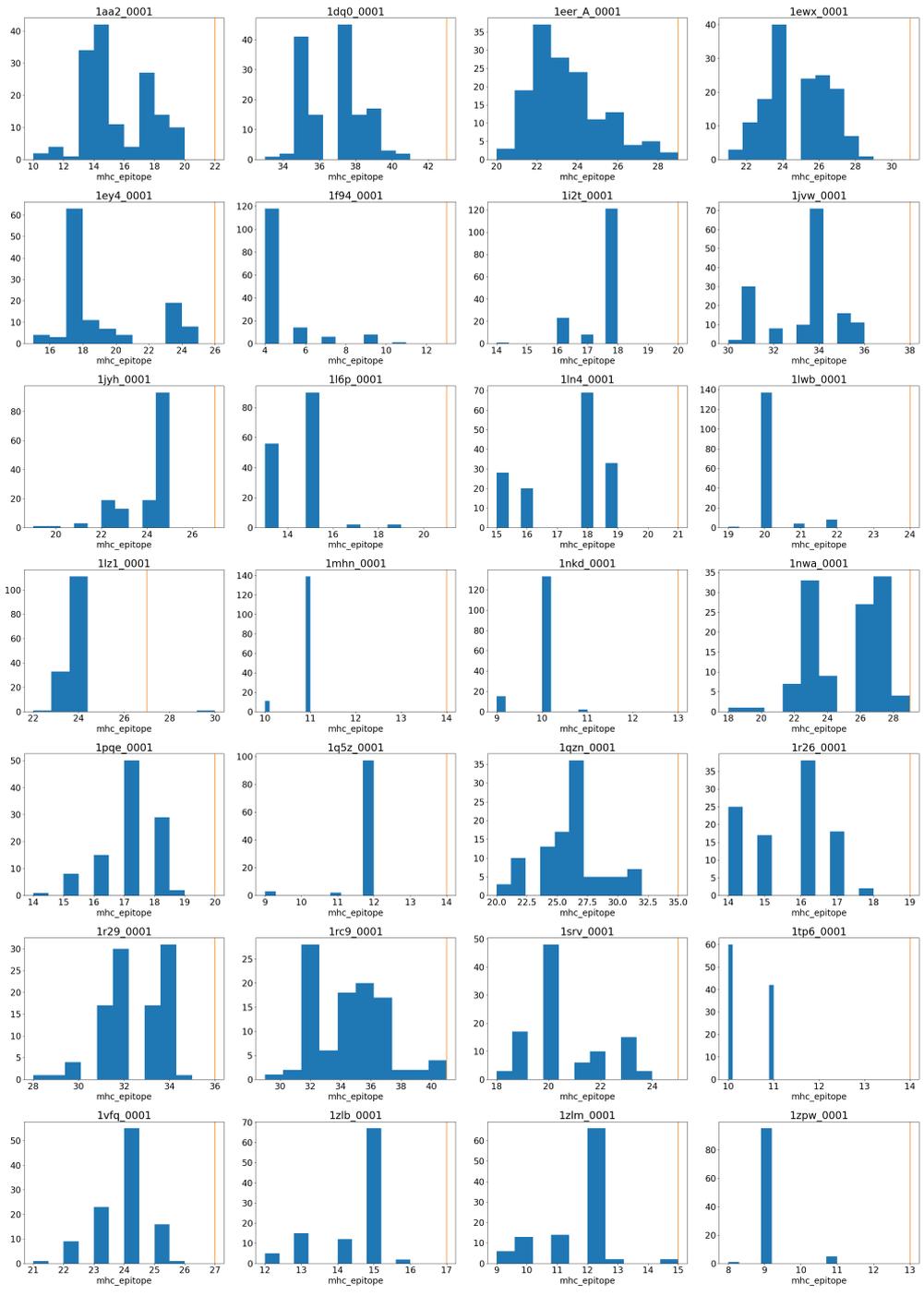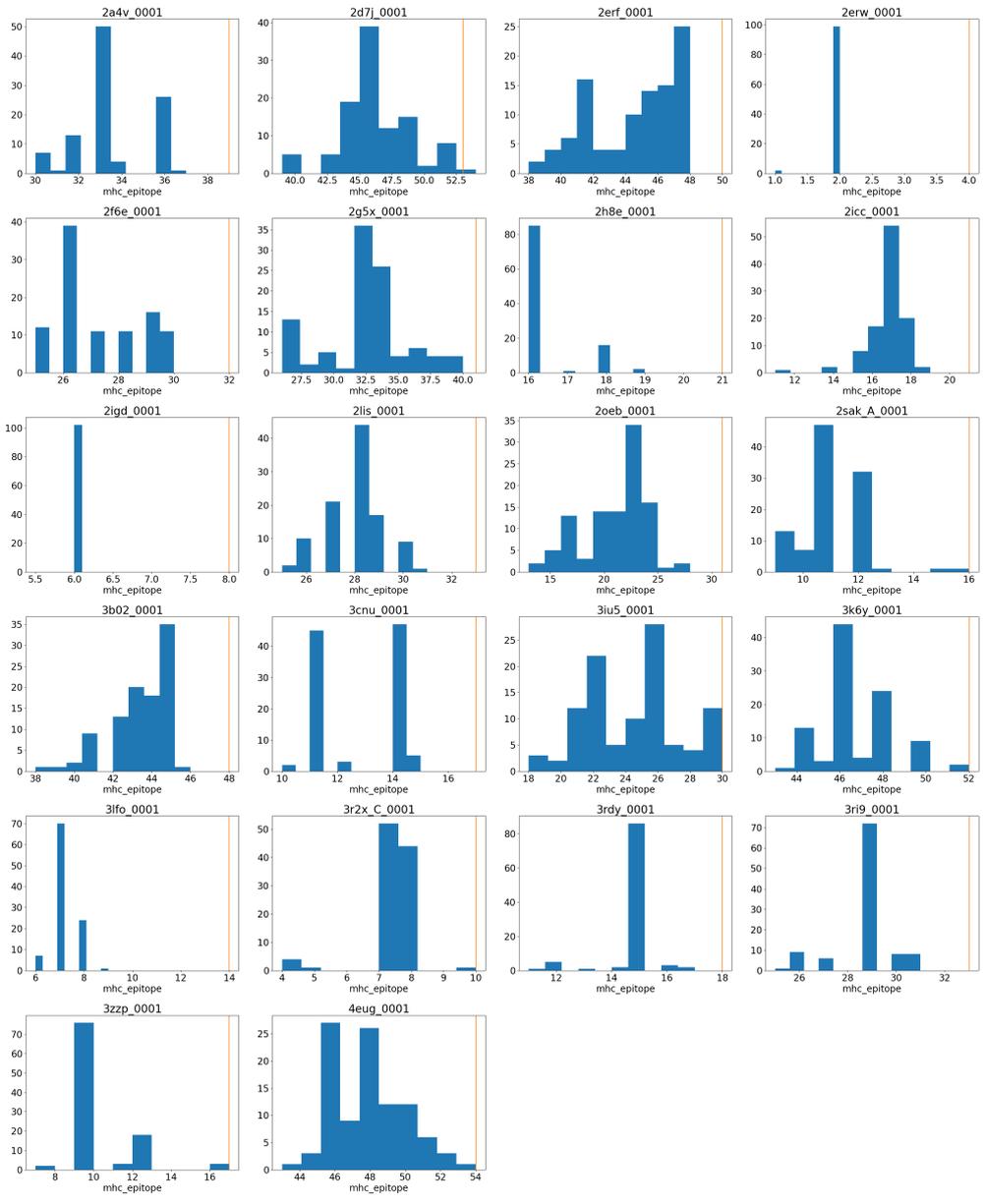# Scientific test: mhc_epitope_energy

