Electronic Supplementary Material (ESI) for Molecular Systems Design & Engineering.

Protocol Capture

Introduction

The following is a protocol capture that demonstrates how to determine the stability at the HC/LC interface using an antibody/antigen co-crystal, and generate an ensemble of HC/LC docked models. In this example, we will be making models of the antibody VRC-PG04.

The version of Rosetta used for the entirety of this study is: Rosetta_2015.12.57698, released on May 5th, 2015.

All input materials for this protocol capture can be downloaded from https://github.com/ac1546/HC_LC_docking.

The ABangle software can be found at http://www.stats.ox.ac.uk/~dunbar/abangle/.

Preparing input structures

The PDB structure 3se9 was downloaded from the Protein Data Bank (https:www.rcsb.org/) and processed manually in PyMol. The gp120 component (Chain G), waters, and salt ions were removed. Next, we remove the constant region of the Fab in order to lessen the time needed to generate models. Here, we removed residues 113-216 of the heavy chain and residues 108-214 of the light chain. The molecule was saved as 3se9_Fv_clean.pdb to denote the type of fragment the pdb contains and whether or not this contains atoms that Rosetta cannot process.

Defining the HC/LC interface

Here, we use the InterfaceAnalyzer application to define an HC/LC interface using the following command:

```
/path_to_rosetta/rosetta/main/source/bin/InterfaceAnalyzer.linuxgccrelease -s
3se9_Fv_clean.pdb -tracer_data_print true -pack_input true -pack_separated true -score:weights
talaris2013.wts
```

Near the end of the output is a PyMol selection defining the residues that comprise the HC/LC interface:

```
select 3se9_Fv_clean_interface,
/3se9_Fv_clean//H/1+3+4+6+35+37+39+43+44+45+46+47+48+49+50+57+58+59+89+91+92+
93+94+99+100+100+100+100+100+100+100+101+102+103+104+105+106+108+ +
/3se9_Fv_clean//L/31+32+33+34+35+36+38+41+42+43+44+45+46+47+48+49+50+51+52+53
+55+56+57+58+85+87+89+90+91+96+97+98+99+100+101+
```

In order to determine which of the HC/LC interface residues do not interact with the antigen, this selection is modified so that it works with the unmodified 3se9 structure:

```
select 3se9_Fv_clean_interface, /3se9
//H/1+3+4+6+35+37+39+43+44+45+46+47+48+49+50+57+58+59+89+91+92+93+94+99+100
+100+100+100+100+100+100+101+102+103+104+105+106+108+ + /3se9
```

//L/31+32+33+34+35+36+38+41+42+43+44+45+46+47+48+49+50+51+52+53+55+56+57+58+85+87+89+90+91+96+97+98+99+100+101+

To ensure that only antigen-distal residues are considered, the paratope residues are defined using the following commands:

select paratope, byres(chain H+L within 5.5 of chain G)
color red, paratope

Next, we identified mutations in the HC/LC interface. To do this, we downloaded the nucleotide sequences for VRC-PG04 from Genbank (accession numbers **JN159466.1** – light chain, and **JN159464.1** – heavy chain). The nucleotide sequences are then entered into IMGT V-Quest (http://www.imgt.org/IMGT_vquest/vquest), and the mutations from germline in the HC/LC interface were identified manually using the resulting alignments. Additionally, we only selected for mutations whose side chains face the interface. The resulting mutations were formatted into a residue file or "resfile", which tells Rosetta which residue to place at any given position in a model. The mutations in the HC/LC interface of VRC-PG04 were reverted to their inferred germline residue using 3se9_germline.resfile:

NATAA
EX 1 EX 2
start
91 H PIKAA Y #F
32 L PIKAA Y #H
34 L PIKAA A #T
38 L PIKAA Q #K
43 L PIKAA A #P
44 L PIKAA R #K
49 L PIKAA Y #F
46 L PIKAA G #A
53 L PIKAA S #K

