new window will open:

- The sequence in one window with the protein name
- A "Region Browser" window with properties of the sequence. The `S` column checkboxes in the Region Browser can be used to show feature annotations as coloured boxes in the sequence window. This window can be closed with the `Close` button. It can be accessed at any time from the sequence window `Info` menu.

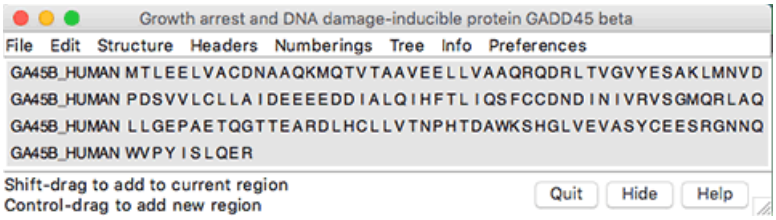

Figure 2: "***Window with 'target' sequence.***"

6.2 Tempate search

Different methods can be employed to search by template, some could be outside of `Chimera`. For this tutorial we will use the `BLAST Protein` tool search built in `Chimera` against the Protein Data Bank in search of sequences similar to the target from this set of known 3D structures.

TASK

This button will invite you to act on **Open `BLAST` search**.

- On the sequence window: `Info` > **'Blast Protein...**
- Click `OK` on the next window with the suggested name `GA45B_HUMAN`
- Click `OK` on the `Blast Parameters` window that opens to keep default settings: "Database `pdb`", "E-value: `3`", "Matrix: `BLOSUM 62`", "List only best-matching chain per PDB entry."
- The results will open in a new window

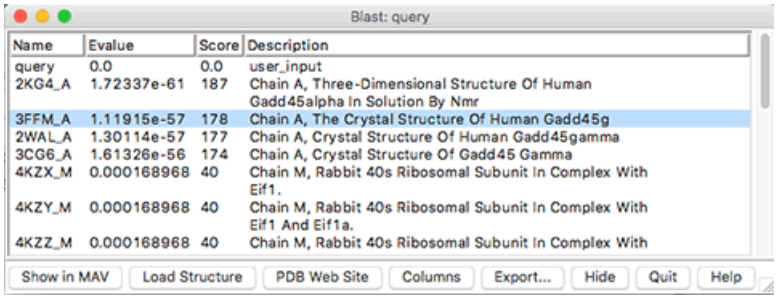

Figure 3: "***BLAST results.***"

The `BLAST` search should propose multiple templates listed from best to worst.

The two best hits are two other members of the same protein family, **GADD45 alpha** and **GADD45 gamma**, that have been already solve, the first by NMR and the second by X-ray crystallography.

(*Note*: In the future, if this exact gene product is solved and deposited in PDB, an entry might be proposed for a "real" structure. For the purpose of learning, that structure could be omitted to continue with the tutorial.)

6.3 Template selection

TASK

This button will invite you to act on **Select template.**.

In the Blast results dialog:

- Click the `Columns` button to reveal several checkboxes for controlling which columns of information are shown.
- Hide (uncheck) `Description`
- Show `Resolution`
- Show `Chain names`.

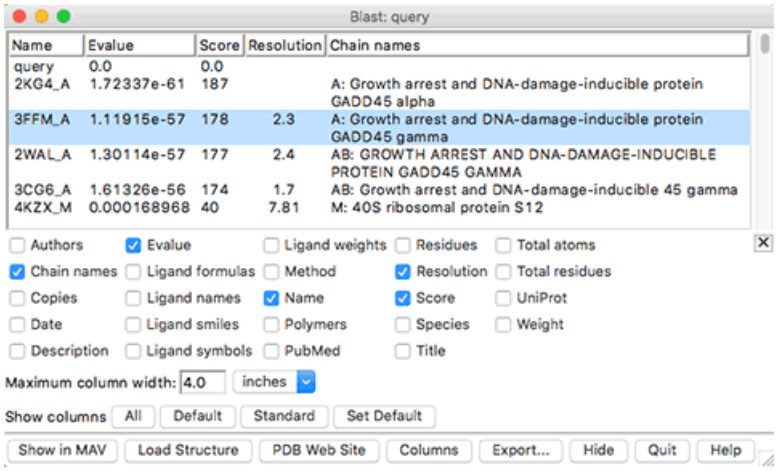

Figure 4: "***Change columns shown on `BLAST` results.***"

Some decisions:

- `MODELLER` can compute models based on multiple templates but we will use only one for simplicity here.

