

# RosettaAntibody3 Server

**Bold text means that these files and/or this information is provided.**

*Italicized text means that this material will NOT be conducted during the workshop*

fixed with text means you should type the command into your terminal

If you want to try making files that already exist (e.g., input files), write them to a different directory.

## Tutorial

This tutorial presents the RosettaAntibody3 program, which can be used to model the 3-D structure of an antibody variable (Fv) region starting from the sequence of the light chain and heavy chain Fv domains. RosettaAntibody uses the Chothia numbering scheme, and residue positions “H” and “L” refer to heavy and light chains, not histidine or leucine residues, respectively.

1. Create a directory within the RosettaAntibody directory called `my_files` and switch to that directory. We will work from this directory for the remainder of the tutorial.

```
mkdir my_files
cd my_files
```

2. Obtain input sequences for modeling with RosettaAntibody.

1. RosettaAntibody needs two inputs, 1) the sequence of the light chain and 2) the sequence of the heavy chain, which we need to prepare. There are a couple of things to keep in mind when using RosettaAntibody:

1. Only sequences from the Fv region can be modeled.
2. The six complementary determining regions (CDRs) are determined by conserved cysteine (Cys, C) and tryptophan (Trp, W) residues that identify the location and length of each CDR. Therefore it is imperative to check the input sequences for the inclusion of these residues so that RosettaAntibody will successfully run. This should include Cys residues at position L22, L92, H22, and H92, as well as Trp residues at position L35, H36, and H103 must be present in the input sequences.

2. As an example of how to prepare your input sequences, download **3O2W.pdb** from the Protein Data Bank and remove all peptide chains besides the heavy and light chains (chains H and L) to prepare the fasta files.

1. Go to <http://pdb.org>, and type in “3O2W” in the search tab. Click enter and once you get to the 3O2W page, click on “Download Files” and select the “PDB Format” option. Move the downloaded **3O2W.pdb** file to your working directory.

```
mv ~/Downloads/3o2w.pdb 3O2W.pdb
```

2. Clean the PDB file. We want only the amino acid sequence from the heavy (H) and light (L) chains from 3O2W. The script listed below removes any non-amino acid atoms from the pdb file for the specified chains - in this case, chains H and L. Because 3O2W is a crystal structure of only the variable region, we do not need to worry about removing Fc sequence from the heavy or light chains.

```
python ~/rosetta_workshop/rosetta/tools/protein_tools/scripts/clean_pdb.py 3O2W HL
```

3. You should have generated three files, **3O2W\_HL.pdb**, **3O2W\_L.fasta**, and **3O2W\_H.fasta**, which include the pdb file and the fasta files for the light and heavy chains, respectively. Check to make sure that the fasta sequences are correct. Below are the correct sequences for 3O2W heavy and light chains.

```

>3O2W_H
QQQLVQSGPELKKPGETVKISCKASGYMFTNYGMNWKQAPGKALKWGWINPYTGESTFADDFKGRFAFFLETSAT
TAYLQINNKNEDTATYFCARGTTIVRAFDYWGQTSVTVSSASTKGPSVFPLAPSSGTAALGCLVKDYFPEPVTVS
WNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEP
>3O2W_L
ELVMTQTPLSLPVSLGDQASISCRSSQSLVHSGNTYLHWYLQKPGQSPKFLIYKVSNRFGVDPDRFSGSGSGTDFI
LKISRVEAEDLGVYFCSQSTHFFPTFGGGTKLEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWK
VDNALQSGNSQESVTEQDSKSTYLSSTLTLSKADYEEKHKVYACEVTHQGLSSPVTKSFNRGE

