Antibody Structural Modeling

An online version of this tutorial can be found at http://csb.vanderbilt.edu/~morettr/AntibodySummerSchool.html.

Antibody Structure Examination

If you are not already familiar with Chimera, please go through the tutorial at https://www.cgl.ucsf.edu/chimera/docs/UsersGuide/tutorials/menutut.html to get familiarized with the controls.

Notation note. Throughout this tutorial menu choices will be presented with arrows like "Select->Chain->A". This means to start with the Select menu in the main Chimera window, then select each indicated menu entry in turn.

Using "File->Fetch by ID", load the structure 1IGT. This is a structure of a complete IgG. Enable Presets->"Interactive 1" to show a ribbon diagram colored by chain. Can you identify which chains are the heavy chains and which are the light chains? Can you identify what's the Fc region, what's the Fab, and where the Fv regions are? Can you identify the individual immunoglobulin domains? (Hint: there's 12 in this structure.)

Enable the display of disulfide bonds with "Select->Residue->CYS" followed by "Actions->Atoms/Bonds->Show". (Once the cysteines are displayed, you can deselect them by ctrl-clicking on the background or by clicking "Select->Clear Selection".) Notice the location of the disulfides connecting the two heavy chains, the light and heavy chains, and stabilizing the fold of each immunoglobulin domain.

Center on chain A with "Select->Chain->A" followed by "Actions->Focus". Now look more closely at the Fv region. Notice the two sheets forming the "beta sandwich structure" of each immunoglobulin domain. Trace the backbone through the two immunoglobulin domains, identifying the CDR loops in each of the light and heavy chains. (Hint: CDR1 starts just after the disulfide, as does CDR3 [but on the other side].) Note that there's an "extra" outward-facing loop between CDR2 and CDR3 that isn't included in the list of CDR loops.

Using "File->Fetch by ID", load the following structures as "PDB (biounit)" (Note that checking "Keep dialog up after Fetch" may make things easier): 4HKX, 4M5Z, 4XRC, 4YJZ, 4YK4, 5DUP, and 5W0D. (That's a zero in the last one.) Loading as a biounit will simplify the display, as it will omit some of the duplicate copies from crystallography. These are all structure of antibodies which bind to the sialic acid binding site of influenza hemagglutinin, determined in the presence of hemagglutinin.

Align all the structures by their Fv domains.

- Use "Tools->Structure Comparison->MatchMaker" to open up the structural alignment tool.
- Select the "Specific chain in reference structure with best-aligning chain in match structure option"
- Select 4YJZ chain L as the reference chain in the upper left panel. (This is a single-chain Fv molecule.)
- Select all the structures in the right panel.
- Click Apply. Chimera will take a bit of time to align all the structures
- "Select->Chain->L->4YJZ" and then "Actions->Focus" to center on the structures.

Bring up the Model Panel with "Favorites->Model Panel". In this view you can enable and disable display of models with the "S" (Show) checkbox. By selectively showing structures, it can help reduce the "noise" in the display.

For now, let's disable the antigen structures

- Clear the current selection state: "Select->Clear Selection"
- "Actions->Atom/Bonds->hide" to hide all side chains/waters/glycans/ligands
- Set selection mode to append: Select->"Selection Mode"->append
- Select chain A of 4M5Z, 5DUP, and 5W0D; chain C of 4M5Z, and all chain Ds and all chain Es. (Via the Select->Chain submenus)
- Name the selection: "Select->Name Selection", then antigen.
- "Actions->Ribbon->hide"

Examine the similarities and differences between the various antibody structures. What's the most noticeable change between the antibodies? Can you identify the location of the CDR loops in the structure?

Re-enable the antigen. ("Select->Named Selection->antigen" and then "Actions->Ribbon->show"). Use the Model Panel to selectively display individual structures. All the antibodies are binding to the same pocket in the antigen, but are they using the same approach? How different are the relative antibody/antigen orientations in these different antibodies? (Hint: pay attention to where the short helix on the antigen is.) What's the primary contact between the antibody and the antigen? What sort of secondary contacts are there, and where do they occur in the antibody structure? (For example, compare 4XRC to 5W0D.)