FAILURES

None
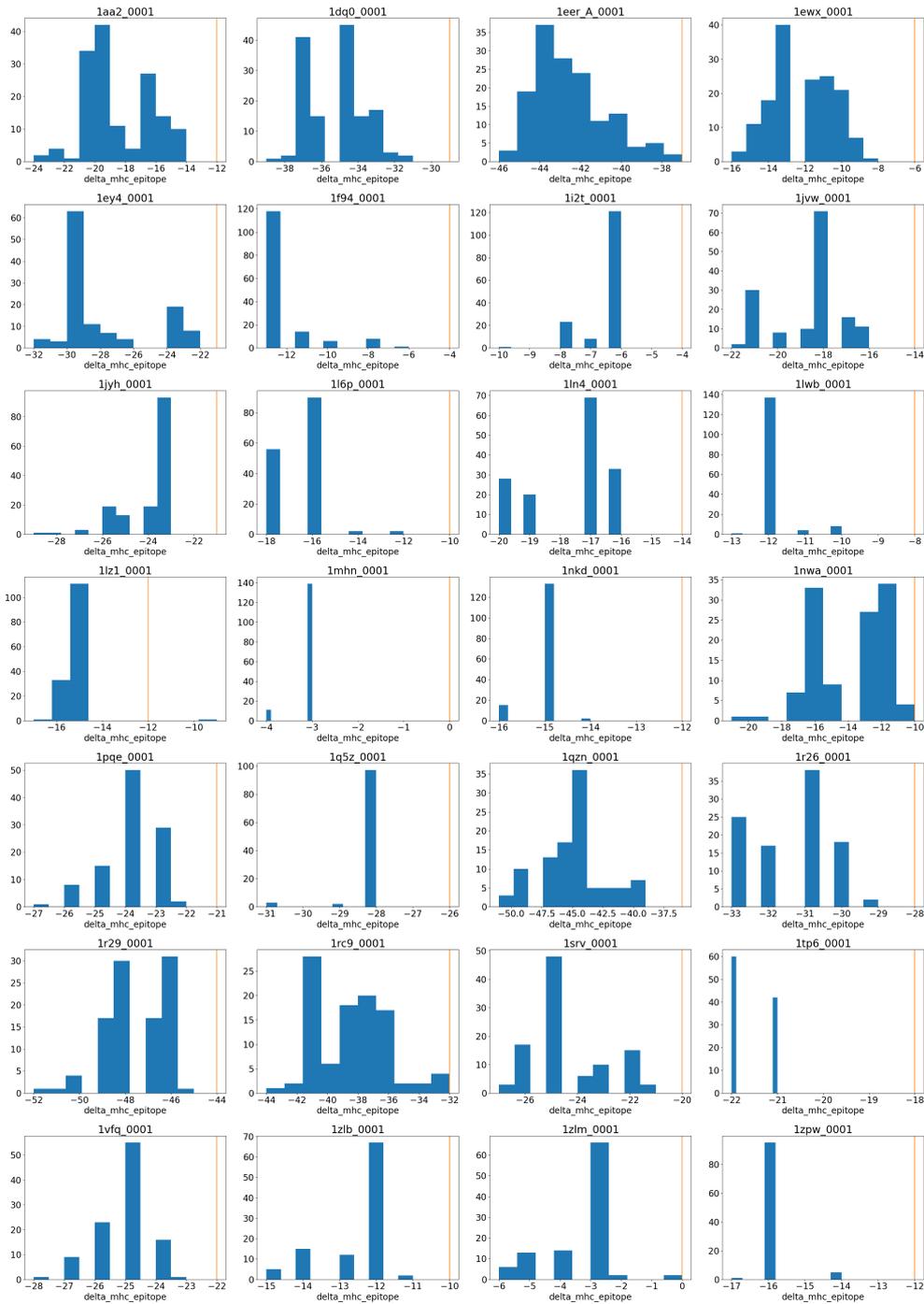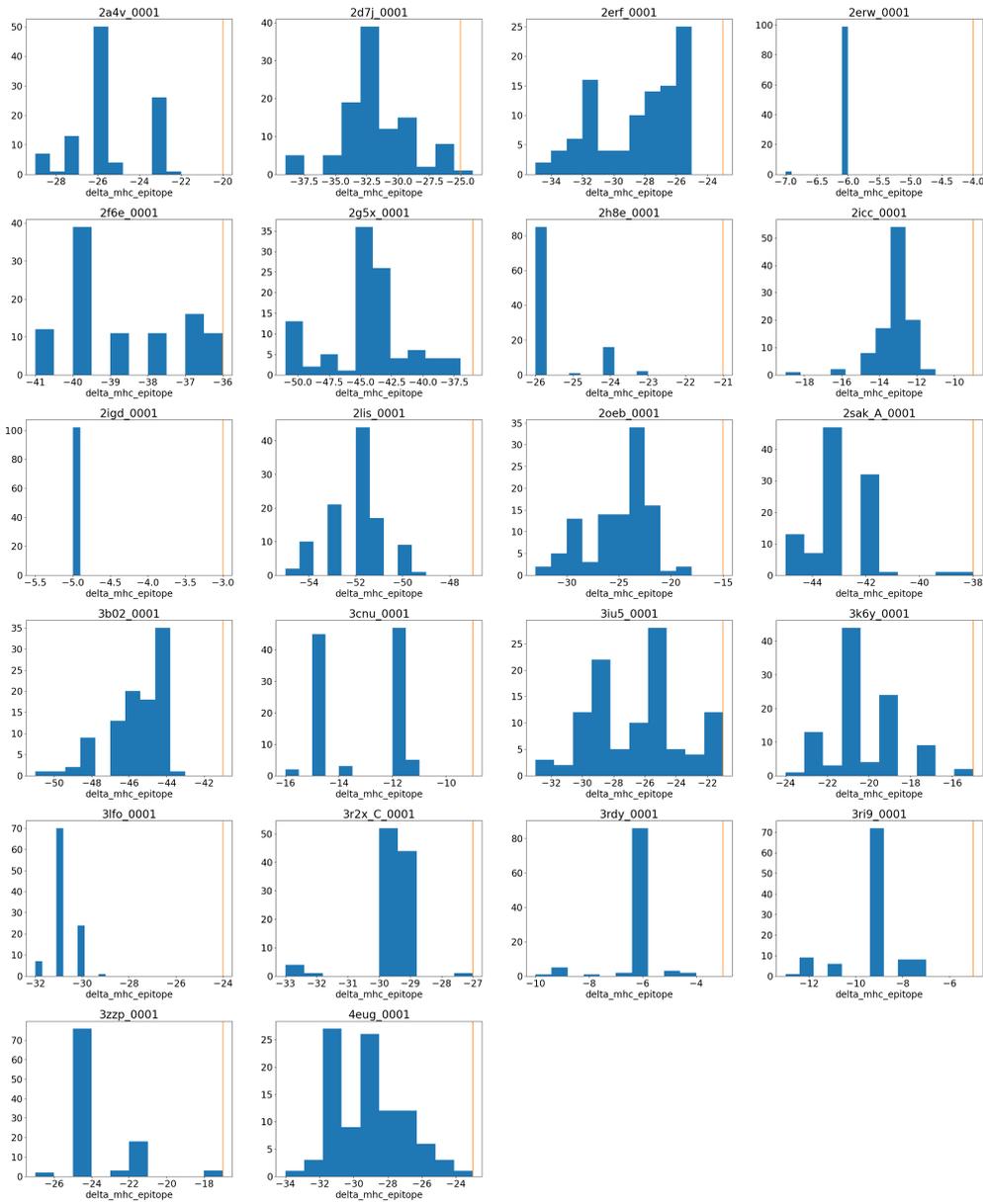
RESULTS

