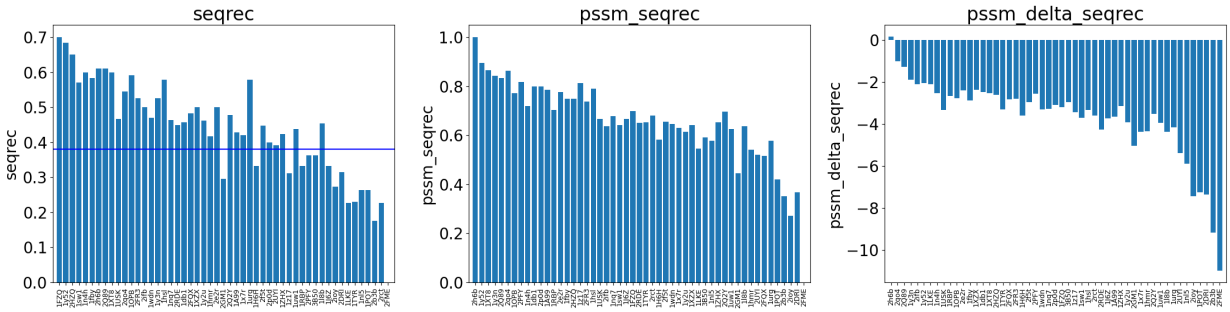
Scientific test: enzyme_design

FAILURES

(NONE)

RESULTS



Protein	pssm_delta_seqrec	pssm_seqrec	seqrec
AVERAGE	-3.660	0.655	0.438
1A99	-3.643	0.786	0.429
1FZQ	-3.200	0.700	0.700
1H6H	-3.583	0.583	0.333
1J6Z	-3.741	0.667	0.333
1LKE	-2.091	0.545	0.227
1OPB	-2.773	0.773	0.591
1POT	-7.263	0.421	0.263
1RBP	-2.667	0.704	0.333
1TYR	-3.308	0.654	0.231
1USK	-3.333	0.667	0.467
1XT8	-2.533	0.867	0.600
1XZX	-2.357	0.643	0.500
1ZHX	-3.154	0.654	0.423
1db1	-2.486	0.800	0.457
1fby	-2.875	0.750	0.583
1hmr	-4.333	0.542	0.417
1hsl	-3.316	0.789	0.579
1l8b	-4.364	0.636	0.455
1n4h	-2.520	0.720	0.600
1nl5	-5.895	0.579	0.263
1nq7	-3.286	0.679	0.464
1sw1	-3.714	0.643	0.571
1urg	-4.158	0.579	0.579

1uw1	-3.938	0.625	0.438
1wdn	-3.294	0.647	0.471
1x7r	-4.368	0.632	0.421
1y2u	-3.923	0.615	0.462
1y3n	-1.895	0.842	0.526
1y52	-2.053	0.895	0.684
1z17	-3.438	0.812	0.312
2DRI	-7.368	0.368	0.316
2FME	-11.000	0.000	0.000
2FQX	-2.828	0.517	0.483
2FR3	-2.789	0.737	0.526
2GM1	-5.037	0.444	0.296
2HZQ	-2.600	0.750	0.650
2PFY	-2.545	0.818	0.364
2Q2Y	-3.522	0.696	0.478
2Q89	-1.278	0.833	0.611
2RDE	-4.250	0.650	0.450
2UYI	-5.391	0.522	0.391
2b3b	-9.176	0.353	0.176
2e2r	-2.389	0.778	0.500
2f5t	-2.966	0.655	0.448
2h6b	0.167	1.000	0.611
2ifb	-2.091	0.636	0.500
2ioy	-7.455	0.273	0.273
2p0d	-3.100	0.800	0.400
2qo4	-1.000	0.864	0.545
2rct	-3.591	0.682	0.227
3B50	-2.955	0.591	0.364

AUTHOR AND DATE

Adapted for the current benchmarking framework by Rocco Moretti (rmorettiase@gmail.com; Meiler Lab), Sep 2018

PURPOSE OF THE TEST

This benchmark tests how well the enzyme design code is able to recapitulate native-like sequences when run over a set of cocrystal structures of small-molecule binding proteins with their native substrates.