Antibody Structure Prediction

We'll be modeling the Fv region of CH67, which is a monoclonal antibody which recognizes the sialic acid binding site of influenza hemagglutinin. This has been crystallized bound to hemagglutinin as PDB entry 4HKX.

The first thing we need is the sequence of the Fv region of the CH67 antibody. Go to https://www.ncbi.nlm.nih.gov/protein and search for CH67. The first two entries (4HKX_A and 4HKX_B) should be the heavy and light chains of the Fab antibody fragment crystalized as PDB id 4HKX. Click the `FASTA` link under each to get the sequence. These can be downloaded by clicking "Send to:" -> "File" -> "Create File". Save the files for 4HKX_A and 4HKX_B as `CH67_heavy.fasta` and `CH67_light.fasta`, respectively, into the working directory of your computer. (You may need to copy and rename from your default download location.)

These sequences correspond to the full Fab fragment. The RosettaAntibody server only models the Fv region of the antibody, so we need to trim the sequences to correspond to only those sections. Open both CH67_heavy.fasta and CH67_light.fasta in a text editor.

Using the text editor, remove the C-terminal portion of the CH67_heavy seqeunce from (and including) `ASTKGP` to the end. This should result in a 126 amino acid sequence, starting with QVQ and ending in `TVSS`.

- Important: The RosettaAntibody server expects the sequence to be on a single line, without spaces or newlines.
- Refer back to the 4HKX model in Chimera. Locate the endpoints and the region being trimmed on the structure.

Likewise, remove the C-terminal portion of the CH67_light sequence from (and including) `PKAAP` to the end. This should result in a 110 amino acid sequence, starting with `QSA` and ending in `TVLGQ`.

- (Note that the 3D coordinates of the 4HKX light chain are missing approximately ten residues. While missing from the 3D coordinates, these residues are present in the sequence we're using. However, such missing residues are something to be careful of when modeling other proteins.)

Go to http://rosie.rosettacommons.org/. Jobs can be submitted anonymously, but logged in users have higher priority and get email notifications. Scroll down to the "Antibody" application. Take a moment to read over the brief "Antibody Server Documentation", and then follow the link to "Submit Antibody task".

This is where you'd submit the task. Enter a short description (to keep track of the job), and upload the edited CH67_light.fasta and CH67_heavy.fasta sequences. In the settings you'll have the option to add an additional step of modeling the H3 loop. As the H3 loop is less conserved (in part due to VDJ joining), the canonical templates approach used for the other 5 CDR loops does not work as well for H3. Enabling "Model H3 loop" will add an additional refinement of the H3 loop to hopefully produce better structures, at a cost of a longer runtime.

Results Depending on server load, it may take a while for your job to run and produce results. In particular, H3 modeling can take several hours. As opposed to waiting for the job to finish, you can examine prepared results. A run without H3 modeling can be found at http://rosie.rosettacommons.org/antibody/viewjob/47831 and one with H3 modeling can be found at http://rosie.rosettacommons.org/antibody/viewjob/47986.

The results page will display a summary of the modeling run, including details of the prediction of where the CDRs are, along with statistics of how well the CDR loops and framework regions match. The result of the grafting will also be available for download as a PDB. If H3 modeling is done, it will also show the top 10 results from the H3 modeling.

- Download all results from the H3 modeling run, and compare the generated structures of the antibodies in Chimera.
  - Use the Tools->Structure Comparison->MatchMaker tool to align all the structures together.
- How close are the framework regions and non-H3 CDR loops to the crystalized structure in 4HKX?
- How much does the H3 loop vary? When comparing the no-H3-modeling result to the with-H3-modeling results, how much does the rest of the structure reorganize around the remodeled H3 loop? How close does the prediction get to the 4HKX structure?

