

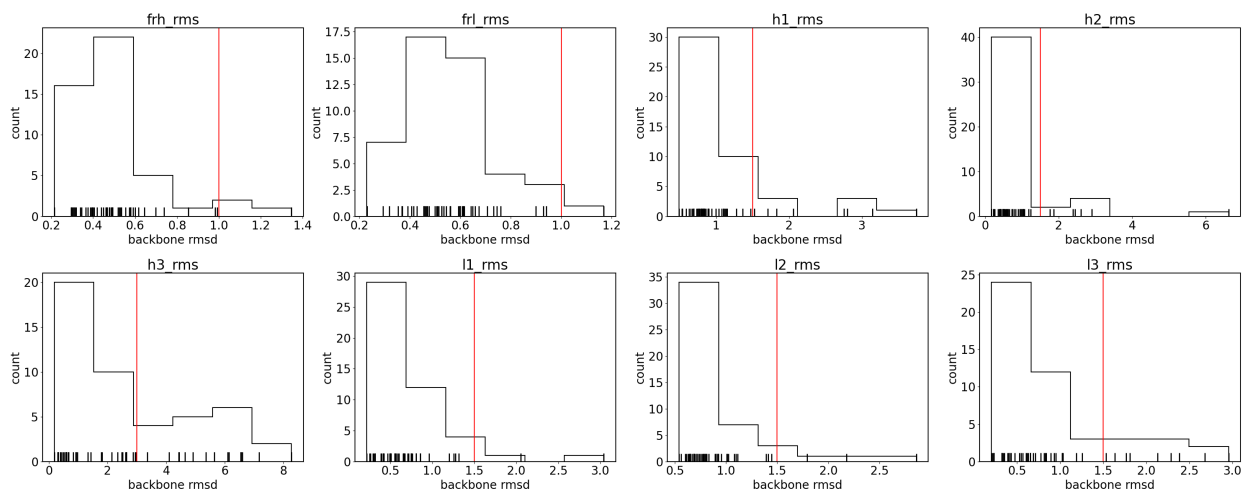
Scientific test: antibody_grafting

FAILURES

None

RESULTS

Briefly, the rmsd histograms plot the results for the benchmark (below) and are separated by structural region. Along the x-axis of each plot are notches showing the rmsd values for each model. The red line is the cutoff. Note that a certain number of models are expected to be above the cutoff, so having just one or a few past the line will not trigger a failure.



Individual PDBs off by more than X A

Listed below are, for each region, the PDBs missing the cutoff. Note these will only result in a test failure of some percentage greater than the cutoff fail. This list is for debugging purposes.

frh_rms > 1.0 A

3nps,

frl_rms > 1.0 A

2fb4,

h1_rms > 1.5 A

1oaq, 2r8s, 3go1, 3lmj, 3nps, 3p0y, 3umt, 4hpy,

h2_rms > 1.5 A

1oaq, 1seq, 2r8s, 3go1, 3lmj, 3nps, 3p0y,

h3_rms > 3.0 A

1gig, 1oaq, 1seq, 2d7t, 2fb4, 2vxv, 2ypv, 3giz, 3go1, 3liz, 3lmj, 3m8o, 3nps, 3umt, 4nzu,

l1_rms > 1.5 A

2r8s, 3giz,

l2_rms > 1.5 A

1gig, 1jfq, 1oaq,

l3_rms > 1.5 A

2e27, 2fbj, 2xwt, 3go1, 3lmj, 3nps, 3oz9, 3p0y, 4h20,

AUTHOR AND DATE

Jeliazko Jeliazkov (jeliazkov@jhu.edu), of Prof. Jeffrey J. Gray's lab, setup this test in April 2019.

PURPOSE OF THE TEST

This test evaluates the RMSDs of regions (FRH, FRL, FRH--FRL, CDR H1, CDR H2, CDR H3, CDR L1, CDR L2, and CDR L3) grafted by the antibody grafting application (apps/public/antibody/antibody.cc).

BENCHMARK DATASET

The test evaluates 48/49 antibodies, first described in (Marze, N. A., Lyskov, S. & Gray, J. J. Improved prediction of antibody VL--VH orientation. Protein Engineering, Design and Selection 29, 409--418 2016). We exclude 3mlr because it has an atypical CDR L3.