- NMR structures contain multiple models which can make things too complicated here.
- Therefore we will just use `3FFM_A` (PDB entry `3FFM`, chain `A`).

6.4 Align and show template structure

TASK

This button will invite you to act on **Input template as sequence and structure.**.

- In the `BLAST` results dialog, click to highlight the row for `3FFM_A`
- At the bottom of the window: click `Show in MAV` to display the query-hit sequence alignment from `BLAST` in another Multalign Viewer (MAV) window that will open.
- click `Load Structure` the `BLAST` results window: this will automatically fetch the `3FFM` PDB structure and display it within the `Chimera` graphics window as a ribbon.
- Click `Quit` to close the `BLAST` results window.

To more easily find the N- and C-terminii we can color the ribbon with a rainbow gradient of colors ranging from bue (Nt) to red (Ct) with either of the following methods:

- Menu: `Presets` > `Interactive 1 (ribbons)`
- Line command: `rainbow`

Letting the mouse "hover" over the graphical display the name of the amino acid, number and chain will be displayed, for example for the existing Nt: Serine 15 from chain A as shown.

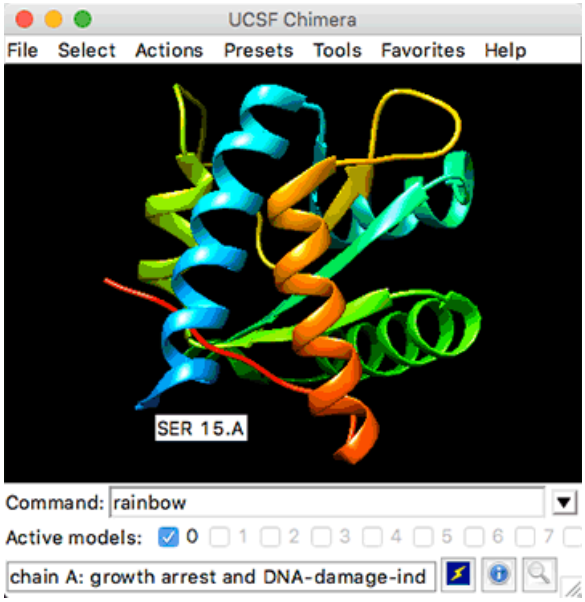

Figure 5: "***Template structure `3FFM` chain A.***"

6.5 Alignment verification

Comparative modeling requires a template structure and a target-template sequence alignment. ***The sequence alignment is crucial; it determines which residues in the template are used to model which residues in the target****, and any inaccuracies in the alignment will result in the use of incorrect constraints during 3D modeling. Regardless of how the sequence alignment was obtained, it should be examined and adjusted as needed before initiating the more computationally intensive 3D modeling calculations.*

An alignment was generated above as a by-product of the `BLAST` *similarity search. However,* `BLAST` *is meant to identify local similarities quickly rather than to give accurate full-length alignments.*

The alignment from `BLAST` is not necessarily adequate for modeling as it may miss segments. In that case the target-template(s) sequence alignment can be generated in some other ways *e.g.* with external software on the web or on the command line (not shown here.)

6.6 Alignment correction/optimization

We will create a new alignment based on the Needleman-Wunsch global alignment algorithm used to to align protein or nucleotide sequences (Needleman and Wunsch 1970).

TASK

This button will invite you to act on **Create new target-template alignment**.

- Click `Quit` to close the sequence alignment from `BLAST`
- Locate the original GA45B_HUMAN (target) sequence window.

If you had closed the sequence window: the target sequence can be fetched again as before or using in the following `Chimera` command line:

```
Command: open uniprot:GA45B_HUMAN
```

- From the sequence window menu choose: **`Edit... > Add Sequence`**.

The resulting dialog contains tabs for different ways of obtaining the sequence.

- Click on the `From Structure` tab.
- Verify that the default sequence name is `3FFM` (at the top) and at the bottom a pull-down menu should show the association of the sequence and the structure as `3FFM (#0) chain A`.
- Click on `OK`.

Within the sequence window, the association of sequence and structure is indicated with a colored box around a sequence name derived from a structure, here `3FFM`.