```

4. Check positions of conserved Cys (C) and Trp (W) residues. The conserved Cys positions should be at chain positions L22, L92, H22, and H92, and the Trp positions should be at chain positions L35, H36, and H103. If you are not familiar with PyMOL, load the PDB file using the first command below, then in the bottom right screen, select the “S” button to view the sequence. Because 3O2W.pdb is not numbered using the Chothia numbering scheme, the conserved residues will not match the aforementioned positions. However, in the “/output\_files/” directory, the 3O2W.pdb file has been renumbered using the Chothia numbering scheme, and you can load this version while using PyMOL to view the renumbered PDB using the second command.

```

pymol ../input_files/3O2W.pdb &
load ../output_files/3O2W.pdb &

```

3. Go to the RosettaAntibody3 server.

To avoid flooding a public server with redundant jobs, we will not be running individual jobs on the ROSIE server. Instead we will use a pre-prepared run under a group account (see below). However, if you want to use the RosettaAntibody server in the future, follow the steps listed below to submit a job.

1. Go to <http://rosie.rosettacommons.org>. If you do not have an account on the ROSIE server, click the “Create an account” tab found on the upper right screen. Fill in the five boxes with your personal information and select the “Create an account” button. Check the email account you provided to register to confirm your email address. Once you have confirmed your email address, log in to the ROSIE server.
2. Go to the RosettaAntibody server by clicking on the RosettaAntibody icon at the bottom of the ROSIE main page.
3. Using 3O2W as an example of how to submit a job to the RosettaAntibody server, you would:
  1. Select the “[Submit Antibody task]” icon.
  2. Enter a job description name.
  3. To enter the Fv light chain sequence, click the “Browse” button and select the “3O2W\_L.fasta” file, which should be found in your working directory. After you have selected the light chain fasta file, click the “Upload” button.
  4. To enter the Fv heavy chain sequence, repeat the previous step, except choosing the “3O2W\_H.fasta” file.
  5. Check the “Model H3 loop” if you want to have the heavy chain CDR3 modeled. This will increase the CPU time, but it will also return a set of ten models that have gone through the high-resolution refinement step during the RosettaAntibody protocol.
  6. Un-checking the “Keep my job-data public” option will de-prioritize the scheduling of your job. Unless necessary, do not un-check this option.
  7. After you have made sure you have entered everything correctly, click the “Submit” button.

4. Analyze the results.

1. Go to <http://rosie.rosettacommons.org> and login to the ROSIE server using the username “Workshop2016” with the password “Participant”.

2. Because the job to predict the 3O2W Fab structure was submitted to the server more than a year ago, the job is no longer stored on the server. However, to become familiar with server interface results, we can still go through some of the completed jobs to view the differences in output between jobs that opted to model the CDR H3 loop or not. For this portion of the workshop, go through the finished jobs and try to find at least one job that did and did not opt to model the CDR H3 loop. To do this, click the “Queue” tab at the top of the page, then under “Current Queue”, click the “finished” and “all” options. To view only RosettaAntibody jobs, select “antibody” using the pull-down menu.
3. Select a job id number, and then scroll down the page to reach the “Results” section.
  1. If the H3 CDR region was not modeled, RosettaAntibody only returns one model, **Grafted-Relaxed-Model.pdb**, which has not gone through the high-refinement stage to reduce steric clashes and optimize dihedral angles or V\_H and V\_L interface distances. Therefore, if you choose to use a grafted model for future experiments, you may want to check for unreasonable bond distances or dihedral angles. To the right of the model in shaded purples and blues, are the CDR regions RosettaAntibody used to perform the BLAST search to identify templates with which to perform grafting. At the bottom of the screen there will be a summary table listing which PDBs were identified as grafting templates, their resolution, sequence percent identity to the input sequences, and BLAST results.
4. For jobs that chose to model the CDR H3 loop, there will also be ten models representing the ten lowest-scoring (in Rosetta Energy Units) models, which have gone through energy minimization and refinement, as well as an energy score table listing the energy score terms and loop-specific RMS values for each model. The score files are *very* big, so do not try to download any of the scorefiles during this tutorial. To save time, all output files that will be used for this tutorial have been downloaded and renamed, and are stored in the “/output\_files/” directory. You can copy these files to your working directory.

```
cp ../output_files/CDR3_model*.pdb .
cp ../output_files/grafted* .
```