Antibody Docking

The antibody CR6261 also recognizes influenza hemagglutinin, but in the stem domain, rather than in the sialic acid binding pocket. It also can neutralize influenza infection by preventing a conformational change required for fusion of the influenza particles with the cells it is infecting. This antibody has been crystallized with both a H5N1 and the 1918 H1N1 hemagglutinin (PDB codes 3GBM and 3GBN, respectively).

We'll be using the ROSIE SnugDock server to dock CR6261 to its binding site on influenza hemagglutinin.

Download the Fv fragment from the RosettaAntibody server. We could certainly use the structure of the crystallized version of the CR6261 antibody. However, SnugDock requires antibodies to be ordered and numbered in a particular way. While possible to do by hand or by script (e.g. by using the http://www.bioinf.org.uk/abs/abnum/ server), for convenience and to demonstrate how protocols can be pipelined, we'll be using RosettaAntibody prediction.

Download the predicted structure from http://rosie.rosettacommons.org/antibody/viewjob/48207 into your working directory, and examine it along with 3GBM and 3GBN in Chimera.

- Using File->"Fetch by ID" load 3GBM and 3GBN into your Chimera session.

Align the 3GBM and 3GBN models to the predicted mode

- Open the Tools->Structure Comparison->MatchMaker dialog
- Select the "Best-aligning pairs of chains between reference and match structure" option
- Select grafted.0.relaxed.model.pdb as the reference structure in the left panel, and 3GBM and 3GBN in the right panel.
- Click OK to align the structures.

Take a moment to examine the structures and the prediction, and see how close the prediction got. Notice that the backbone of this structure matches much more closely than that for the CH67 prediction, especially in the HCDR3 region. There's two reasons for that. The first is that the HCDR3 loop in CR6261 is smaller than that of CH67, making it easier to model. The second is that 3GBN is being picked as a template structure in modeling. (Go back to http://rosie.rosettacommons.org/antibody/viewjob/48207 and examine the report of the template selection.)

The docking approach we're taking is a cross-docking one. That is, we're using the model of the antibody bound to the 1918 H1 (3GBN) and using the general binding mode of the antibody bound to H5 (3GBM). As we already know what the answer should be, we can see this happens to be an "easy" case, as the structures are rather similar. However, if this weren't a benchmarking case, we might not know how CR6261 binds to H5, and would use known structures (e.g. bound to 1918 H1) as a starting point for refinement.

Align the model light chain with the 3GBN light chain.

- Open the Tools->Structure Comparison->MatchMaker dialog
- Select the "Specific chain(s) in reference structure with specific chain(s) in match structure" option
- Select 3GBN chain L as the reference chain in the left panel.
- Select grafted.0.relaxed.model.pdb chain L in the drop down in the right panel
- Click Apply

Align the H5 hemagglutinin with the 3GBN hemagglutinin.

- With Tools->Structure Comparison->MatchMaker still open, with "Specific chain(s) in reference structure with specific chain(s) in match structure"
- Select 3GBN chain B as the reference chain in the left panel
- Select 3GBM chain B asin the drop down in the right panel
- Click OK

Now we have all the parts aligned, we just need to combine them.

- Set selection mode to append: Select->"Selection Mode"->append
- Select chain L of grafted.0.relaxed.model.pdb: Select->Chain->L->grafted.0.relaxed.model.pdb
- Select chain H of grafted.0.relaxed.model.pdb: Select->Chain->H->grafted.0.relaxed.model.pdb
- Select chain A of 3GBM: Select->Chain->A->GBM
- Select chain B of 3GBM: Select->Chain->B->GBM
- Set selection mode to subtract: Select->"Selection Mode"->subtract
- Remove non-protein residues: Select->Residue->"all nonstandard"

Save the combined selections

- File->"Save PDB"
- Make sure the "Save selected atoms only" box is checked, and "Save multiple models in [a single file]"
- Enter the filename as "combined.pdb" in the working directory, and hit save.

SnugDock expects the protein chains in a particular order: Antibody light chain, Antibody heavy chain, antigen chain(s). Open the combined.pdb file in a text editor, and make sure the ATOM records in the PDB are in that order. Use the text editor to rearrange the order, if necessary. (Note that most lines which aren't ATOM records will be ignored by the SnugDock server, and can be safely deleted.)