Note: PDB files headers contain the complete sequence of a solved protein (`SEQRES` lines) even if not all amino acids are present in the 3D Structure. **Red outline** boxes enclose residues that are in the sequence but not in the associated structure. 14 residues are missing at the N-terminus in the structure.

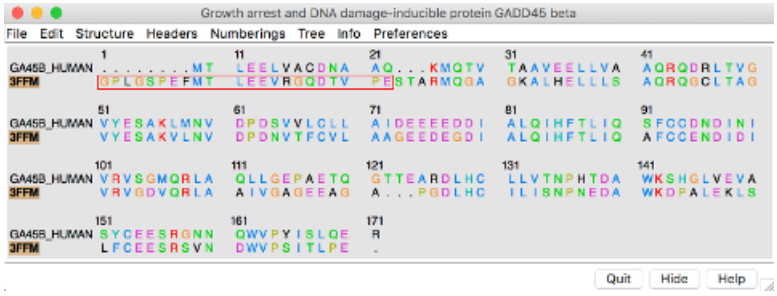

Figure 6: "***Sequence for structure*** `3FFM` ***chain A is added. AA not part of the structure are boxed in red (Nt).***"

Note: The default color scheme of the sequence alignment is that of `ClustalX` and can be changed with the menu found on the sequence alignment window: **`Preferences > Appearance`**, then at the bottom left in the box named `Residue letter coloring - Color scheme:` the options are : `black`, `ClustalX`, `ribbon`, `Kyte-Doolilttle hydrophobicity`. There also more options to change the appearance of the sequence: font size, line length etc.

Coloring the sequence to match the structure ribbon is only one of several approaches for sequence-structure mapping. For example:

1. *highlighting residues in the sequence with the mouse selects them in the structure*
2. *selecting residues in the structure highlights them in the sequence (green boxes)*
3. *structure helix and strand assignments can be shown on the sequence with* menu **`Structure... > Secondary Structure... > show actual`** - this will place a yellowish box behind the sequences that are in alpha helix and a greenish box behind those in beta sheet.

6.7 Model building

We are now ready to ask `MODELLER` to build a model for our target sequence based on the structure template.

There are 2 ways to run `MODELLER`:

1. natively on your computer if it is installed (it is on the classroom iMacs)
2. via a web service, which could be slower.

Note: Registered users (free for academics) obtain a `KEY` that is used upon installation of `MODELLER` on your computer, or provided to `Chimera` if the web service is requested.

TASK

This button will invite you to act on **Run Modeller**.

From the sequence window menu:

- Choose **`Structure... > Modeller (homology)...`** to open the `Chimera` interface to comparative modeling with `MODELLER`.
- The target should be set to `GA45B_HUMAN` (top right)
- Click `3FFM` in the dialog to choose it as the template.
- Click the Advanced Options button to reveal additional settings. This is where the number of alternate models can be requested. We will keep the default of `5`.

We will run `MODELLER` locally and it is necessary to specify where the software is located: **`/usr/local/bin/mod9.18`**. Note that if `MODELLER` is updated the numbers following `mod` would be updated as well.

- Click `Run Modeller locally` and enter `/usr/local/bin/mod9.18`
- Click `OK` to launch the model building.
- You can see the progress of calculation at the bottom of the main graphics window reported as percent completion.

When all models have been generated they will be opened within the `Chimera` graphics and a `Modeller Results` window will open with the evaluation scores.

By default the last model is selected and shown within the graphics display and highlighted within the result window. The models can be viewed individually or collectively by choosing rows in the dialog with the mouse (use `SHIFT` to select contiguous or `CTRL` to select non contiguous list names.)

The different scores from Modeller use different criteria and will not necessarily agree on which models are best:

- `GA341` - *model score derived from statistical potentials; a value > 0.7 generally indicates a reliable model, >95% probability of having the correct fold*
- `zDOPE` - *normalized Discrete Optimized Protein Energy (DOPE), an atomic distance-dependent statistical score; negative values indicate better models.*