mhc_epitope

delta_mhc_epitope

base_total_score_vs_mhc_epitope
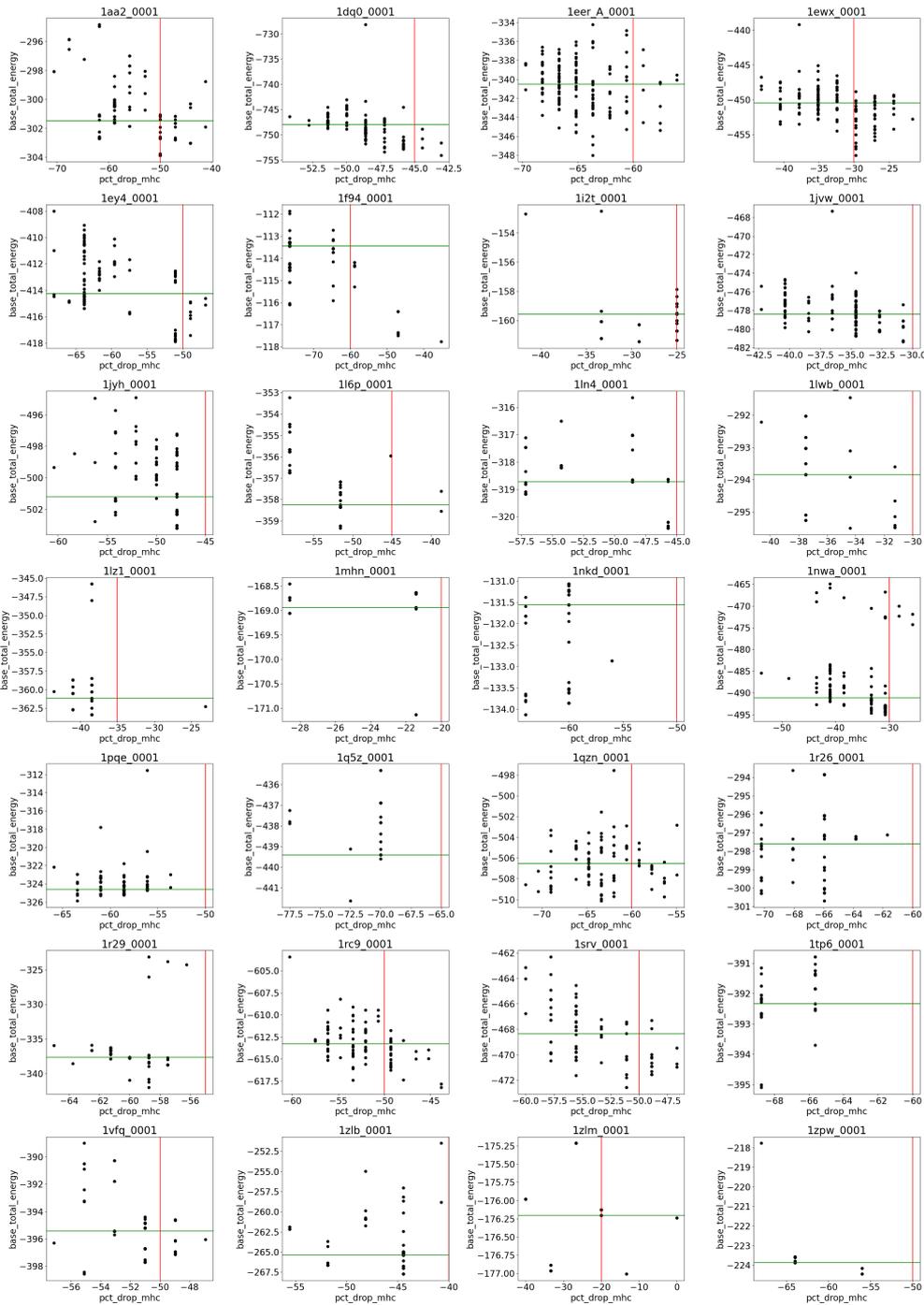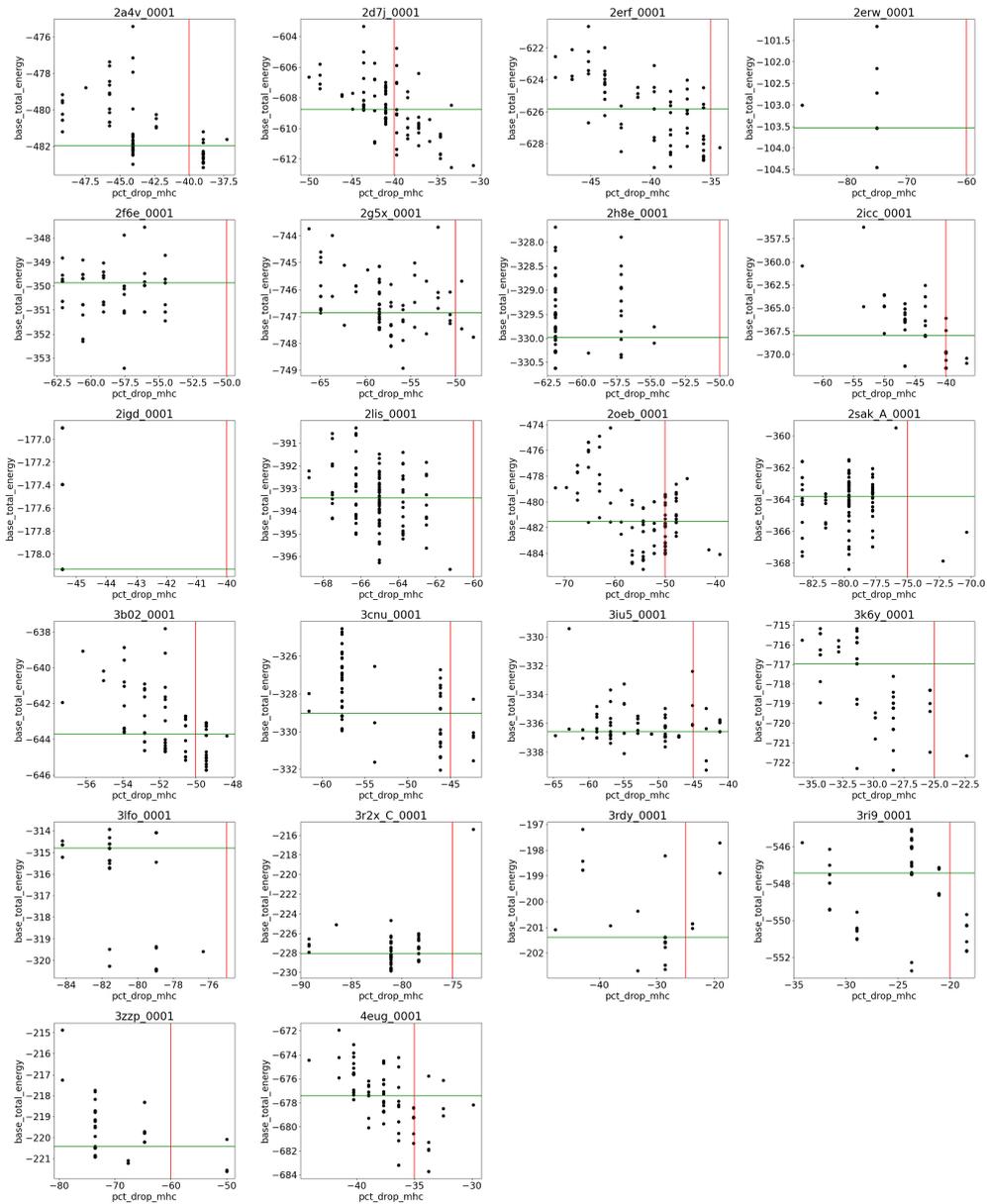
sequence_recovery_vs_mhc_epitope