- In addition, the PDBs must be numbered according to the Chothia numbering scheme. This is already the case for the results of the RosettaAntibody server, but if you're using an externally-sourced antibody structure, you may need to renumber the structure. See https://www.rosettacommons.org/docs/latest/application_documentation/antibody/General-Antibody-Options-and-Tips for more information.

Reopen the edited combined.pdb in a new Chimera window to make sure all the chains are present and aligned properly.

Go to http://rosie.rosettacommons.org/. Scroll down to the "SnugDock" application. Take a moment to read over the brief "SnugDock Server Documentation", and then follow the link to "Submit SnugDock task".

This is where you'd submit the task. Enter a short description (to keep track of the job), and upload the edited combined.pdb. You need to also specify the chain designations. Here it would be "LH_AB". This should match the order of the chains in the input PDB, with an underscore between the antibody and the antigen chains. Note that SnugDock requires the first two chains to be the light and heavy chains of the antibody, respectively, so this will always start with "LH_"

You can also specify the type of protocol to use. The thorough protocol is only marginally more accurate than the fast protocol, despite taking much more time. For this class or for trial runs, please select the fast protocol to save on server time.

Results Depending on server load, it may take a while for your job to run and produce results. As opposed to waiting for the job to finish, you can examine prepared results. See http://rosie.rosettacommons.org/snug_dock/viewjob/48217 for an example of a SnugDock run.

The SnugDock server will provide you with a selection of 10 models which it thinks are good.

It will also provide score-versus-rmsd "funnel plots". In these plots, each output model from the repeated Monte Carlo runs will be a point, with the root mean squared deviation (RMSD) to starting structure plotted on the x-axis, and the Rosetta score (more negative better) plotted on the y-axis. If the reference (starting) structure is native-like, one expects to see a "funnel" of points, where the low-energy (good) models correspond to low rmsd (native-like) structures, and as one gets progressively further away you get progressively higher energy (worse). If your reference structure isn't native like, you would hope to get low energy structures clustering around a particular rmsd which corresponds to the native structure..

The SnugDock server provides two score-versus-rmsd plots. The second ("total score models") plots the total score of the entire complex against the whole-complex rmsd. The first plots just the antibody-antigen interface scores versus the rmsd for just the (reference structure) interface residues. Each focuses on a particular aspect of the docking - either the quality of the overall docking, or the quality of the antibody-antigen interface. Ideally both should show a good "funnel", with low scoring structures by either total or interface score clustering around a single rmsd range. Additionally, which structures are low scoring for interface energy should match those which are low scoring for total energy. (This is often the case, but not always.) You can examine which structures correspond to which dots by hovering over the dots.

Download the top 10 models and examine them compared to the starting structure and the 3GBM crystal structure. (Note that the downloaded structures may be lacking the '.pdb' file extension, so they may need to be renamed, or the file format explicitly chosen in the Chimera open dialog.) To best illustrate the variation, align to chain B of the combined.pdb starting structure with Tools->Structure Comparison->MatchMaker.

How well did the server predict the overall position of the antibody? Did it adjust the HCDR3 positioning and heavy/light chain orientation to be closer to the 3GBM model versus the starting structure? (Did it need to?)

In addition to the overall structure and position, take a look at the atomistic contacts. To make things less confusing, turn off the display of hydrogens with "Select->chemistry->element-H" and then "Actions->Atom/Bonds->Hide". (Then deselect the hydrogens by control-clicking on the background.) – Alternatively, this is easier by typing "~display element.H" at the Chimera command line.

How well was SnugDock able to recapitulate the binding of the side chains? There are a number of critical aromatic residues that bind into the pocket. In the top docked model are these closer to the starting model or the 3GBM structure?

Another thing to look at is the range of sampling. Download some of the high rmsd structures from the SnugDock site. (Hover over the high rmsd points on the graph to get a tooltip with a download link for that structure in the popup.) This example was a relatively "easy" one, so models are quite converged. Other docks will have greater variation, but SnugDock is primarily a "refinement" protocol, so if there's a large difference in binding site or orientation between the starting model and the native, other approaches which do a more general search may be warranted.