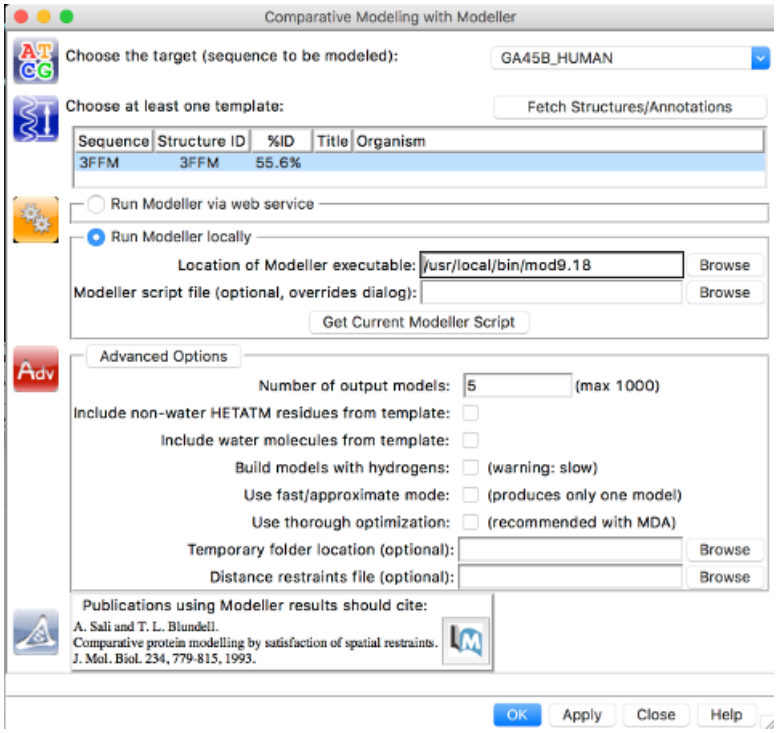

Figure 7: "*Comparative* `MODELLER` *interface.*"

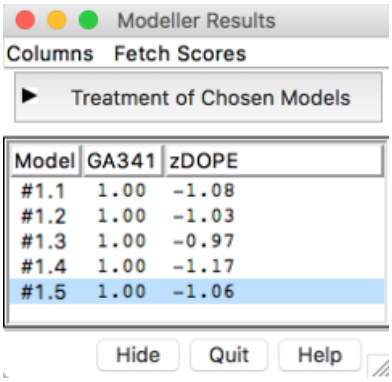

Figure 8: "`Modeller Results` *window displays evaluation scores..*"

6.8 Remove untemplated terminii

TASK

This button will invite you to act on **Remove untemplated N- and C-termii**.

> *Note*: the `GADD45 beta` sequence was longer than the template structure at the Nterminal (see above and red-lined boxes). *Therefore the model termini extend beyond the template structure and consequently cannot be modeled reliably. Displaying all the models at once shows little conformational variability except in the N terminus (GADD sequence residues 1 to 11). This conclusion is reinforced by the RMSD histogram in the sequence window, where bar heights indicate root-mean-square distances among the atoms of the residues associated with a column.*

Although there is also a `Chimera` *interface to* `MODELLER` *for untemplated building and refinement, in this tutorial we will simply remove the termini and rescore the models.*

Note: The NMR model `2KG4` that we did not use as a template is an NMR structure with 20 NMR model solutions, where it could be noted that the Nt are similarly moving in all directions and not not exhibit specific secondary structure. however they have a more consistent direction. Therefore it is for now best to remove the terminii.

Click into the sequence window, then move the cursor over the residues to see the corresponding structure residue numbers near the bottom of the window. In the comparative models (#1.1-5), Lys-15 is aligned with the first residue in the template structure (Arg-18) and Glu-159 is aligned with the last residue in the template structure.

Delete the termini in the comparative models that extend beyond the template. This will delete 14 aa at Nt **and** one aa at Ct:

```
Command: del #1:start-14,160-end
```

All models are contained within model `#1`. You can see this by opening the model panel with the main menu: `Favorites` > `Model Panel`.

6.9 Rescore models

TASK

This button will invite you to act on **Rescore models**.

To rescore the models find the `Modeller Results`:

- Choose `Fetch Scores` > `zDOPE and Estimated RMSD/Overlap`
- Rescoring uses a web service provided by the Sali lab at UCSF.
- **NOTE**: THIS IS A CASE WHERE YOU NEED TO ENTER THE `KEY` LICENSE

After a minute or few, more favorable zDOPE values are obtained, along with the additional scores:

- *Estimated RMSD - TSVMod-predicted C-alpha root-mean-square deviation (RMSD) of the model from the native structure*
- *Estimated Overlap (3.5 Å) - TSVMod-predicted native overlap (3.5 Å), fraction of C-alpha atoms in the model within 3.5 Å of the corresponding atoms in the native structure after rigid-body superposition*

6.10 Model analysis in `Chimera`

The comparative models are calculated at the atomic level and can be further analysed and displayed within `Chimera` as any structure downloaded from the PDB.