5. The output of RosettaAntibody ranks the predicted models based on an energy score that is supposed to reflect relative changes in Gibbs free energy, where the lowest total energy-scoring models reflect the most stable structure. This does not necessarily indicate that the predicted structure matches the native structure, and it is important to select models that are both energetically favorable as well as close to the native structure. A good starting point when trying to identify accurate models is to perform a total score v. RMSD comparison to the native structure and selecting for models that have the lowest combined total score and RMSD. However, because RosettaAntibody does not have an input starting structure, it is not possible to calculate the RMSD from the native structure unless a starting structure already exists. In the case of this tutorial, we have the capability to calculate the RMSD using 3O2W as the native structure. For design purposes, you would want to compare at least one hundred predicted models’ score and RMSD values, but RosettaAntibody only outputs the coordinate files for the lowest ten energy-scoring models. We can only calculate the RMSD for these ten model structures, and because the models are numbered using the Chothia numbering scheme, we will use PyMOL align to manually calculate the RMSD for each model. If you do not want to manually calculate the RMSD of each model, the table **output\_RMSD.tsv** in the “/output\_files/” directory summarizes the all-atom RMSD for each model. After looking at the RMSD of each model, you will notice that the energy refinement step improves the RMSD of only one of the lowest-energy scoring models, and that all models have an RMSD of greater than three Angstroms to the original 3O2W structure. According to Sivasubramanian et al., an accurate prediction should have an all-atom RMSD of less than 1.5 Angstroms, indicating that the RosettaAntibody server failed to predict the 3O2W structure. However, one likely reason for this is that 3O2W is the antibody 1E9 bound to a transition state analog. The dataset used to cluster CDR backbone conformations for RosettaAntibody structure prediction typically includes apo forms of Fabs. In the case of the provided 3O2W predicted models, the clustering database also includes 3O2V, the apo form of 1E9, and 3O2V was used as

the top template for three of the CDR loop conformations, whereas 3O2W was used only once. Notably, 3O2V has an RMSD of 3.526 Angstroms to 3O2W, whereas 3O2V has an RMSD of 1.035 Angstroms from the Grafted-Relaxed-Model.pdb.

1. Open PyMOL to visually compare the predicted models to 3O2W.pdb.  
`pymol *pdb &`
2. Manually calculate the RMSD of each model using the PyMOL GUI.  
`align 3O2W, CDR3_model11, cycles=0`  
`align 3O2W, CDR3_model12, cycles=0`  
`align 3O2W, CDR3_model13, cycles=0`  
`align 3O2W, CDR3_model14, cycles=0`  
`align 3O2W, CDR3_model15, cycles=0`  
`align 3O2W, CDR3_model16, cycles=0`  
`align 3O2W, CDR3_model17, cycles=0`  
`align 3O2W, CDR3_model18, cycles=0`  
`align 3O2W, CDR3_model19, cycles=0`  
`align 3O2W, CDR3_model110, cycles=0`  
`align 3O2W, grafted, cycles=0`  
`align 3O2W, grafted.relaxed, cycles=0`
3. Use gedit, or your text editor of choice, to view **output\_RMSD.tsv**.  
`gedit output_RMSD.tsv &`
6. In the case where the native structure is unknown, you can look at conserved features of Fv regions that can help indicate whether or not the predicted models are realistic models or not.
  1. Once in Pymol, on the far right side, there should be a bar labeled “all” with 5 buttons labeled, “A”, “S”, “H”, “L”, and “C”. Click the button “S” in the “all” bar, and select the first “as” option, which will list additional options. Select the “As cartoon” option, which will illustrate the backbone and secondary structural elements as a ribbon diagram. All the PDBs should already be aligned.
  2. On the bottom right of the Pymol session, there should be another “S” button. Select this button which will pull up the sequences of all PDBs on the top of the screen. There is also a line “Selecting Residues”. Click on “Residues” until it the line reads “Selecting Chains”. In the upper screen where the sequences are, select the 3O2W light chain, which will highlight the light chain sequence in the viewer. From this you can see the location of the light chains and that there is little structural deviation between 3O2W and the predicted models.
  3. Go back to the “Selecting Chains” line and click on “Chains” until it reads “Selecting residues” again. This time, click on the most dissimilarly aligned region between all eleven structures. You should see that this highlights residues that are located in the CDR H3 loop (H92-H105), illustrating that RosettaAntibody is not as consistent with predicting CDR H3 conformations, but is otherwise fairly accurate in predicting all other CDR loops.
  4. In the right panel, click on all CDR3\_models except CDR3\_model1 to have only CDR3\_model1 displayed. Click on L36 (Y), L46 (F), and L49 (Y) residues to highlight their positions. Next to the “<sele>” bar, select the “S” button, then “side chain” -> “sticks”. Next select the “A” button, then “find” -> “polar contacts” -> “to others excluding solvent”. This should create yellow dashed lines of polar contacts made by L36, L46, and L49 to side chains in the heavy chain. Find which heavy chain residues these light chain residues make contact with. L36 should make a contact with the HCDR3 loop, and either L46 or L49 should make another contact with the HCDR3 loop. You should also see that the H3 loop is in a bulged conformation. Re-select 3O2W, and you will see that the torso of the model H3 loop does not match the angle of rotation for 3O2W exactly. Since we know the structure of 3O2W, we know that this is not the native conformation of the H3 loop, but all other model CDR loops should nearly match the 3O2W CDR loops.
  5. Time permitting, you can go through all ten CDR3 models repeating the above step, as well as comparing the grafted model and relaxed grafted model to 3O2W. To load the grafted