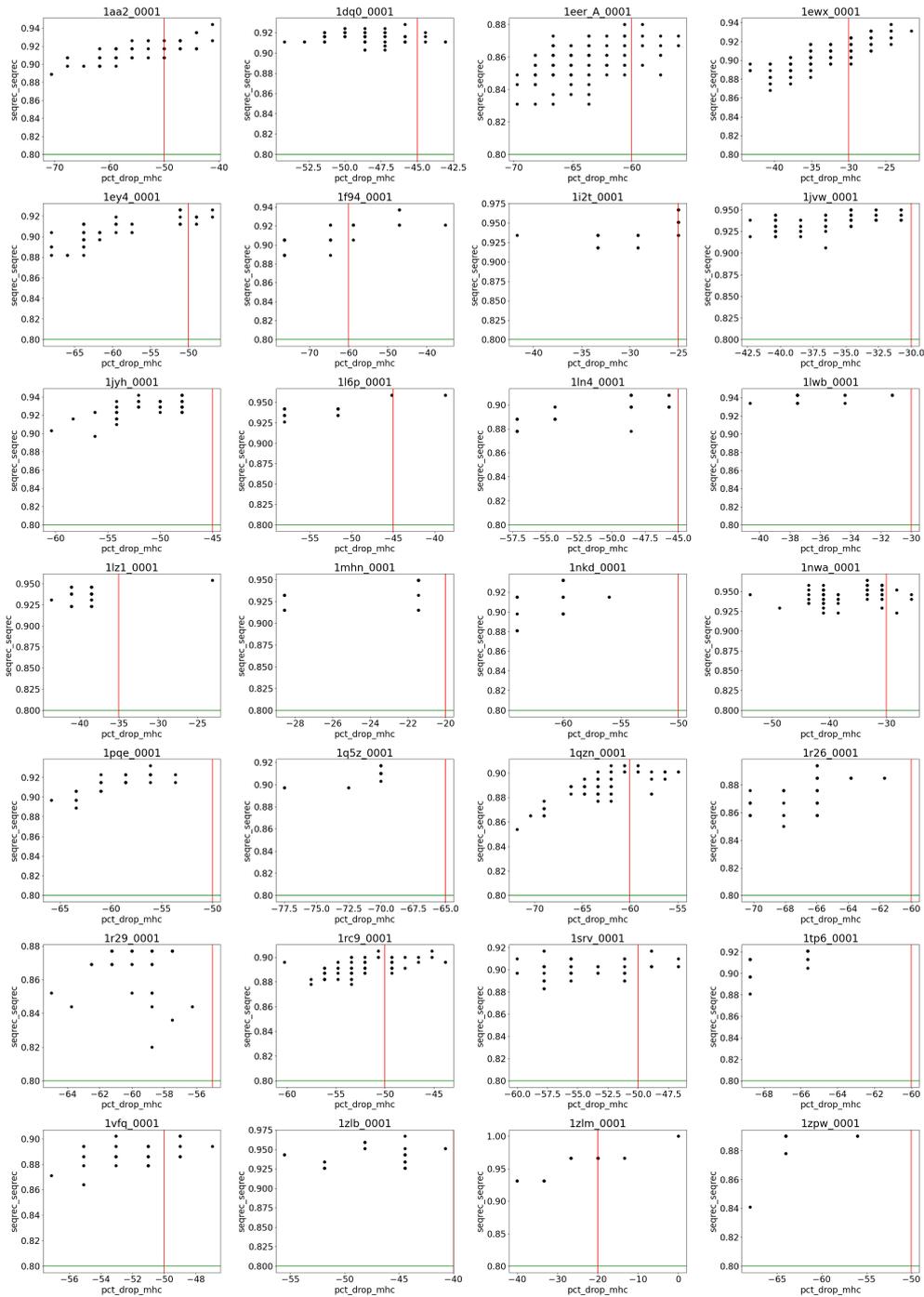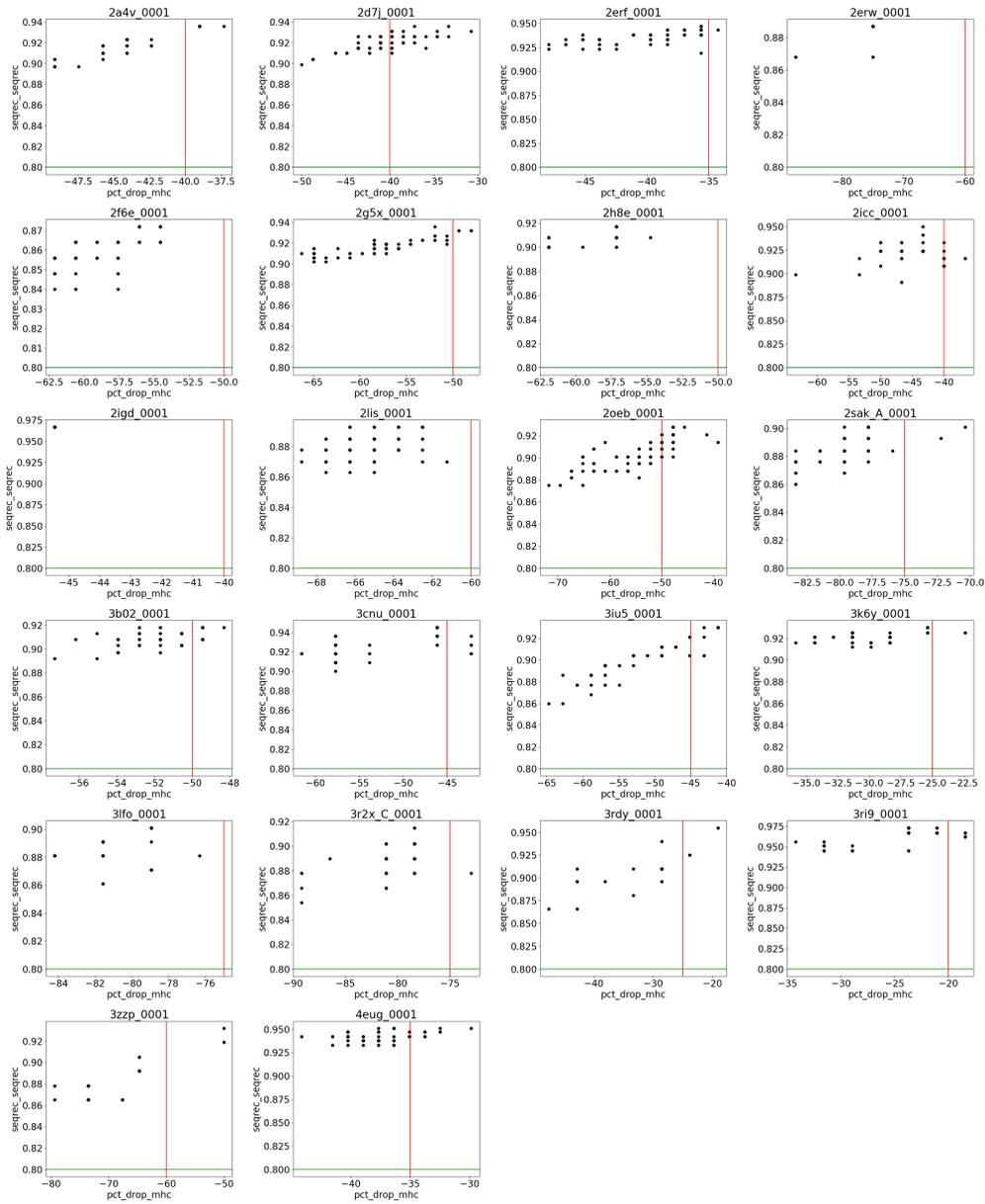
core_sequence_recovery_vs_mhc_epitope
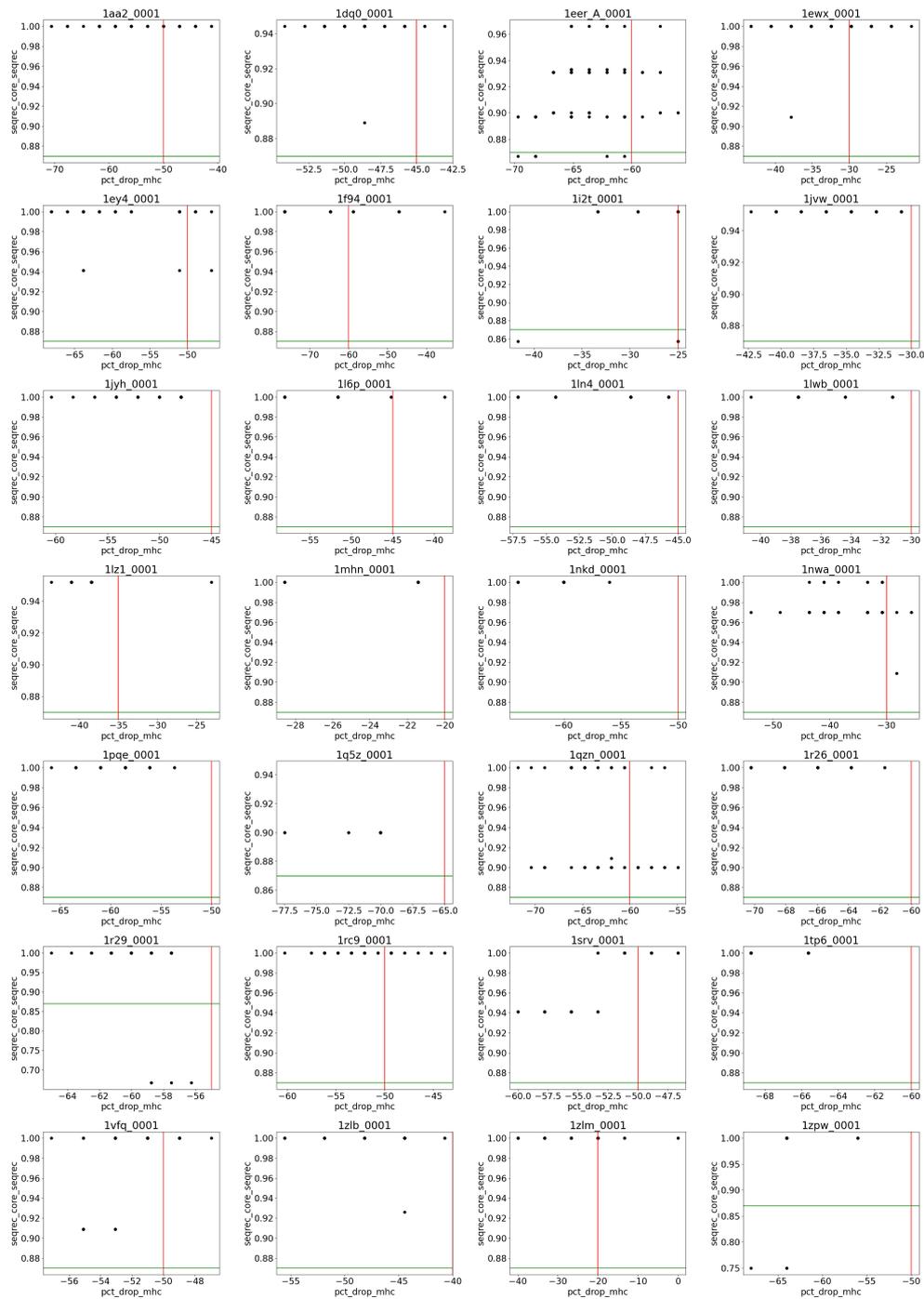
delta_packstat_vs_mhc_epitope