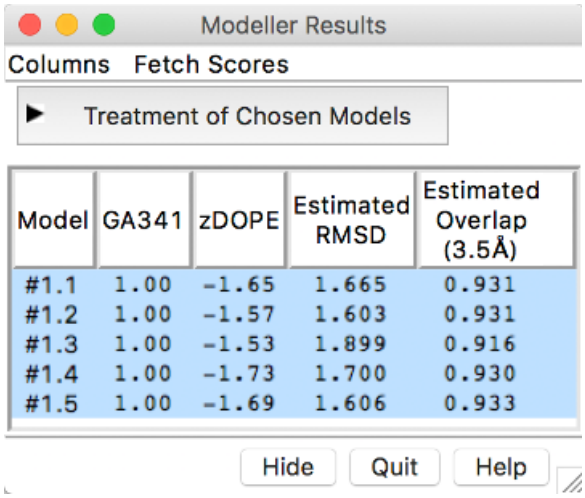

Figure 9: "***`Modeller Results` window displays evaluation scores.***."

Here is an example of showing the hydrophobicity of the model. A similar display with molecular surface could be had via the main menu **`Presets` > `Interactive 3 (hydrophobicity surface)`** but the following method uses the line command:

```
Command: ~modeldisp #0
Command: disp
Command: ~ribbon
Command: rangecol kdHydrophobicity min dodger blue mid white max orange red
Command: preset apply pub 1
Command: repr sphere
```

Commands explanations: `~` creates the opposite effect of any command *e.g.* `~disp` UN-display

- `~modeldisp #0`: UN-display model 0
- `disp`: display stick models for all amino acids
- `~ribbon`: UN-display the ribbon for all models.
- `rangecol...`: command to create a color range (blue white orange) based on hydrophobicity property
- `preset apply pub 1`: equivallent to mouse menu **`Presets` > `Publication 1 (silhouette, rounded ribbon)`**
- `repr sphere`: show atoms as sphere. Equivallent to mouse menu **`Actions` > `Atoms/Bonds` > `sphere`**

The coloring and representation will apply equally to models 1-5 as part of model 1 (`#1`) which can be selected to view from the `Modeller Results` panel.

7 Homology modeling of naja toxin

Now that we know the steps to create a model and IF there is additional time during the workshop we could explore more options based on a tutorial by Jean-Didier Maréchal[3] "Tutorial 1.3 - Homology modeling."

The gists of the tutorial is very similar to the exercise we have just done and is reproduced almost *verbatim* with minimal editing. No **TASK** button is added, just follow the flow of instructions.

[3]http://gent.uab.cat/jdidier/content/tutorial-13

7.1 Learning objectives:

The objective of this tutorial is to generate a 3D model of a snake toxin (naja snake) using the homology modeling program Modeller.

1. Generate 3D models using a comparative modeling software 2.Appreciate the quality of the models produced in function of the sequence similarity between the target and the template
2. See how missing elements in the sequence alignment influence the resulting models
3. Run tools for the validation of the structural quality of the results

7.2 Target sequence

Pir file of the sequence of naja toxin

Provided on the web site:

```
>P1;3176829
3176829
LKCNKLVPLFYKTCPAGKNLCYKMYMVAMPKVPVKRGCIDVCPKSSLLVKYVCCNTDRCN*
```

Alternate FASTA format contains 2 more aa at the Nt:

```
>AAC61316.1 cardiotoxin 4B precursor, partial [Naja sputatrix]
YTLKCNKLVPLFYKTCPAGKNLCYKMYMVAMPKVPVKRGCIDVCPKSSLLVKYVCCNTDRCN
```

7.3 Template search

Note: Difference from pervious exercise: the sequence is copied/pasted rather than fetched from Uniprot.

- Copy the sequence of the target.
- Open Chimera.
- Go to **`Tools` > `sequence` > `blast protein`**, and select `Plain Text`
- Paste the sequence into the `Sequence` box
- Choose Program: `psiblast`
  Database: `pdb`

- Click `Apply`.