There are two components to the test: (1) Chothia-numbered structures downloaded from the antibody database SAbDab and truncated to the Fv region (~residues 1--112) and (2) FASTA sequence files derived from these structures. The sequences are used as protocol input and the structures are used for rms evaluation.

This set of antibodies was also used for CDR-H3 loop modeling evaluation in (Weitzner, B. D. & Gray, J. J. Accurate Structure Prediction of CDR H3 Loops Enabled by a Novel Structure-Based C-Terminal Constraint. The Journal of Immunology 198, 505--515 2016).

PROTOCOL

The antibody modeling protocol is described in (Weitzner, B. D. et al. Modeling and docking of antibody structures with Rosetta. Nature Protocols 12, 401--416 2017).

The original protocol comes from (Sivasubramanian, A., Sircar, A., Chaudhury, S., & Gray, J. J. Toward high-resolution homology modeling of antibody F v regions and application to antibody-antigen docking. Proteins 74, 497-514 2009).

Here we test only the grafting step, in which an input antibody sequence is decomposed into structural regions (FRs and CDRs) then homologs for those regions are selected (by BLAST) and combined into a *single model*. Grafting

is based on sequence identity and is deterministic, unless the sequences in the database change, so only a single model is produced per simulation. Grafting takes ~5 mins to run. This benchmark should take at most 200 CPU minutes.

Special note: in the full antibody modeling protocol, models are relaxed and we generate multiple models with different light-heavy chain orientations. Since we are not testing the full protocol here, there is no need to do this. We are only testing template selection and grafting.

PERFORMANCE METRICS

A single grafted model is produced per target antibody. The quality of a grafted antibody model is evaluated over four structural regions in each of two loops. The regions are a framework (FR) and three complementarity determining regions (CDRs). The metric of interest is backbone RMSD. For the FRs, this is evaluated after aligning on the regions. For the CDRs, this is calculated after FR alignment, so the CDR loops themselves are not aligned. The alignments are done with the `cdr_backbone_rmsds()` function in `protocols/antibody/metrics.hh/cc`.

For this benchmark, we ask what percentage of model regions are within X Angstrom of the native. FRs are structurally conserved across different antibodies. We expect little variation in this region, setting the threshold at 95% sub-Angstrom models. CDRs are a bit trickier. In the most recent blind evaluation of antibody grafting we modeled 46/55 (84%) non-H3 CDRs to sub-1.5-Angstrom accuracy. This value varies by CDR however, L3/H2 are the most erroneous whereas L1 & L2 rarely err, so we use a range of thresholds around there (L1 & L2 @ 90%, L3 @ 85%, H1 & H2 @ 80%). Finally, there is a special CDR loop (the H3) that is exceptionally diverse. It is rare to find structural homologs for this loops so we set the threshold at 65% sub-3 A models. Thresholds for the benchmark are summarized below:

FRs: 95% sub-Angstrom

non-H3 CDRs: 80--90% sub-1.5 A (depending on the CDR)

CDR H3: 60% sub-3 A

If the benchmark produces a lower percentage of sub-X-Angstrom models than described in the cutoffs, then the analysis scripts will signal "failure".

The latest assessment results are here (Weitzner, B. D., Kuroda, D., Marze, N., Xu, J. & Gray, J. J. Blind prediction performance of RosettaAntibody 3.0: Grafting, relaxation, kinematic loop modeling, and full CDR optimization. *Proteins: Structure, Function and Bioinformatics* 82, 1611--1623 2014). Although, these are for a small assessment.

For a full evaluation of the original protocol see: (Sivasubramanian, A., Sircar, A., Chaudhury, S., & Gray, J. J. Toward high-resolution homology modeling of antibody Fv regions and application to antibody-antigen docking. *Proteins* 74, 497-514 2009).

KEY RESULTS

Most models should have ~ 1 Angstrom RMSD in most regions. See Blind prediction performance of RosettaAntibody 3.0 above for details.

DEFINITIONS AND COMMENTS

PDB ID 3MLR was removed from the benchmark because of its peculiar L3, but could potentially be investigated further.

LIMITATIONS

I have not tested if relaxing improves/worsens results.