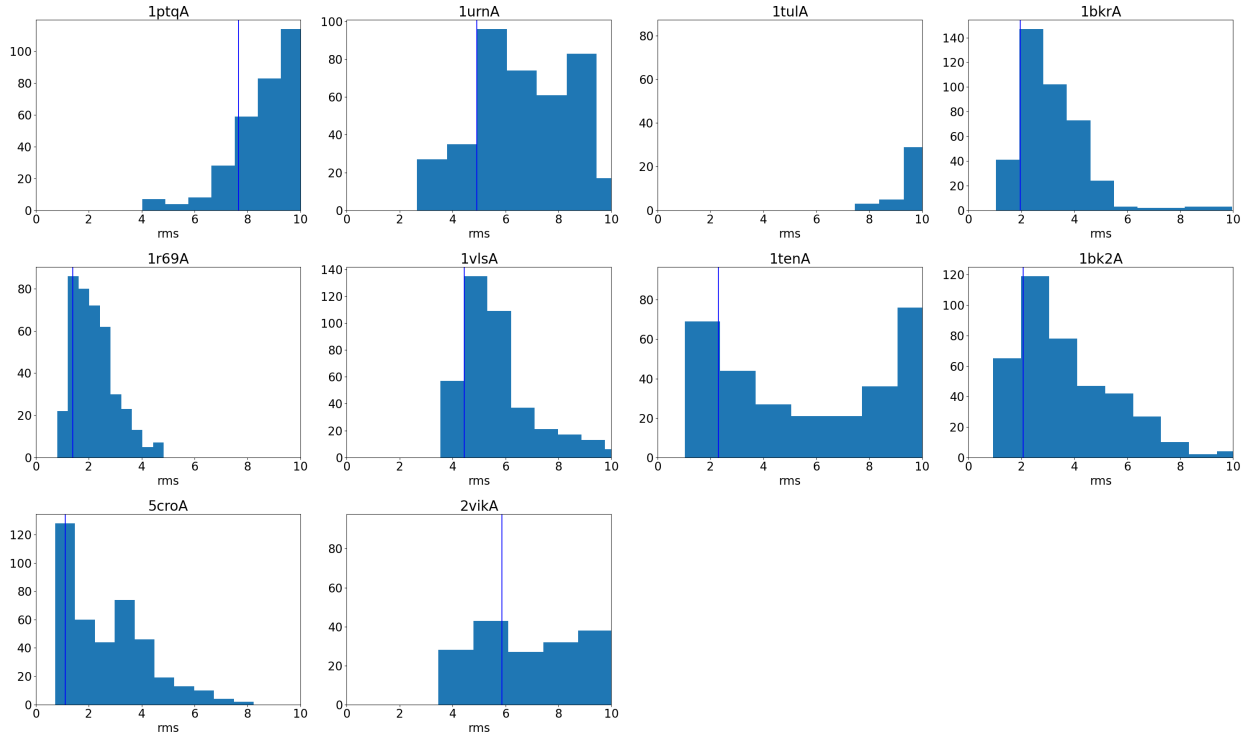
Scientific test: fragments_picking

FAILURES

None

RESULTS



AUTHOR AND DATE

The test was set up by Justyna Krys (juchxd@gmail.com) in Dominik Gront's (dgront@gmail.com) lab on June 2019.

PURPOSE OF THE TEST

We test if the FragmentPicker picks fragments correctly. After running the FragmentPicker we run the ab-initio protocol with the RMSD to test the quality of the fragments.

BENCHMARK DATASET

The set has 10 proteins and contains alpha, beta and alpha-beta proteins with various lengths.

We pick some proteins from the BENCHMARK62 database.

Input files for the FragmentPiker are: pdb file, sequence profile in .profile format from blast, secondary structure predictions for proteins from three

predictors (psipred, porter and rdb(=SAM)), homolog file to excluded them from the VALL database, scoring-multirama.wgths, quota.def, .vall database.

Input files for the ab initio protocol are the pdb file, ideal-pdb which is referenced to calculate rmsd, fragment files generated in the first step of test.

PROTOCOL

The FragmentPicker picks structural fragments for single proteins from a database according to sequence and structure (from structure predictors') similarity. Then fragments are used in the ab initio protocol to build a model of protein structure.

The FragmentPicker protocol is described in (Gront D, Kulp DW, Vernon RM, Strauss CEM, Baker D (2011) Generalized Fragment Picking in Rosetta: Design, Protocols and Applications. PLOS ONE 6(8))

This protocol requires about 2,000 CPU hours.

PERFORMANCE METRICS

We use the rmsd between the native and decoys calculated in ab initio simulations. The score function contains the rmsd to the native as one of the scoring terms. Such a simulation therefore reflects how well the native can be reconstructed by the given set of fragments.

10th percentile of rmsd should be lower then cutoff to pass the test. If its higher then the test will fail.

The ab initio protocol was run and statistics from 10,000 decoys where made to define proper cutoffs.

KEY RESULTS

The baseline is based on fragments published in (Gront D, Kulp DW, Vernon RM, Strauss CEM, Baker D (2011) Generalized Fragment Picking in Rosetta: Design, Protocols and Applications. PLOS ONE 6(8))

There are no outliers in the dataset.

DEFINITIONS AND COMMENTS

The test is sensitive to input files - sequence profile and secondary structure predicions.

LIMITATIONS

The benchmark assumes the input files remains constant. The benchmark doesn't test the impact of secondary structure predictions, sequence databases and psiblast parameters.

One could include the full fragment picking procedure in the test - this would require psiblast runs. This is actually part of another test called make_fragments, set up by Dan Farrell.

One could include more and more varied protein cases in the benchmark.