The list of templates of your sequence is listed and sorted in function of their score. You can now download their structure to your chimera environment by selecting them and clicking on `Load Structure` on the bottom of the window.

Note: Since the original publication of the tutorial the proposed template is no longer the first on the list.

We know from the previous exercise how to add columns on the results. This will be useful to continue selection:

- Click on `Columns` at the bottom
- UN-select `Description`
- Click on: `Name`, `Resolution`, `Total residues`
- Note that the length of the `P1` sequence we provided is `60` and therefore a template with `60` residues would be a good match.
- Select `2CRT_A`
- Click `Load Structure` and look at its geometrical features within the graphical window.

7.4 Generate alignment

After template(s) have been determined we first have to generate the alignment between the target and the template(s). The alignment could be done with many programs, but here we'll use the `UCSF Chimera` alignment interface

- Select Select `2CRT_A` on the `BLAST` results
- Click `Show in MAV` button, at the left bottom of the window.
- The `MultiAlgin Viewer` window now appears.
- Look at the alignment.
- Discuss its quality.

Note: The alignment will show a consensus and just a few differences can be found

```
query       LKCNKLVPLFYKTCPAGKNLCYKMYMVAMPKVPVKRGCIDVCPKSSLLVK 50
2CRT_A      LKCNKLVPLFYKTCPAGKNLCYKMFMVATPKVPVKRGCIDVCPKSSLLVK 50
            ************************:*** ********************

query       YVCCNTDRCN 60
2CRT_A      YVCCNTDRCN 60
            **********
```

Note: Due to the high identity of sequences and lack of gap(s), it seems not necessary to perform a better alignment as we did on the previous exercise.

7.5 Generate model

Now, we can perform the modeling.

- On the MAV window, select `Structure` **>Modeller(Homology)‘**.
- This will open the `Comparative Modeling with Modeller` window.
- On the top, `Choose the target` should be already set to `query`. Since we pasted the sequence earlier it is simply called `query`.
- `Choose at least one template:` Click on `2CRT_A` within the list.

Note that the identity level is extremely high at 96.7%

- Choose `Run Modeller locally` and enter the location of the program at `/usr/local/bin/mod9.18` (likely already there from the previous exercise.)
- Click `OK`
- Calculation progress appear as `%` progress at the bottom of the graphical window and should be very quick.
- When `MODELLER` is finished a results window with scores appear with 5 models.

7.6 Advanced modes

The tutorial continues with more advanced options that we may not have time to review in class. They are below for complete records.

7.6.1 Alternate templates

Repeat the same process and the same discussions with pdb `3PLC` and `1JGK`.

Notes: - `3PLC` contains three identical chains. - `1JGK` is an NMR structure with `19` versions of the structure.

Suggestion: rather than use one template at a time try to opt for creating a model with multiple templates, perhaps using some others than those suggested here.

7.6.2 Advanced options

In the window from **`Structure` >Modeller(Homology)`** click on`Advanced Options‘ at the bottom.

- Ask a calculation performing `Build models with hydrogens`.
- Appreciate the time required for this option regarding the former one.
- Analyse the differences betweens models generated with and without hydrogen atoms in the force field.

Suggestion: if you have time you can try the option `Use thorough optimization: (recommended with MDA)`

`MDA`: Multidomain Assembler (Hertig et al. 2015)

7.6.3 Applying restraints

`MODELLER` works by a procedure called of satisfaction of spacial restraints. Sometime the user can benefit from several experimental or common knowledge to guide certain restrictions. To do so, `MODELLER` offers a while panel of user driven restraint.

In `Chimera` you can provide your own constraints in a file that you can import into the interface. The file should have the format like:

```
23  89  30  0.1
435-677 901-989 50  0.1
55  443-502 78  0.1
```

corresponding to:

```
res1  res2  dist  stdev
```

where `res1` and `res2` are residue numbers or ranges of residue numbers in the target sequence, `dist` is the distance in Å, and `stdev` is the standard deviation.

If a single residue is specified, its C-alpha will be used to anchor the restraint. If a residue range (*e.g.* `233-275`) is specified, the center of mass of the range will be used to anchor the restraint.

With the best model you obtained, try a constraint between atoms in two successive loops.

7.6.4 Analysis

You can pursue with analytical tools as provided in the slides on the quality of the model.

7.6.5 Loop Modelling

The final part of our modelling, will be dedicated to optimize the structural quality of the loops. We will start with a X-ray structure which has been partially derived.