Antibody Design.

To see if we can further refine the antibody to better bind the antigen, we'll attempt to redesign the antibody. We'll be using the RosettaDesign fixed-backbone redesign server.

In this design, we'll redesign CH67 (which was crystalized bound to a type H1 hemagglutinin) to increase affinity toward an H3 hemagglutinin (specifically A/Perth/142/2007)

Using either the RCSB website or the File->"Fetch by ID" feature of Chimera, download both the 4HKX (CH67 bound to hemagglutinin H1 Solomon Islands/03/2006) and 4WE6 (a dimer of the head domain of hemagglutinin H3 Perth/142/2007)

Align the hemagglutinin domain to each other.

- Open the Tools->Structure Comparison->MatchMaker tool
- Click the "Specific chain(s) in reference structure with specific chain(s) in match structure" option.
- Select the 4hkx chain E in the top left panel.
- Select the 4we6 chain A in the dropdown in the upper right panel.
- Click Apply

Select the relevant substructures to combine.

- Set selection mode to append: Select->“Selection Mode”->append
- Select chain A of the antibody: Select->Chain->A->4hkx
- Select chain B of the antibody: Select->Chain->B->4hkx
- Select chain A of the antigen: Select->Chain->A->4we6
- Set selection mode to subtract: Select->“Selection Mode”->subtract
- Remove non-protein residues: Select->Residue->“all nonstandard”

Rename the chains for the antibody. This is necessary to chang(We do this after selection, as there’s a bug in some versions of Chimera with selecting chains after renaming.)

- Open the Tools->“Structure Editing”->“Change Chain IDs” dialog
- Select the 4hkx chain A and chain B
- Rename chain A to H
- Rename chain B to L
- Click OK

Save the combined selections

- File->“Save PDB”
- Make sure the “Save selected atoms only” box is checked, and “Save multiple models in [a single file]”
- Enter the filename as “combined_design.pdb” in the working directory, and hit save.

The RosettaDesign server requires a particular format for multichain interface design. Open the combined_design.pdb file in a text editor.

- Remove all lines up to the first line starting with ATOM.
- Remove all non-ATOM lines between the H and L chains.
- Remove all lines after the last ATOM line.
- Replace all non-ATOM lines between the antigen (chain A) and antibody (chain H or L) lines with a single `TER` line.
- Save the edited file.

The RosettaDesign server has two modes relevant for antibody-antigen predictions. One is the “Increase binding affinity” mode, which looks for specific point mutations which increase the buried hydrophobic surface area - one way of increasing affinity. The other mode is a standard design mode which does combinatorial optimization of side chains.

Submit the interface design task.

- Go to http://rosettadesign.med.unc.edu
- Log in with a user name and password.
  - Accounts are free for academic users.
  - If you just want to look around, you can use the following account: user: “bu_mib_2018” password: “Participant”
- Select the “Increase binding affinity” job type (Take a moment to read over the brief increase binding affinity help page)
- Select the combined_design.pdb file to upload
- Leave the rest of the options default. (The server will automatically detect the interface.)
- Submit your job.

Submit a standard design task.

In order to submit a standard task, you'll need to specify which residues and identities you wish to use. This is particularly important for antibody/antigen designs, as you typically only want to design one side of the interface.

This specification is in the Rosetta resfile format. See https://www.rosettacommons.org/docs/latest/rosetta_basics/file_types/resfiles for the full format. The basics of the format is to specify the residues ranges you wish to act on and the flexibility you wish to allow.

Based on the combined_design.pdb, the following is a resfile which will design the antibody in the interface.

```
NATRO
start
30-33 H ALLAA
50-58 H ALLAA
99-112 H ALLAA
25-30 L ALLAA
90-95 L ALLAA
58 A NATAA
91-105 A NATAA
113-119 A NATAA
147-156 A NATAA
182-187 A NATAA
```

The NATRO at the top says to default to keeping the residue sidechains fixed ("native rotamer"). The subsequent lines specify residue ranges in the protein, and whether to mutate the position ("all amino acids") or to allow sidechain movement but not design ("native amino acid"). Finer control such as picking specific sets of amino acids is possible. The residue ranges were determined manually, by examining the structure within Chimera.