delta_buried_unsat_vs_mhc_epitope
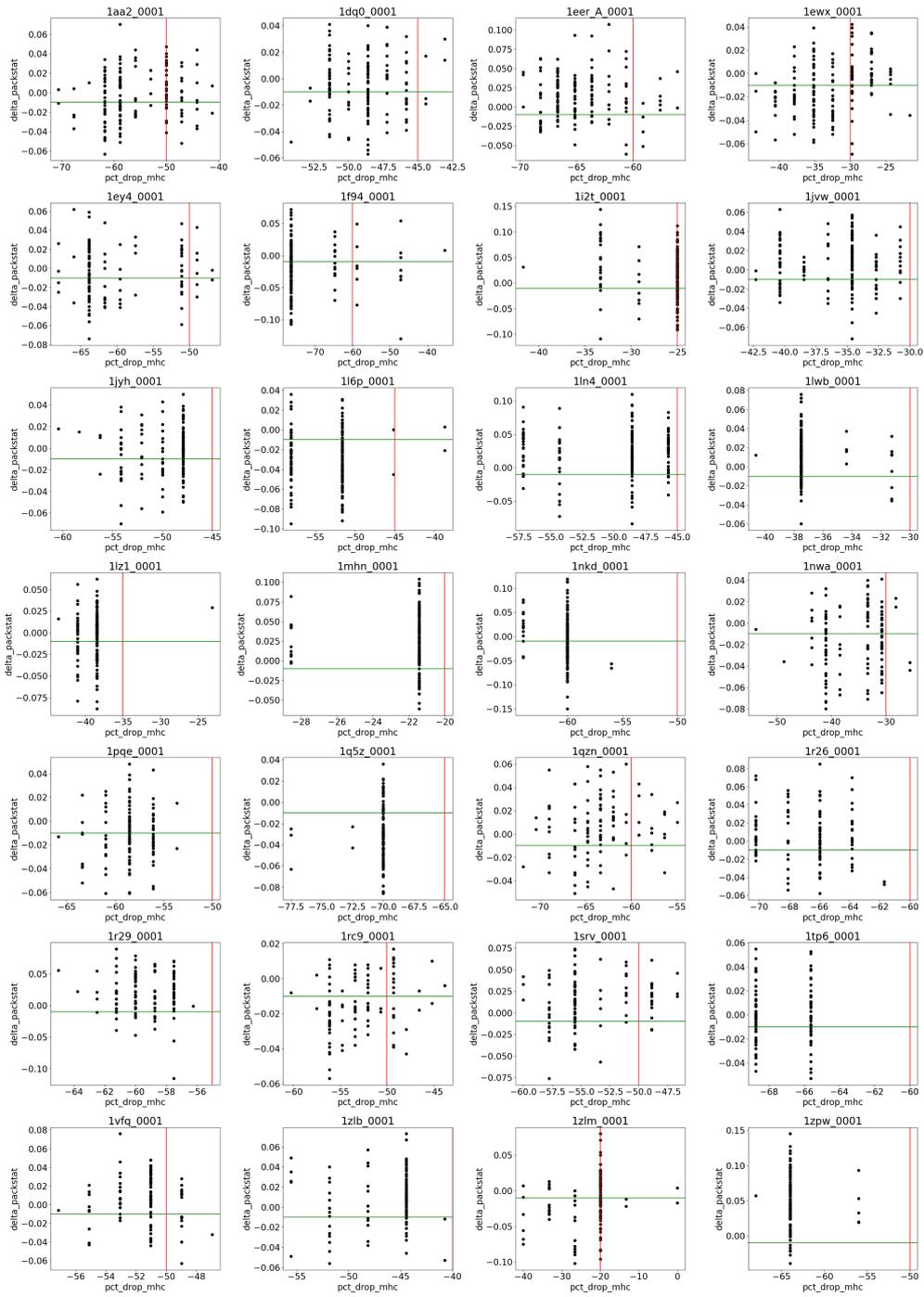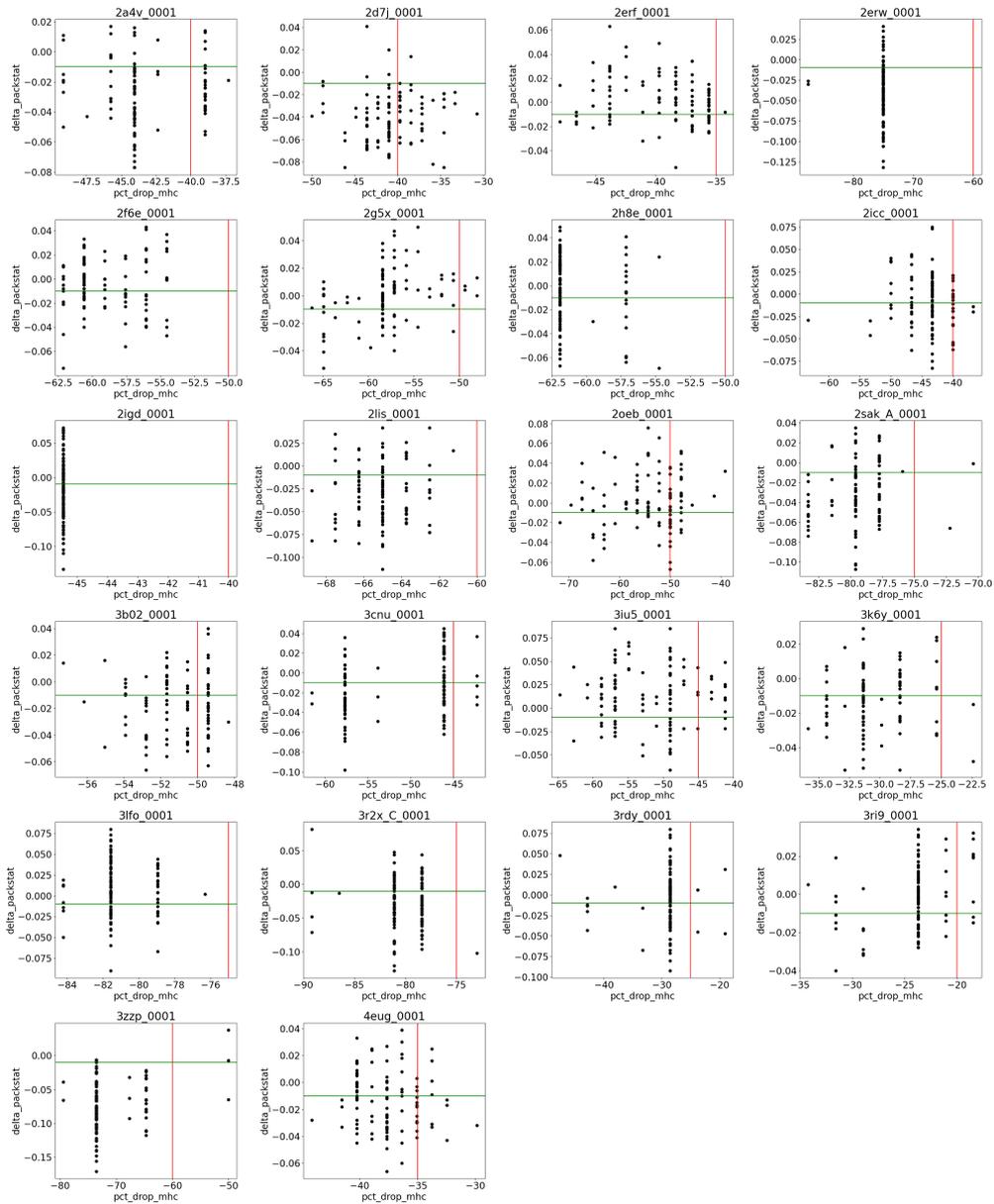