- Open the `3PLC` structure.
- From command line: `open 3PLC`
- Select two of the units and remove them.
- From command line (after the selection is done) : `del sel`
- Then focus on the final one.
- From command line: `focus`

You will see that there is a missing part of the molecule (from 32 to 36). We will try to generate a model which completes this part of the molecule.

Loop modeling interface can be reached in the `MAV` window or from the main UCSF `Chimera` menu. In the laster case, go to **`Tools` > `Structure Editing` > Model/Loop refinment‘**.

New windows pop up: one with the sequence, the other with the loop modelling interface.

The sequence window already shows you the missing part of the pdb file (in red blocks). This will be one of the priority of the work.

The Loop modelling window gives you all the options for performing the simulation and works quite similarly to the pure `MODELLER` application.

Discuss with the teacher the different options for the modeling and test the quality of the results.

8 Model validations (general info)

The quality of the model will depend on the quality of the alignment and that of the template(s) used amongst many other variables.

The validity and plausibility of the model can be assessed with various software shown below.

Quoting from: https://www.researchgate.net/post/How_to_validate_our_homology_model [accessed Apr 11, 2017].

You can use:

- `PROSA`, `ANOLEA`, `VERIFY3D`, `RAMPAGE`, `ERRAT`, `SWISSPROT Tools`.

If you want to check/know whether the model predicted is secondary structure you can use `GOR V`, `JPRED`, `SOPMA`.

Note: UCLA has a functional web site with many tools cited above http://services.mbi.ucla.edu/SAVES/ with additional: `WHAT_CHECK`

Additional suggestions:

`PROSA2` https://prosa.services.came.sbg.ac.at/prosa.php : compare the global energy profile of model to energy profiles of a non redundant set of good quality models.

`PBDsum` will generate a full set of PDBsum structural analyses http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html

9 Further resources

9.1 Course

- Complete protein homology modeling online course: http://swift.cmbi.ru.nl/teach/HOMMOD/

9.2 Alternate software

- List of alternate homology protein software and web servers: https://bioinformatictools.wordpress.com/2012/04/10/homology-modeling-of-proteins/ (archived: http://bit.ly/2prhNkX)

Brief table summary:

Table 1: Protein structure prediction software

Software	Web link	Notes
CPHmodel	http://www.cbs.dtu.dk/services/CPHmodels/	use of single template
MODELLER	http://www.salilab.org/modeller/	modeling by satisfaction of spatial restraints
SWISS-MODEL	http://swissmodel.expasy.org/	fully automated protein structure homology-modeling server
Phyre2	http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index	Protein Homology/AnalogY Recognition Engine - pronounced as 'fire' - web-based services
HHpred	http://toolkit.tuebingen.mpg.de/hhpred	more sensitive in finding remote homologs
LOMATES	http://zhanglab.ccmb.med.umich.edu/LOMETS/	Local Meta-Threading-Server - web service collecting high-scoring target-to-template alignments from 8 locally-installed threading programs (FUGUE, HHsearch, MUSTER, PPA, PROSPECT2, SAM-T02, SPARKS, SP3).
MODBASE	http://salilab.org/modbase	database of annotated comparative protein structure models.
Robetta	http://www.robetta.org/	Robetta provides both *ab initio* and comparative models of protein domains.
chunk-TASSER	http://cssb.biology.gatech.edu/skolnick/webservice/chunk-TASSER/index.html	combines threading templates from SP3 and *ab initio* folded chunk structures (three consecutive segments of regular secondary structures). It is better for extreme hard targets
PSiFR	http://psifr.cssb.biology.gatech.edu/	Protein Structure and Function predicton Resource: integrated tools for protein tertiary structure prediction and structure and sequence-based function annotation. Various methods used : TASSER (Threading/ASSembly/Refinement), TASSER-Lite, PROSPECTOR_3, METATASSER, SPARKS2
ESyPred3D	http://www.fundp.ac.be/sciences/biologie/urbm/bioinfo/esypred/	increased alignment performances - 3D model made by MODELLER
PMP	http://www.proteinmodelportal.org/PMP	Protein Model Portal access to different partner sites for model building, and quality assessment.
ProModel	http://www.vlifesciences.com/products/VLifeMDS/Protein_Modeller.php	complete package for modeling proteins for Linux and Windows
SCWRL4	http://dunbrack.fccc.edu/scwrl4/index.php	use of rotamer library.
VADAR	http://vadar.wishartlab.com/	Volume, Area, Dihedral Angle Reporter is a compilation of more than 15 different algorithms and programs

