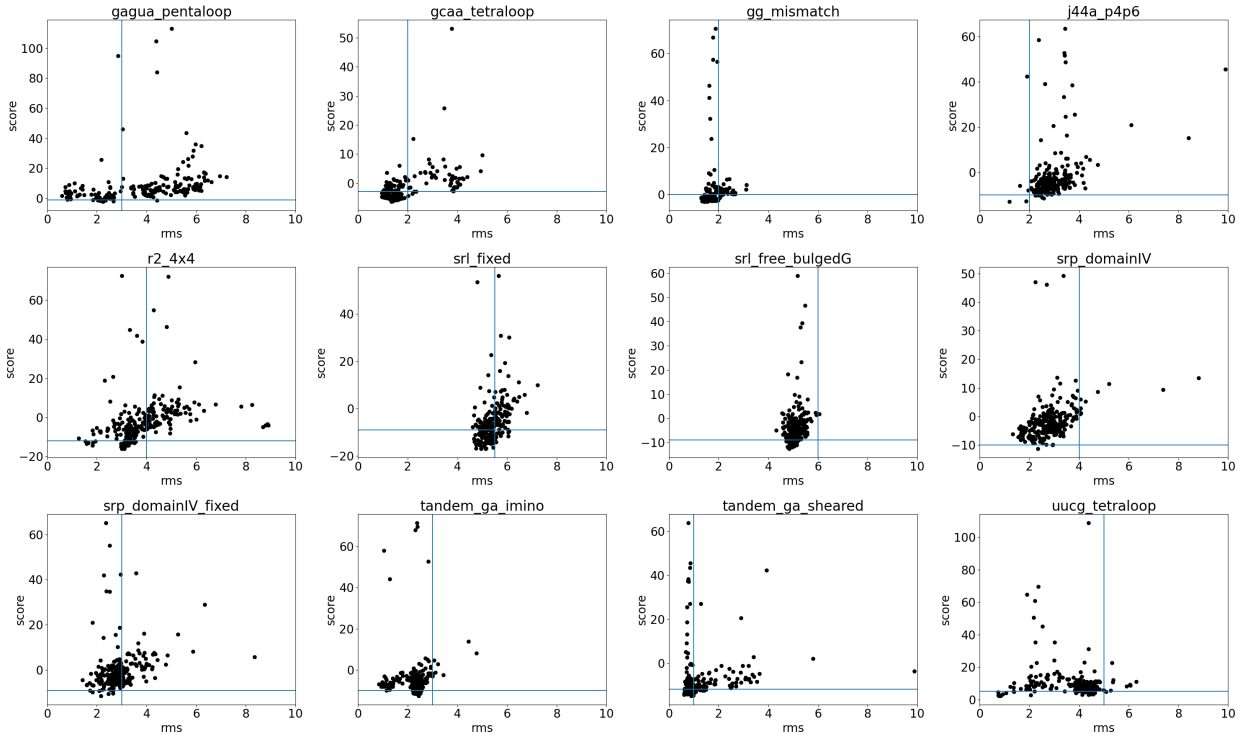
Scientific test: rna_denovo_favorites

FAILURES

None

RESULTS



AUTHOR AND DATE

Andrew Watkins (andy.watkins2@gmail.com); Rhiju Das (October 2018)

PURPOSE OF THE TEST

This test assesses the stability of the FARFAR algorithm's performance. FARFAR stands for Fragment Assembly of RNA with Full-Atom Refinement; it is a mirror of the protein 'ab initio' application but it is also well equipped for homology modeling using fragments of input structure.

BENCHMARK DATASET

The benchmark set contains the 12 loop modeling problems from favorites.txt, a benchmark first established in Watkins et al., Science Advances 2018. The input files are ideal A-form RNA helices and fasta files; the FARFAR algorithm takes these input pieces of PDB structure and attempts to predict the remaining residues defined by the FASTA file; 'correctness' is defined as similarity to portions of crystal structures.

PROTOCOL

See (Das, R., Karanicolas, J. & Baker, D. Atomic accuracy in predicting and designing noncanonical RNA structure. Nat. Methods 7, 291â€“294 (2010)) for a description of the FARFAR protocol. The benchmark takes 6 hours on the test server (~120 CPU-hours).

PERFORMANCE METRICS

Benchmark performance is based on the minimum RMSD sampled and minimum energy sampled; each case has a separate threshold. Since sampling is limited for each test, cutoffs are determined conservatively through repeated runs. A superior measure, at least in theory, might be to compute pnear and compare it to some threshold value.

KEY RESULTS

Because of the high-energy outliers endemic to the fragment assembly method, real "funnel-like" energy surfaces won't be visible for this test. That said, the lowest energy overall should lie to the left of the RMSD threshold and below the scoring threshold for each test; stochastic failures where a low-energy model or two appear slightly past the RMSD threshold should be unsurprising, but a failure to return sufficiently low-energy models at all should be viewed with suspicion.

DEFINITIONS AND COMMENTS

n/a

LIMITATIONS

The benchmark could be expanded to include favorites2.txt, challenges.txt, followups.txt, and more of the benchmarking challenges already developed for stepwise Monte Carlo. These files are located in the publicly available Github repo DasLab/rna_benchmark. Furthermore, the protocol could also target the RNA-Puzzles dataset (the equivalent of CASP), where we have pretty predictable performance characteristics as long as we can generate a few thousand models. More CPU power would permit better sampling and therefore more precise cutoffs.