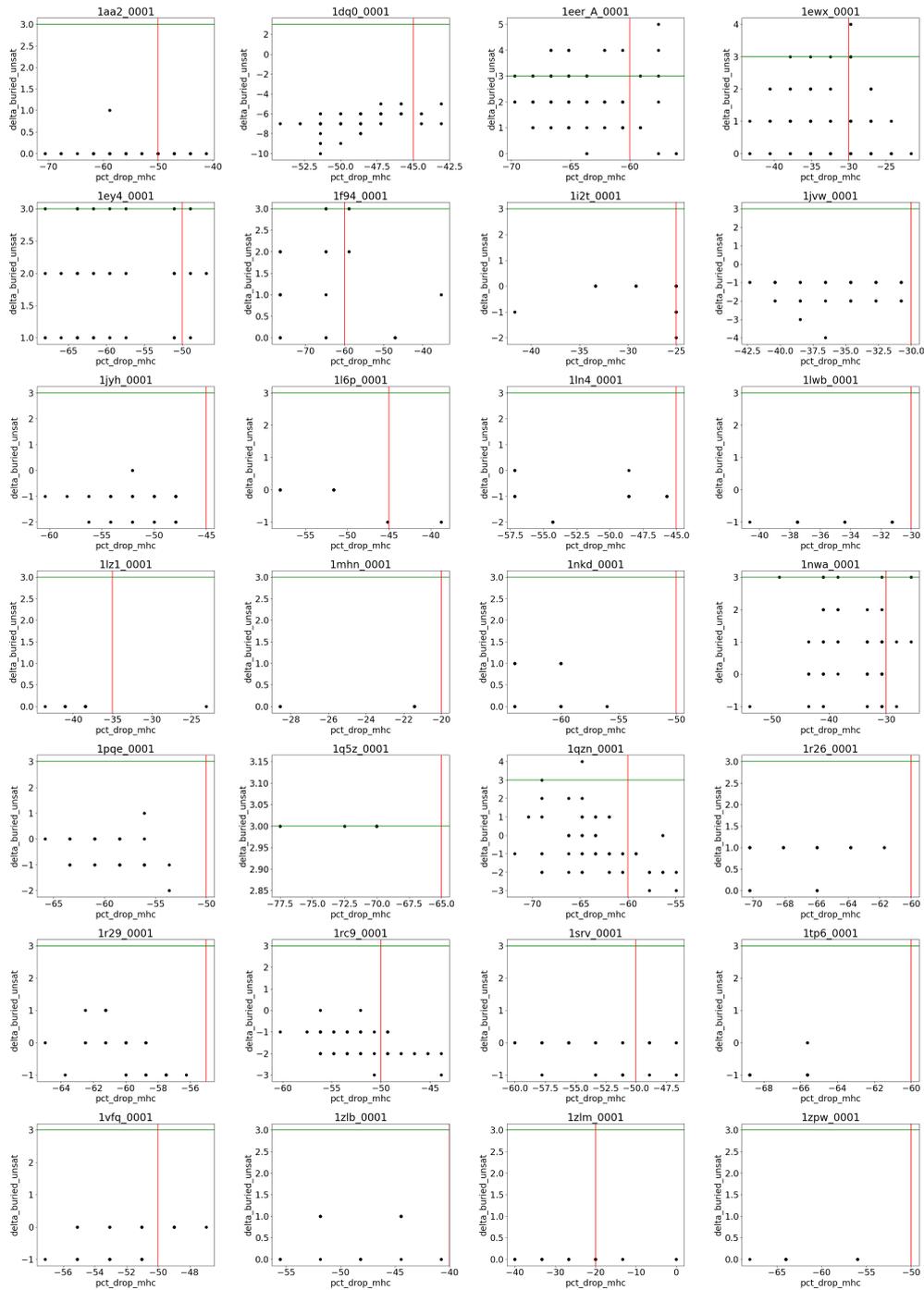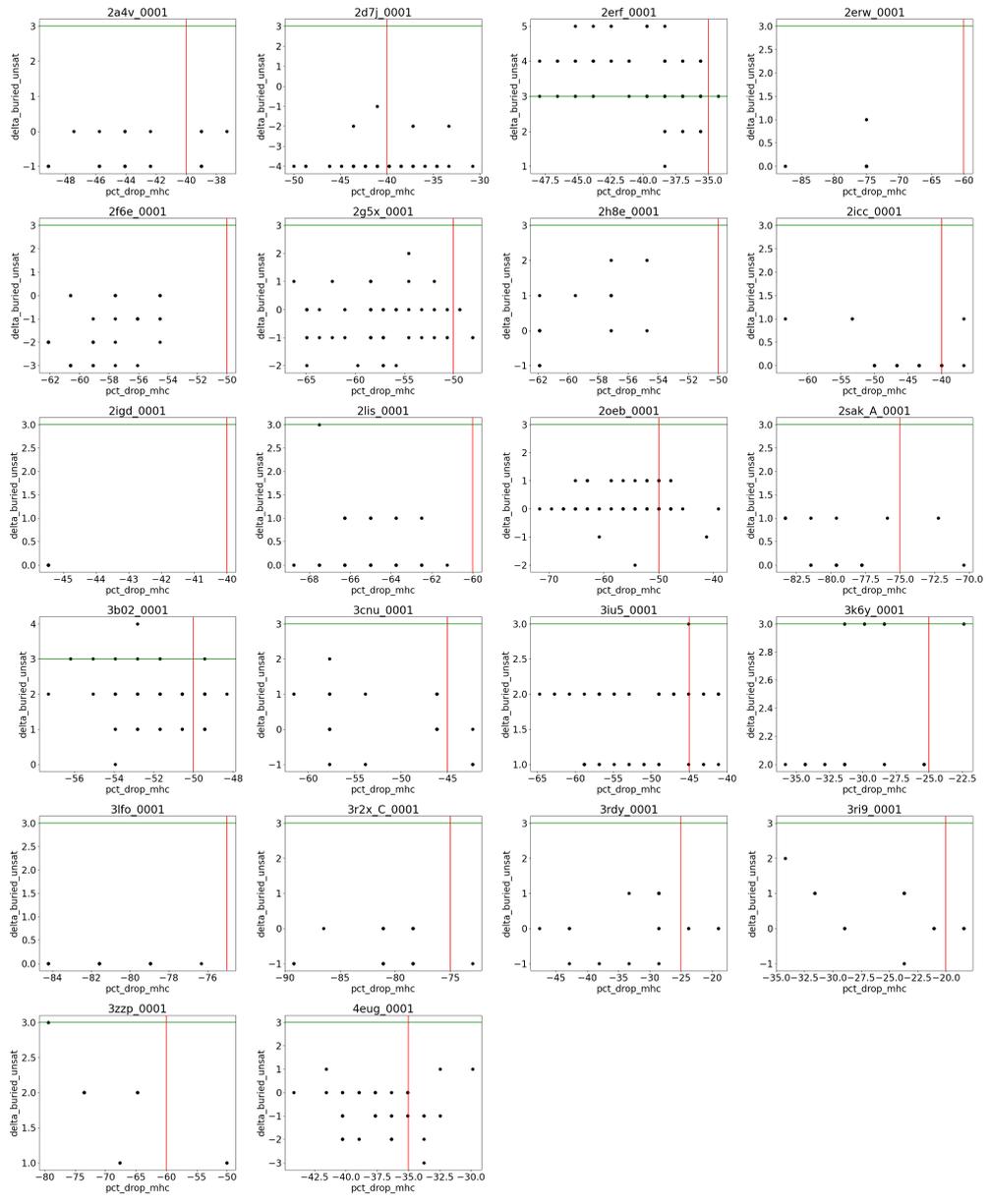