

Cyrus Bench: Protein Loop Rebuilding in Rosetta

Overview

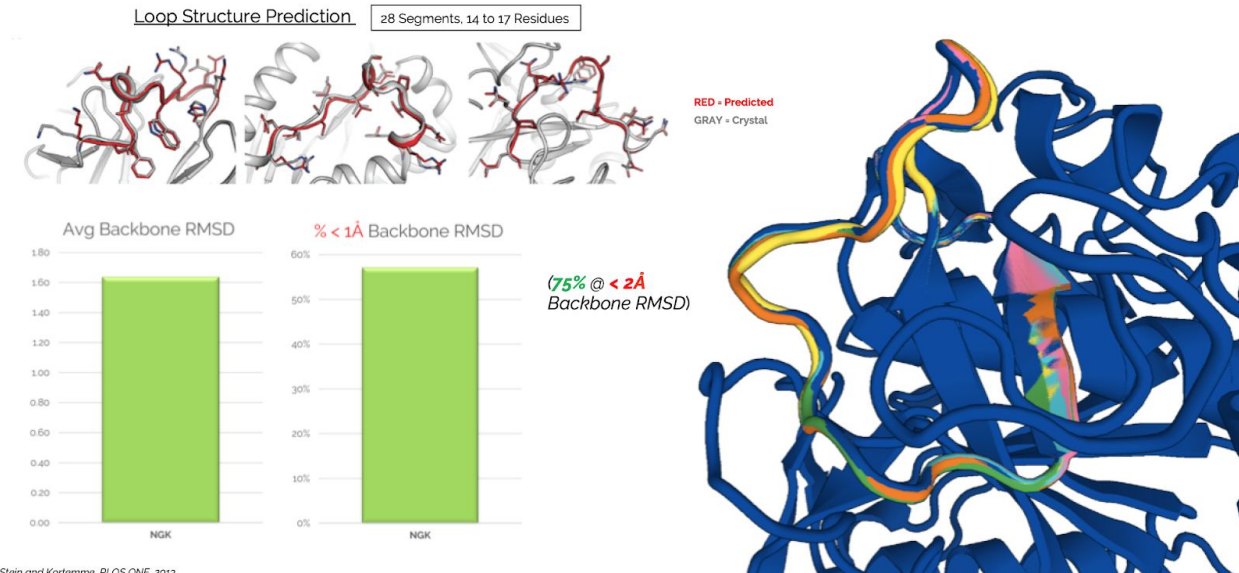
The Loop Rebuild tool is suitable for resampling the conformations of existing loops. This can be used to focus conformational resampling to flexible, exposed surface loops as an alternative to the Relax tool, which operates on the whole protein. Paired with the Design tool, Loop Rebuild can sample new backbones for new loop sequences. Finally, Loop Rebuild is useful for repairing loops with suspicious geometry (i.e. from weaker resolution crystal structures).

Cyrus's loop rebuild tool is a reimplementation of the "NGK, talaris 2013" protocol as benchmarked by the Kortemme lab [1], to make it compatible with and run stably in the current code-base of Rosetta. With apologies for the alphabet soup, "NGK" (Rosetta documentation: [2]) (paper: [3]) stands for "Next Generation KIC", where "KIC" is Rosetta's 'kinematic' loop closure algorithm, developed by the Kortemme lab (paper: [4]). At the time of this release Talaris2013 is the Rosetta scorefunction used throughout Bench.

Performance

The Loop Rebuild NGK tool in Rosetta is highly accurate in benchmark testing. In a test over 28 segments of 14 to 17 residues the tool gives an average backbone RMSD of 1.6 Angstroms with 75% of loops under 2 Angstroms RMSD. The figure below summarizes these benchmark results. This highly accurate tool has been used throughout Rosetta in both protein design and structure prediction.

Loop Modeling & Design



Loop selection

When selecting your loop endpoints for Loop Rebuild, keep these pointers in mind:

- Loops must be a minimum of 4 residues. This is a caveat of the equation solving underlying KIC -- it solves a system of 6 equations and needs 6 variables to do it: phi and psi for 3 residues. Cyrus Bench has chosen to impose a minimum of 4 rather than 3 to improve the user experience.
- Loop Rebuild performs best on loops of 12 residues or fewer, but has been used to generate very accurate loops for longer loops. Longer loops require more sampling (more repeats) to generate good results.
- All backbone atoms, torsions, and angles within the loop are sampled during loop modeling. The N-terminal nitrogen and C-terminal carbonyl carbon will not move; all other atoms internal to the loop may.

Input preparation

To ensure that analysis of Loop Rebuild models will be smooth, we encourage you to run a Repack operation first, before Loop Rebuild - this will make score comparison more straightforward. All sidechain positions in the protein are subject to repacking during Loop Rebuild (this is part of the NGK protocol).

Aggressive sampling mode

Cyrus Bench offers two modes for Loop Rebuild: normal and aggressive sampling. In normal sampling, the input loop is maintained, and each Monte Carlo cycle within loop remodeling makes minor perturbations from that starting point. In aggressive sampling, the starting loop conformation is deleted, and loop closure proceeds from scratch. (This is the “extended boolean” in Rosetta parlance). Aggressive sampling was used for the loop recovery benchmark to make it a more rigorous test. Broadly, aggressive mode is useful for finding conformations further from the starting model, but at the expense of requiring many more trajectories to produce high-quality models. Normal mode is better for refining loops and finding nearby conformations.

Number of repeats

For normal sampling of loops less than 12 residues, 50 repeats is likely to provide sufficient sampling, scaling up as the loop length increases. For aggressive sampling, 12mer loops are benchmarked at 500 repeats, but many structures find quality models with only a few hundred repeats. Again, as loop length increases, the number of repeats for sufficient sampling increases.

Broken loops

A side effect of the aggressive sampling mode is that input bond lengths are overwritten, with the side effect that a broken loop (represented by an overlong bond length) can be closed. This is currently more of a side effect than a feature and may be removed in the future when we add a new feature set for working with missing and broken loops. If you want to try closing broken loops, we encourage you to set the repeats to no more than 5; if those models return closed loops, use those as inputs for further rounds of loop modeling.

Missing loops

Loop Rebuild cannot currently change the lengths of loops. This includes loops that were present in the crystal but not present in the model - in other words, “missing” loops due to poor density. We are working on a separate loop modeling tool for broken and missing loops, allowing easy insertions/deletions and closure of broken loops.