model and relaxed grafted model in the PyMol session, in the command line at the bottom of the PyMol session, type

```
load grafted.pdb
load grafted.relaxed.pdb
```

5. For better explanation of the rules used to select HCDR loop conformations, please refer to these papers:

1. Kuroda, D., Shirai, H., Kibori, M., Nakamura, H. (2008) "Structural Classification of CDR-H3 Revisited: A Lesson in Antibody Modeling". *Proteins: Struct. Funct. Bioinf* **73**(3): 608-620.
2. Morea, V., Tramontano, A., Rustici, M., Chothia, C., Lesk, A.M. (1998) "Conformations of the Third Hypervariable Region in the VH Domain of Immunoglobulins". *J. Mol. Biol.* **275**(2): 269-294.
3. North, B., Lehmann, A., Dunbrack, R.L. (2011) "A New Clustering of Antibody CDR Loop Conformations". *J. Mol. Biol.* **406**(2): 228-256.
4. Shirai, H., Kidera, A., Nakamura, N. (1999) "H3-rules: Identification of CDR-H3 Structures in Antibodies". *Febs Letters* **455**(1-2): 188-197.

6. For better explanation of the scoring and methods used in RosettaAntibody, please refer to these papers:

1. Sivasubramanian A., Sircar, A., Chaudhury, S., Gray, J.J. (2009) "Toward high-resolution homology modeling of antibody Fv regions and applications to antibody-antigen docking". *Proteins* **74**(2): 497-514.
2. Lyskov, S., Chou, F.C., Conchuir, S.O., Der, B.S., Drew, K., Kuroda, D., Xu, J., Weitzner, B.D., Renfrew, P.D., Sripakdeevong, P., Borgo, B., Havranek, J.J., Kuhlman, B., Kortemme, T., Bonneau, R., Gray, J.J., Das, R. (2013) "**Serverification of Molecular Modeling Applications: The Rosetta Online Server That Includes Everyone (ROSIE)**" *PLoS One* **8**(5): e63906. doi: 10.1371/journal.pone.0063906. Print 2013.
3. Sircar, A., Kim, E., Gray, J.J. (2009) "RosettaAntibody: Antibody Variable Region Homology Modeling Server". *Nucleic Acids Research* **37**(Web Server Issue): W474-W479.