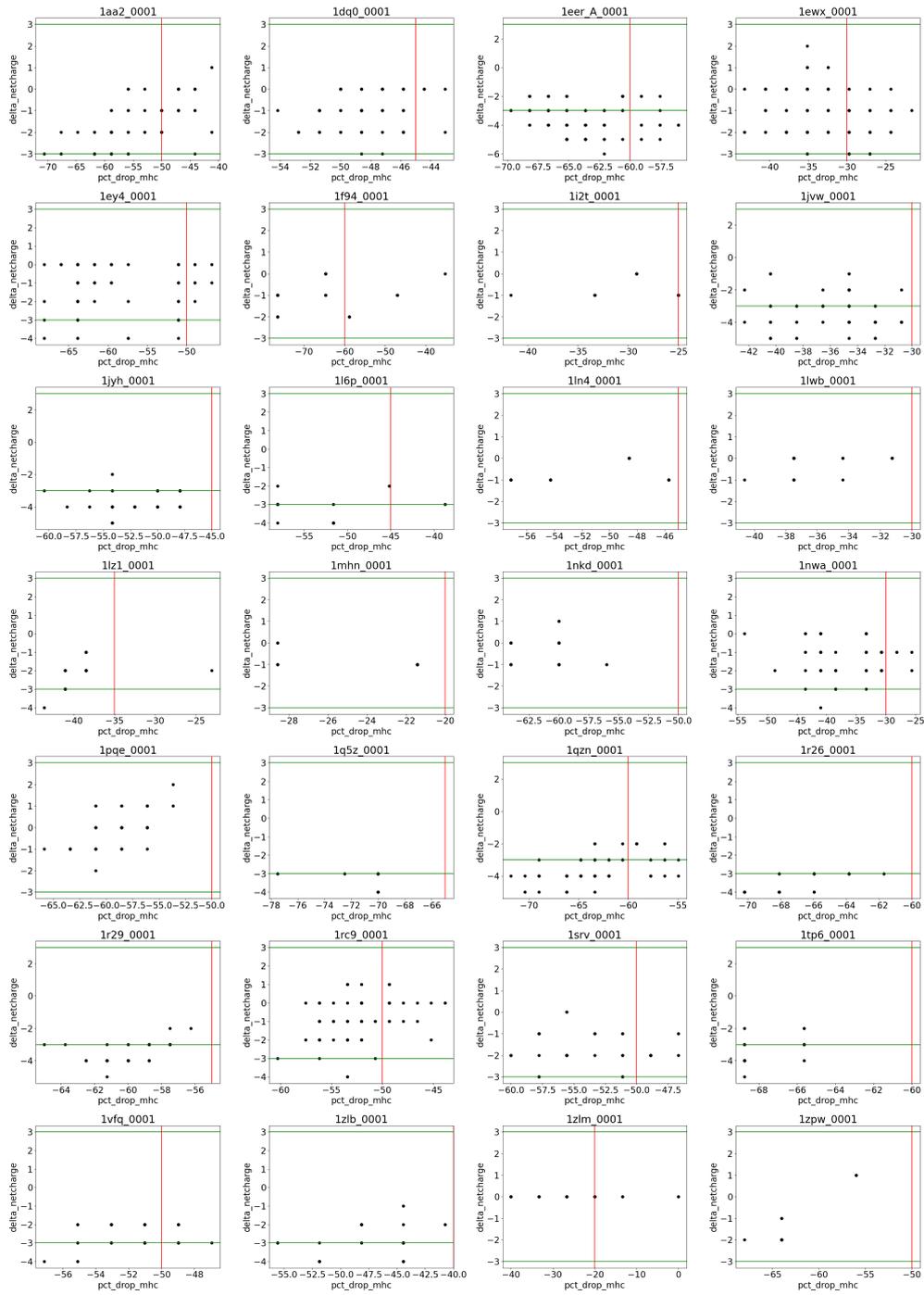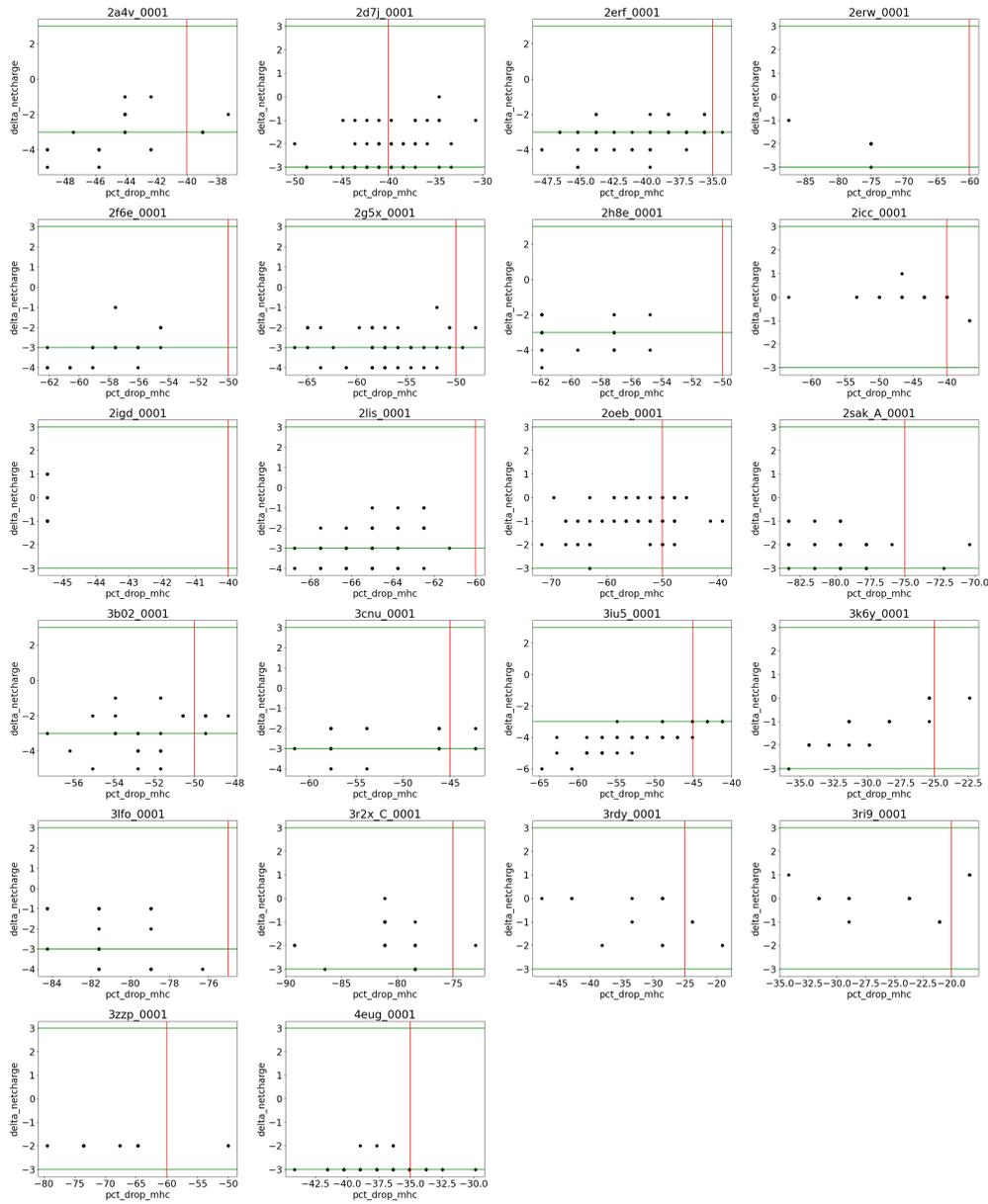
delta_netcharge_vs_mhc_epitope

