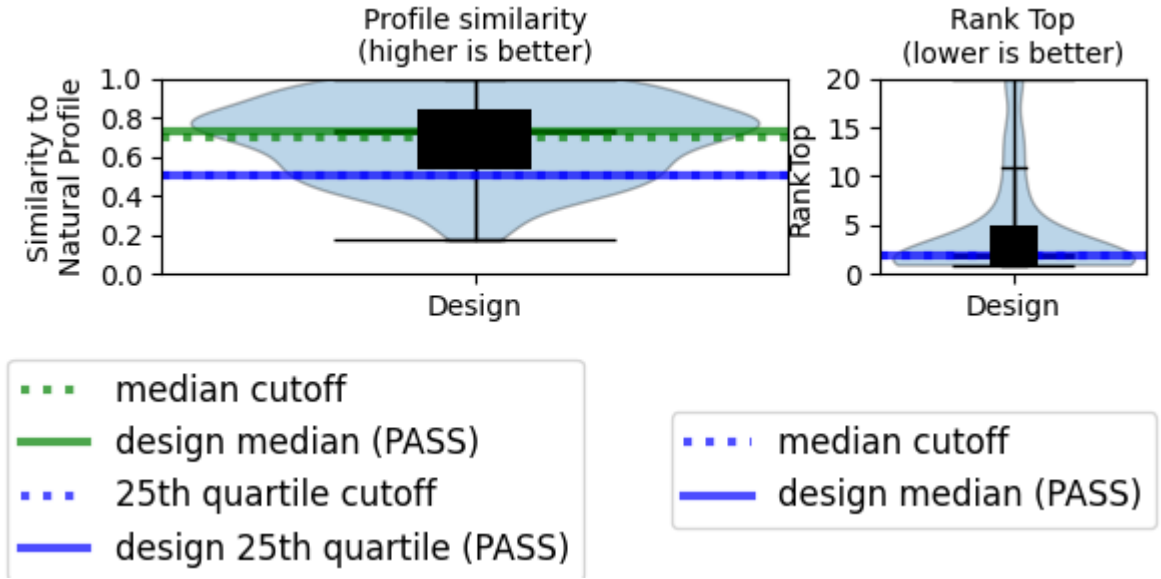
Scientific test: cofactor_binding_sites

FAILURE:

False
The test passed! You can learn more about CoupledMoves at https://www.rosettacommons.org/docs/latest/coupled-moves

RESULTS



AUTHOR AND DATE

Author: Amanda Loshbaugh (aloshbau@gmail.com)

PI: Tanja Kortemme (kortemme@cgl.ucsf.edu)

Date test first added to Scientific Benchmark: February 2019

PURPOSE OF THE TEST

This benchmark describes the ability of the Rosetta protocol CoupledMoves to recapitulate sequence profiles of naturally evolved proteins. The benchmark set consists of evolutionary sequence alignments for seven naturally occurring protein families that each bind a specific cofactor. Using sequence profiles has the advantage of clearly identifying consensus binding positions.

BENCHMARK DATASET

The cofactor binding sites dataset is comprised of seven separate protein families, with the conserved small molecule cofactor binding sites of each family. This dataset is described in detail elsewhere.[3]

In brief, each protein family satisfies three criteria: 1) there exists "at least one representative crystal structure bound to the cognate ligand to use as input for design, 2) the protein family has a large number of diverse sequences such that the binding site is not completely conserved, and 3) all members of the protein family are capable of binding the cognate ligand using the same ligand binding site. Ligand binding positions are defined as any amino acid position with a side-chain heavy atom within 6Å of any heavy atom in the co-factor ligand. Natural sequences of these binding sites were obtained using the protein family alignment from Pfam and filtered to remove all redundant sequences.[3]

This dataset that does not involve experimental characterization of binding affinity. Because natural proteins are under many different selection pressures, the affinity of natural sites is usually just good enough rather than as tight as possible.

PROTOCOL

Coupled Moves is a flexible backbone design method meant to be used for designing small-molecule binding sites, protein-protein interfaces, and protein-peptide binding sites. It handles ligands and waters. It was originally developed for designing enzyme active sites.

The CoupledMoves protocol is described at https://www.rosettacommons.org/docs/wiki/coupled-moves.

A publication describing the Coupled Moves is (Loshbaugh, A. L. & Kortemme, T. Comparison of Rosetta flexible-backbone computational protein design methods on binding interactions. Proteins Struct. Funct. Bioinforma. 88, 206â€“226 (2020))

The test runs for about 170 CPU hours.

PERFORMANCE METRICS

Tolerated sequence space, for each position in a given protein, describes the diversity of amino acids compatible with the protein's function,[2] and is calculated from the known natural sequence profiles. We quantify Rosetta's ability to recapitulate the natural sequence profiles at the defined ligand binding positions by two metrics, position profile similarity (PPS) and rank top (RT). Metrics are calculated per position.

PPS measures the similarity of the distribution of amino acid frequencies between two datasets, the known profile (from natural sequence alignment) and the profile generated by Rosetta design.

RT measures the rank, in the design profile, of the amino acid most frequently observed in the known profile.

When benchmarked on this dataset in [1], CoupledMoves achieved PPS with median ~0.72 on this dataset, and 25th percentile ~0.56. Accordingly,

thresholds of 0.7 (green) and 0.5 (blue) are chosen for the median and 25th percentile in this test.

KEY RESULTS

This benchmark describes the ability of the Rosetta protocol CoupledMoves to recapitulate sequence profiles of naturally evolved proteins.

LIMITATIONS

Sequence profiles in the cofactor dataset result from natural evolution, rather than experimental screening. Natural evolution includes additional selection pressures beyond affinity (function, stability, kinetics), the full consequences of which are difficult to categorize and quantify, meaning that the sequence profiles for natural binding site positions may be influenced by factors beyond those modeled by Rosetta. However, using evolutionary sequence profiles of cofactor-binding protein families has the advantage of clearly identifying consensus binding positions.

More data, in the form of more natural ligand binding protein families, would improve this benchmark.

REFERENCES

[1] Loshbaugh, A. L. & Kortemme, T. Comparison of Rosetta flexible-backbone computational protein design methods on binding interactions. Proteins Struct. Funct. Bioinforma. 88, 206â€“226 (2020)

[2] Humphris, E.L. and T. Kortemme, Prediction of protein-protein interface sequence diversity using flexible backbone computational protein design. Structure, 2008. 16(12): p. 1777-88.

[3] Ollikainen, N., de Jong, R. M., and Kortemme, T. Coupling protein side-chain and backbone flexibility improves the re-design of protein-ligand specificity. PLoS Comput Biol, 2015.