BENCHMARK DATASET

There are 50 proteins in this set, chosen for being a high-quality structures of proteins binding to their native substrates. This benchmark set (and the basic protocol) is partly described in Nivon et al. (2014) "Automating human intuition for protein design." The input PDBs are from the previous benchmark tests, so their provenance is not 100% clear, but I believe that they have been downloaded from the RCSBV, minimally cleaned, and subjected to the all-atom constrainted relax protocol of Nivon et al. (2013) (probably under the score12/enzdes scorefunction).

PROTOCOL

The protocol follows more-or-less that of Nivon et al. (2014), updated for RosettaScripts XML. Briefly, the residues surrounding the ligand (design within 6 Ang (8 if pointed toward ligand) and repack within 10 (12) Ang) are subjected to 2 cycles of softpack/hardmin followed by 1 cycle of hardpack/hardmin. The protein-ligand interactions are upweighted by 1.8-fold for those residues being designed.

One big change from the publication is that the scorefunction being used is the (currently default) REF2015, rather than the older scorefunction used in the paper/in previous benchmarks. (It's the intention that the score function be updated based on whatever the current default is.)

Currently we are only running one output structure for each input, which results in the test taking somewhere around 25-50 CPU hours.

PERFORMANCE METRICS

The original test only looked at percent sequence recovery at designed positions; whether or not the design recaputulated the identical residue type as the input. ("seqrec") The concept is that, as the structures are native binders, their current sequence should be close to optimal for binding the ligand.

This reimplementation added two new metrics, based on matching the design result to a PSSM of the input protein. This PSSM was generated with the 2.6.0 version of psiblast, using the BLAST nr database from 21May2014 (yes, both the psiblast and nr database were old when this was done in late 2018) with the following command:

psiblast -query 1A99.fasta -db /path/to/db.21May2014/nr -out_pssm 1A99.chk -out_ascii_pssm 1A99.pssm -save_pssm_after_last_round

The first PSSM-based metric ("pssm_seqrec") is a percent recovery metric, but instead of attempting to match the input sequence exactly, it counts as a "success" matching any amino acid which has a favorable score in the PSSM. This metric is described in DeLuca et al. 2011. This is a 0-100% normalized scale which better allows for modest changes (e.g. S->T) which are permitted evolutionarily.

The second PSSM-based metric ("pssm_delta_seqrec") looks at the per-residue change in the PSSM score compared to the "native" input sequence. (Where more positive is a "better" design.) This attempts to capture the magnitude of the mutational change.

The benchmark set is summarized by averaging the scores for each protein. (This average is on a per-protein basis, and is not weighted by the number of positions mutated.)

The current tested cutoffs is based only on the match/fail seqrec metric, and is taken directly from the previous iteration of this benchmark. This value is somewhat arbitrary, and was set based on what didn't give overly noisy results when the benchmark was being run.

KEY RESULTS

I am unaware of any objective level one could compare the results to. Generally you'd be limited to a comparative performance based on prior results.

There's currently no analysis of outliers or per-protein performance, over and above anything discussed in Nivon et al. (2014).

DEFINITIONS AND COMMENTS

Comparing different runs and different scorefunctions, you're probably looking for something which is consistenty higher.

There's a bit of noise in the runs, so you may need to compare several runs to get better sense of how different scorefunctions compare.

LIMITATIONS

While the dataset attempted to be comprehensive (all structures which fit the criteria) when it was produced, the quality/breadth of structures these days may be better.

The preparation of the input structures may be another issue. The input structures may have been minimized/relax with an older version of the scorefunction. Updating the structure preparation may improve benchmark performance.

The run-to-run variablility is rather high, which might indicate that the benchmark could be improved by running several output structures for each input, rather than just the current one.

The exact protocol being used may not be optimal for ligand binding design, and a different packing/minimization scheme might improve performance.

Finally, keep in mind the intrinsic limitations of sequence-recovery based protocols. While low values are likely bad, you wouldn't necessarily expect the metric to saturate near a "perfect" score, as native proteins are optimized for more than just ligand-binding affinity, so even family PSSM profiles may not match the ideal design that Rosetta would be aiming for.