Software	Web link	Notes
IntFOLD	http://www.reading.ac.uk/bioinf/IntFOLD/	3D modeling, quality assessment, [...] Prediction of protein-ligand binding residues.
PEPstr	http://www.imtech.res.in/raghava/pepstr/	server predicts the tertiary structure of small peptides with sequence length varying between 7 to 25 residues.
BSR	http://cssb.biology.gatech.edu/BSR	Binding Site Refinement: template-based method for the local refinement of ligand-binding regions in protein models
KeyRecep	http://www.immd.co.jp/en/product_2.html	rational molecular design when the 3D structure of the target protein is unknown. [...] estimate the characteristics of the binding site
PROTEUS2	http://wks16338.biology.ualberta.ca/proteus2/	web server, average prediction takes ~2 minutes per query sequence. able to achieve among the highest reported levels of predictive accuracy for signal peptides (Q2=94%), membrane spanning helices (Q2=87%) and secondary structure (Q3 score of 81.3%).
PSIPRED	http://bioinf.cs.ucl.ac.uk/psipred/	simple and accurate secondary structure prediction method
I-TASSER	http://zhanglab.ccmb.med.umich.edu/I-TASSER/	Web server - ranked as the No 1 server for protein structure prediction in recent CASP7, CASP8 and CASP9 experiments.
JPred	http://www.compbio.dundee.ac.uk/www-jpred/	Protein Secondary Structure Prediction server

9.3 Online tutorials:

- Homology modeling with Copenhagen Models and HHpred: http://www.cbs.dtu.dk/~blicher/Courses/Homology_modelling_tutorial.pdf
- Algorithms in Structural Bioinformatics Homology Modeling Practical: http://www.lcqb.upmc.fr/laine/STRUCT/TME/TP_homol.pdf
- Homework #10 Simple homology modeling with hhpred and Modeller: http://faculty.washington.edu/dimaio/files/hw10.pdf

10 Acknowledgments

This tutorial is based on the following online resources:

1. homology modelling with Chimera https://github.com/allegravia/homology_modelling by Allegra Via and Domenico Raimondo.
2. Tutorial 1.3 - Homology modeling http://gent.uab.cat/jdidier/content/tutorial-13 by Jean-Didier Maréchal
3. Chimera tutorial: Comparative Modeling Tutorial https://www.cgl.ucsf.edu/chimera/docs/UsersGuide/tutorials/dor.html
4. Video demonstration: Homology modeling with Modeller https://www.cgl.ucsf.edu/chimera/videodoc/Modeller/

Note: Larger chunks of mostly undedited text from the original tutorial will be printed in italics

REFERENCES

Fiser, A., R. K. Do, and A. Sali. 2000. "Modeling of loops in protein structures." *Protein Sci.* 9 (9): 1753–73.

Hertig, S., T. D. Goddard, G. T. Johnson, and T. E. Ferrin. 2015. "Multidomain Assembler (MDA) Generates Models of Large Multidomain Proteins." *Biophys. J.* 108 (9): 2097–2102. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4423039/.

Marti-Renom, M. A., A. C. Stuart, A. Fiser, R. Sanchez, F. Melo, and A. Sali. 2000. "Comparative protein structure modeling of genes and genomes." *Annu Rev Biophys Biomol Struct* 29: 291–325.

Needleman, S. B., and C. D. Wunsch. 1970. "A general method applicable to the search for similarities in the amino acid sequence of two proteins." *J. Mol. Biol.* 48 (3): 443–53. https://www.ncbi.nlm.nih.gov/pubmed/5420325.

Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin. 2004. "UCSF Chimera–a visualization system for exploratory research and analysis." *J Comput Chem* 25 (13): 1605–12. http://www.cgl.ucsf.edu/chimera.

Sali, A., and T. L. Blundell. 1993. "Comparative protein modelling by satisfaction of spatial restraints." *J. Mol. Biol.* 234 (3): 779–815.

Webb, B., and A. Sali. 2016. "Comparative Protein Structure Modeling Using MODELLER." *Curr Protoc Bioinformatics* 54 (June): 1–5.