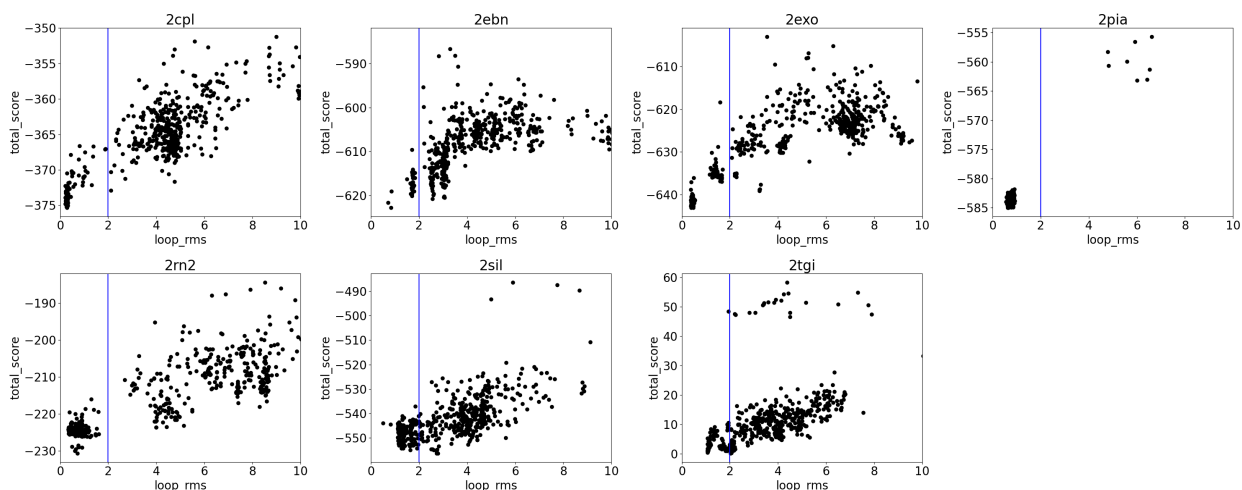# Scientific test: loop_modeling_ngk_12res

FAILURES

None

RESULTS

AUTHOR AND DATE

Phuong T. Nguyen, tranphuonguns@gmail.com, PI: Vladimir Yarov-Yarovoy, Jun 2019

PURPOSE OF THE TEST

This scientific benchmark tests loop modeling protocol using different algorithms:

- classical methods: cyclic coordinate descent (CCD) and kinematic loop closure (KIC)

- more recent kinematic-based methods: Next-generation kic (NGK) and kic with fragments (KIC-FRAG)

BENCHMARK DATASET

A standard 12-residue mini benchmark set taken from the Kortemme's lab full benchmark set: https://guybrush.ucsf.edu/benchmarks/benchmarks/loop_modeling

This set of proteins was manually curated to select 7 proteins with different sizes ranging from 100-400 amino acids.

Input proteins were preminimized with Rosetta, native coordinates of loop regions were removed along with sidechains of nearby residues within 10A.

Non-native conformations of these loops were then added as a starting conformation. Reference native structures are also provided for analysis.

PROTOCOL

The NGK and KIC-FRAG protocols run LoopModelerMover through RosettaScripts interface. This Mover currently only supports kinematic-based sampling. This was used in the Kortemme lab's benchmark sets, but with much older Rosetta version and scoring functions (score12, talaris2013, talaris2014). Here ref2015 scoring function is used. Note that, despite simulation configuration is called as config="kic", this mover appears to invoke ramping on repulsive and rama by default, which is NGK method.

RMSDMetric was used to calculate the difference in generated loop conformations compared to native.

No superposition was performed as the protocol does not move other parts of input proteins.

The classic protocols CCD and KIC run through loopmodel application. Rmsd calculation is integrated in the application.

The CCD and KIC-FRAG used fragments insertion while classical KIC and NGK do not.

Runtime (for nstruct = 500):

Total: 2456 CPU hours

NGK: 571 CPU hours | KIC-FRAG:766 CPU hours | KIC: 621 CPU hours | CCD: 498 CPU hours

PERFORMANCE METRICS

The total_score and rmsd (loop_rms) to the native were used as metrics. These values are plotted with expectation of forming a funnel shape. However, in actual modeling process, we don't know the native conformation, which makes the score vs rmsd plot not very useful. We rely mostly on the far-perfect Rosetta scoring function to make a decision. Here we chose 10% cutoff for score vs rmsd to describe ultimate pass/fail. The test will pass if we can find any structures in the top 10% scoring models that has rmsd to the native less than a cutoff value. We use a cutoff of 2.0A for all loop modeling protocols and all targets.

KEY RESULTS

Using nstruct = 500, we were able to observe somewhat funnel shapes in the total_score vs loop_rms plots across different targets with all four algorithms.

We also observed sub-angstrom accuracy for several targets.

The 2TGI structure appeared to be a difficult case and required a higher cutoff value to pass.

DEFINITIONS AND COMMENTS

All methods need both ends of loop regions to perform kinematic sampling, thus cannot model loop regions at termini. Cyclic coordinate descent (CCD) is the only method can do the job, with the condition not using kinematic sampling in the refinement stage. To do that, users need to specify "refine_ccd", instead of "refine_kic" in the running flags.

Phuong T. Nguyen (tigerous) is not an expert in Rosetta loop modeling.

LIMITATIONS

The dataset is not big and diverse enough to draw a full conclusion, but appear to be decent enough to describe a working algorithm.