Construction of Rosetta models of interface-reverted CD4BS antibodies

To determine the effects that these naturally occurring somatic mutations had on the antibody bound conformation, we used Rosetta to construct ensembles of models for the HC/LC interface germline-reverted antibodies. Since we want to understand how antigen-distal mutations contribute to the bound conformation, the protocol was limited to a rigid-body threading. The following command was used to generate models for the mature antibody:

/path_to_rosetta/rosetta/main/source/bin/relax.default.linuxgccrelease -flip_HNQ -no_optH false -relax:constrain_relax_to_start_coords -score:weights talaris2013.wts -relax:ramp_constraints false -s 3se8_Fv_clean.pdb -nstruct 100 -scorefile 3se9.fasc -out:suffix "_mature"

Interface reverted models were generated using the following command:

/path_to_rosetta/rosetta/main/source/bin/relax.default.linuxgccrelease -flip_HNQ -no_optH false -relax:constrain_relax_to_start_coords -score:weights talaris2013.wts -relax:ramp_constraints false -s 3se9_Fv_clean.pdb -nstruct 100 -out:suffix "_revert" -relax:respect_resfile 1 -packing:resfile 3se9_germline.resfile -scorefile 3se9_reverted.fasc

At this point, both the mature and interface reverted models have been generated, but we still want to evaluate the effect that the mutations have on the interface. To ensure that we calculate across the same interface that we defined earlier, we're going to use the Interface Analyzer application again, but in the form of a mover. For the sake of this example, it isn't necessary, but it makes batch processing much easier. This method takes a "flags" or "options" file, an xml file, and the input pdb, all of which are available in the "inputs" folder on the github page. The interface energy, ΔΔG, was calculated for each model using the following commands:

/dors/meilerlab/apps/rosetta/rosetta_2015.12.57698/main/source/bin/rosetta_scripts.default.linuxgccrelease @iface_analyzer.flags -s *mature*pdb -parser:protocol iface_analyzer_VH_VL.xml -out:file:score_only -scorefile iface_3se9_mature.fasc

/dors/meilerlab/apps/rosetta/rosetta_2015.12.57698/main/source/bin/rosetta_scripts.default.linuxgccrelease @iface_analyzer.flags -s *revert*pdb -parser:protocol iface_analyzer_VH_VL.xml -out:file:score_only -scorefile iface_3se9_reverted.fasc

Next, the top 10 scoring models for each treatment were identified and their metrics collected using these commands:

cat iface_3se9_mature.fasc | sort –nk 2 | head -10 > top10_mature.fasc
cat iface_3se9_reverted.fasc| sort –nk 2 | head -10 > top10_reverted.fasc

The sixth column in these "top10" scorefiles represents the value for interface energy. The change in average interface energy, ΔΔΔG, is equal to Mature ΔΔG – Reverted ΔΔG. Negative values indicate a more favourable interface in the bound conformation.

HC/LC docking

Next, we determined how the mutations in the HC/LC interface affect HC/LC orientation by performing small perturbation docking. The docking step also takes an options file, and xml, and the starting model. In order to provide a direct comparison between mature and reverted interfaces, we restricted the docking protocol so that it does not alter the structural integrity of the domains; the minimization step employed an atom coordinate constraint, ensuring that the relax protocol itself would not skew angle measurements. We use this to generate 1000 models for each category, and analyse the top %5 for each, ranking by ΔΔG. Small perturbation docking was enacted using the following commands for the mature and reverted models:

Mature
/path_to_rosetta/rosetta/main/source/bin/rosetta_scripts.default.linuxgccrelease @docking.flags -s 3se9_Fv_clean.pdb -parser:protocol small_pert.xml -out:file:scorefile 3se9_dock_mature.fasc -nstruct 1000 -out:suffix "_mature"