Submit the design task

- Go to http://rosettadesign.med.unc.edu
- Log in with a user name and password.
- Select the "Protein design" job type (Take a moment to read over the brief increase binding affinity help page)
- Select the combined_design.pdb file to upload
- Select the "Selected residues" radio button, and upload the design_resfile.txt file (contents listed above)
- Select the number of structures to make. RosettaDesign is stochastic, so different output structures should give slightly different results.
- Submit your job.

Results Interface binding affinity results

Depending on server load, it may take a while for your job to run and produce results. As opposed to waiting for the job to finish, you can examine prepared results. See http://rosettadesign.med.unc.edu/D/35960.zip for an example of a RosettaDesign interface run.

For the interface design runs, the results consist of two files, with data tables in each file. The [jobid]_table.txt file is the main output result, giving a curated list of mutations, and their predicted effects on binding.

Mutation	ddG_bind	ddG_partnerA	ddG_partnerB	#Neighbors	ddG_h_bond
185A N>Y:	-9.6	0.0	0.0	14	0.0

The mutation is given by the residue number and chain in the PDB, followed by the predicted effect on binding, the number of surrounding residues (a measure of burial) and the effects on hydrogen bonds. Note that the binding energy effect might be exaggerated by the fixed-backbone nature of the protocol - particular residues may start highly disfavorable due to how the complex is put together, resulting in mutations that are overly favorable.

This table is specialized, though, specifically looking to convert hydrophilic residues in the interface into hydrophobic ones, while avoiding clashes and broken hydrogen bonds. For our example, these types of mutations are primarily on the antigen side (Chain A), with few to none on the antibody side.

Looking at the [jobid]*intout.txt file can give more details about potential mutations. In particular, look at the* `score` *column. (The table should be sorted by this column.) More negative numbers indicate likely better mutations. Identify likely mutations on the antibody (H or L chains), then see where they fall on the combined*design.pdb input structure.

- You can use Tools->Structure Editing->Rotamers in Chimera to model the new structure on the input structure.

Standard design results

Depending on server load, it may take a while for your job to run and produce results. As opposed to waiting for the job to finish, you can examine prepared results. See http://rosettadesign.med.unc.edu/D/35975.zip for an example of a standard RosettaDesign run.

For the standard design runs, the primary results consist of your set of designed PDB. Open up the designs in Chimera alongside the starting structure (combined_design.pdb).

Note that RosettaDesign is a fixed-backbone protocol, so a backbone cartoon trace shouldn't show much difference between the structures. (If there's backbone difference between your starting structures and the example outputs, that's because your starting structure is slightly different than the starting structure which generated them.)

You'll probably want to turn on the side chain display with Actions->Atom/Bonds->Show. To make things less confusing, turn off the display of hydrogens with "Select->chemistry->element-H" and then "Actions->Atom/Bonds->Hide". (Then deselect the hydrogens by control-clicking on the background.) – Alternatively, this is easier by typing "~display element.H" at the Chimera command line.

Use the Model Panel (Tools->General Controls->Model Panel) to show ('S' box clicked) or hide each structure. Doing this you can examine the differences between the starting model and each design, and between the designs. Depending on the particular structure and how large and complex the interface is, there might not be all that much difference between the various designs. This sort of convergence is due in part to the fixed backbone nature of the RosettaDesign server. More advanced protocols with backbone flexibility (or running the fixed-backbone design with different starting structures) typically result in a larger range of design variants.

One limitation of the input we used was that we manually placed the two binding partners next to each other. This means that there are likely small "frustrations" in the backbone placement which can't necessarily be resolved by the design run. Using a docked complex, or a more advanced protocol which allows backbone movement during design would likely correct that.