AUTHOR

This test was set up by Brahm Yachnin (Sagar Khare's lab, Rutgers Universty), in collaboration with Chris Bailey-Kellogg (Dartmouth College).

brahm.yachnin@rutgers.edu

cbk@cs.dartmouth.edu

khare@chem.rutgers.edu

PURPOSE OF THE TEST

Protein de-immunization, involving the removal of T-cell epitopes that can trigger an immune response, is critical to the success of protein-based therapeutics/biologics. This test is designed to monitor Rosetta's ability to de-immunize proteins while maintaining good structural properties. In addition to watching for a decrease in mhc_epitope score, a ProPred-based metric for how immunogenic the protein is, we are monitoring five additional parameters that track the "goodness" of the resulting structure (see below).

BENCHMARK DATASET

The benchmark dataset includes 50 proteins, which were originally selected for a similar study (Choi et al. (2013) Structure-based Redesign of Proteins for Minimal T-cell Epitope Content. J. Comp. Chem. 34:879-891.), in which a different Rosetta scoreterm (described in the aforementioned paper, but never merged) was used for computational de-immunization. Our scientific test aims to mirror this study using the new (as of 2019) Rosetta implementation of packer-compatible ProPred-based de-immunization.

The following manuscript describes this benchmark:

Yachnin BJ, Mulligan VK, Khare SD, and Bailey-Kellogg C. (2021). MHCEpitopeEnergy, a flexible Rosetta-based biotherapeutic deimmunization platform. J Chem Inf Model 61(5):2368-2382. doi: 10.1021/acs.jcim.1c00056.

The input PDB files were downloaded from the PDB and relaxed using fast_relax with coordinate constraints using the ref2015 scorefunction. The resfiles were derived from PSI-BLAST PSSMs, generated using the online PSI-BLAST tools (version 2.8.1). The default PSI-BLAST settings were used, and three iterations of PSI-BLAST were performed to obtain the PSSM. Resfiles were generated using the mhc_gen_db.py tool located in tools/mhc_energy_tools using the PSSMs as input, as follows:

mhc_gen_db.py --propred --pssm PDB_NAME.pssm --pssm_thresh 1 --res_out PDB_NAME_thresh1.res --pdb PDB_NAME.pdb --firstres NUMBER_OF_FIRST_RESIDUE_IN_PDB

(PSSM thresholds of 2 and 3 were generated in the same way.)

The .comp files, used to setup the AAComposition score, were generated for each PDB file based on the number of native positive and negative charges for that protein.

PROTOCOL

This test designs (all residue fixed backbone design) and minimizes a set of proteins using the ref2015 scorefunction with the MHCEpitopeEnergy scoreterm (mhc_epitope) turned on to perform immune epitope prediction and elimination. The test uses the "base"/default ProPred configuration. In order to follow our recommended settings, FavorNative constraints are turned on, and a PSSM-based resfile is used to restrict design space to evolutionarily

probable residues. In addition, the AAComposition scoreterm is used to maintain the number of positive and negative charges in the protein, thereby preventing the explosion of positively and negatively charged residues which is typical in de-immunization protocols. The PSSM was generated using PSI-BLAST, and at each position, any residue type with a score of 1 or higher in the PSSM (plus the native identity) is allowed. A number of quality metrics are evaluated and compared to the native structure after design.

Note that in "debug" mode, the PSSM threshold is increased to 3, which results in a much quicker runtime because of the smaller allowed design space.

The publication of this protocol is currently (February 2020) in preparation.

The protocol takes approximately 2000 CPU hours in release mode, running the entire benchmark with an nstruct of 100.

PERFORMANCE METRICS

First, we assess whether the "degree" of de-immunization is maintained by comparing the distribution of mhc_epitope and delta_mhc_epitope scores to a cutoff determined for each target based on typical results. 90% of decoys should be lower than the preset cutoff. The cutoff (determined per target) is denoted by the ORANGE lines in the plots below.

In addition to mhc_epitope, which is used to measure "how deimmunized" the protein is, we looked at five orthogonal metrics to measure the "goodness" of the resulting designs. We are essentially asking if some of the more de-immunized designs also "good" as evaluated by these metrics. For a subtest to be considered a pass, a certain fraction of poses (determined for each subtest) must have a sufficient percent drop in mhc_epitope score (determined for each target, and indicated by the RED lines in the plots below) AND must ALSO meet the following thresholds (denoted by the GREEN lines in the plots below):

1. total_score (ref2015 without any contributions from constraints) must be in the top 50th percentile of the total_score values obtained during design of that target.

2. Sequence recovery (is the sequence native-like?) must be greater than 80%, and sequence recovery of "core" regions must be greater than 87%.

3. packstat (is the protein well-packed?) must be no more than 0.01 worse than native.

4. Buried unsatisfied hydrogen bonds (are more of these introduced as a result of de-immunization?) must increase by no more than 3.

5. Net charge (does de-immunization of the protein introduce extra charges?) must change by no more than 3 in either direction.

total_score and packstat are measurements of the "goodness" of the protein structure and packing. Sequence recovery is important as we want to

minimally disrupt the protein sequence in this design case, as the goal is to maintain structure and activity. As de-immunization tends to replace hydrophobic residues with polar and charged residues, buried unsatisfied hydrogen bonds is a metric used to identify introduction of unsatisfied polar residues in the protein core. Finally, because de-immunization tends to often replace residues with Asp or Glu, net charge is used to monitor if the sequence incorporates a large number of additional charged residues which may destabilize the packing or activity of the protein.

The test passes overall if all subtests pass. Note that while the subtests are classified as passes or fails in debug mode, these results are not meaningful and should be ignored. (The "overall" test will automatically pass in debug mode, as long as the test scripts complete successfully.)

KEY RESULTS

The aim of this test is to ensure that we can continue to get good performance in both protein de-immunization and orthogonal structure quality metrics. We do not have an experimentally validated result to compare against, nor an established computational test set to compare to. The key result should be that we are able to obtain similarly de-immunized proteins in structures that also show high quality metrics (as compared to the native). If future versions of Rosetta fail to obtain structures that are both de-immunized and good quality structures, this implies that the underlying protocol must be modified to restore this balanced behaviour.

The cutoffs that are selected are somewhat arbitrary, aiming to capture the current performance. If performance improves in the future, these cutoffs could be made to be more stringent to reflect this improvement.

In the future, major updates to Rosetta (e.g. a new scorefunction) should be modified within the rosetta_script used in this test to determine if this results in better or worse behaviour during de-immunization. If there is no change or results improve, the test should be permanently modified to make use of the new data.

DEFINITIONS AND COMMENTS

De-immunization, in this context, means the act of removal T-cell epitopes from a protein. With the ongoing rise of biologics in the pharmaceutical industry, as well as in academic research, the ability to de-immunize targets without sacrificing protein structure, function, or stability is an increasingly important field of research. To maximize results, it is critical to remove the strongest epitopes while minimally mutating the native protein.

LIMITATIONS

Because this benchmark is based entirely on computationally derived trends, we do not have the ability to show if we are generating real, active, de-immunized proteins. A case where we assess if Rosetta is able to generate structures that match experimentally validated cases would be a substantial

improvement, but would require extensive work on a large number of these targets.