Reverted

/path_to_rosetta/rosetta/main /source/bin/rosetta_scripts.default.linuxgccrelease @docking.flags -s 3se9_Fv_clean.pdb -resfile 3se9_germline.resfile -parser:protocol small_pert_revert.xml -out:file:scorefile 3se9_dock_revert.fasc -nstruct 1000 -out:suffix _dock_revert

The models are then ranked by ΔΔG, and the top %5 are used to evaluate change in orientation.

cat 3se9_dock_mature.fasc | sort –nk 10 | head -50 > 3se9Top50Mature.fasc
cat 3se9_dock_revert.fasc | sort –nk 10 | head -50 > 3se9Top50Revert.fasc

In order to determine how somatic mutations in the HC/LC interface affect heavy and light chain relative orientation, we used ABangle to calculate the relative HC/LC orientation for each of the top-scoring models. This software calculates six parameters by mapping two reference planes onto the Fv domains, drawing a distance vector between them, and measuring five angles – a torsion angle and four bend angles, between the two planes while using the distance vector as a pivot axis. Additionally, ABangle can take in a list of pdbs to evaluate in the form of .dat files. Generating the .dat file is accomplished through the following commands:

cat 3se9Top50*fasc | grep dock | awk ‘{print($NF”.pdb”}’ > 3se9Top50.dat

ABangle was used to calculate relative orientation through the following command:

ABangle –i 3se9top50.dat –usernumbered

The resulting angles are found in /path_to_ABangle/ABangleData/UserAngles.dat. The average values, standard deviations, and standard error were calculated for each type of model (mature and revertant) across the six ABangle parameters. The resulting values were used in the following equations to calculate the shift in average angle, and tightening of each distribution.

(1) ***Normalized shift*** =

$$\frac{1}{6}\sum\frac{\left|\bar{X}_{Reverted(HL,etc\ldots)} - \bar{X}_{mature(HL,etc\ldots)}\right|}{\sigma_{Reverted(HL,etc\ldots)} + \sigma_{mature(HL,etc\ldots)}}$$

where $\bar{X}_{Reverted(HL,etc\ldots)}$ is the mean ABangle value for any angle distribution generated by docking a reverted HC/LC interface, $\bar{X}_{mature(HL,etc\ldots)}$ is the corresponding mean ABangle value for the mature antibody, where $\sigma_{Reverted(HL,etc\ldots)}$ is the standard deviation for any given angle distribution generated by HC/LC docking at a reverted interface, and $\sigma_{mature(HL,etc\ldots)}$is the standard deviation for the corresponding mature antibody distribution. The Normalized Shift metric provides an estimate of how much the orientation distributions differ between any given mature antibody and

its reverted counterpart as a whole. Values greater than one suggest a shift in each category by an average of 1 standard deviation.

(2) ***Tightening*** =

$$\frac{\sigma_{Reverted(HL,etc\ldots)}}{\sigma_{mature(HL,etc\ldots)}}$$

The tightening equation generates a ratio of standard deviations. Values greater than 1 suggest that the mature antibody models embody a tighter angle distribution during HC/LC docking.

The standard error (SE) for the shift was calculated using error propagation rules for addition where:

shift in HL angle = $\left|\bar{X}_{Reverted(HL,etc\ldots)} - \bar{X}_{mature(HL,etc\ldots)}\right|$

$\boldsymbol{SE_{(HL)}}$ **=**

$$\sqrt{SE^{2}_{Reverted(HL)} + SE^{2}_{mature(HL)}}$$

normalized $\boldsymbol{SE_{(HL)}}$ =

$$\frac{\sqrt{SE^{2}_{Reverted(HL)} + SE^{2}_{mature(HL)}}}{\sigma_{Reverted(HL)} + \sigma_{mature(HL)}}$$

and $\boldsymbol{SE_{(Normalized\ Shift)}}$ =

$$\frac{1}{6}\sqrt{SE^{2}_{(HL)} + SE^{2}_{(HC1)\ldots}}$$