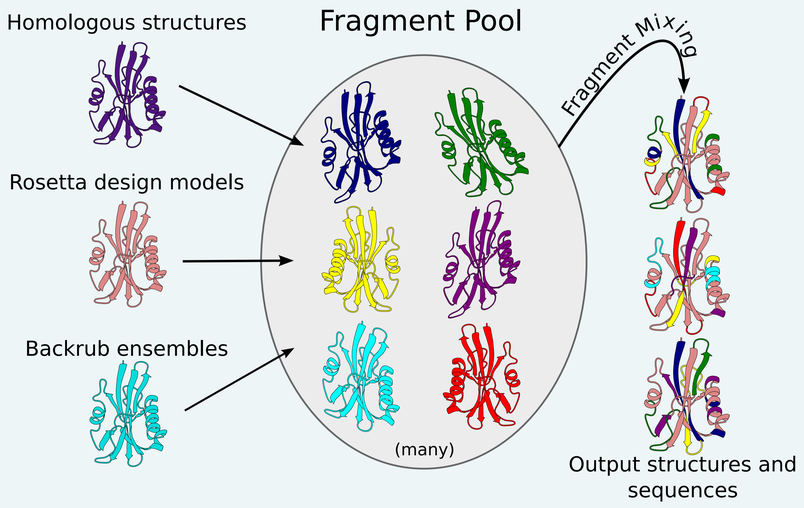
Fragment Mixing: Efficient computational protein structure modeling and design

Kyle Barlow and Tanja Kortemme — University of California, San Francisco

Overview



Fragments, data structures that typically store phi and psi torsion angles for short stretches of protein backbones, are used successfully within the protein modeling software suite Rosetta as part of protocols such as ab-initio prediction and loop modeling. In these protocols, a library of fragments is assembled by BLASTing a target sequence against the PDB and stealing fragments from homologous sequences. Fragments are then inserted into a target by setting the backbone ϕ and ψ torsion degrees of freedom to those stored in the fragment.

I have expanded upon this fragment implementation by storing fragments as coordinates, thus capturing all degrees of freedom, including amino acid identity, side chain torsion, bond lengths, and bond angles. "Inserting" these fragments by superimposing them onto a model in full-atom mode implicitly creates plausible coupled side chain/backbone moves, which can be difficult to sample using other methods. These fragment mixing moves can be used for protein design in a manner analogous to gene shuffling.

This protocol, "fragment mixing," remains under active development, but I have so far applied it to predicting tolerated sequence profiles, studying CypA conformational dynamics, and predicting structural changes in designed proteins.

Current fragment mixing protocol

- **Step 0:** Generate a fragment library by stealing coordinates from a list of input PDBs. These could include homologous structures, a backrub ensemble, or an ensemble of designed models.
- **Step 1:** Pick a random site in the starting model for fragment superposition/insertion (including potential mutation). Replace the residues in this site with the coordinates from the fragment.
- **Step 2:** Cartesian minimization. All atoms but those in fragment are free to move. This step also closes the chain break.
- **Step 3:** Side chain packing protocol: either the packer or rotamer trials protocol in a 4Å radius around the fragment insertion site.
- **Step 4:** Monte Carlo accept/reject based on score
- Repeat ∼100x-200,000x, with ramping temperature

Fragment mixing generates ensembles of predicted protein structures

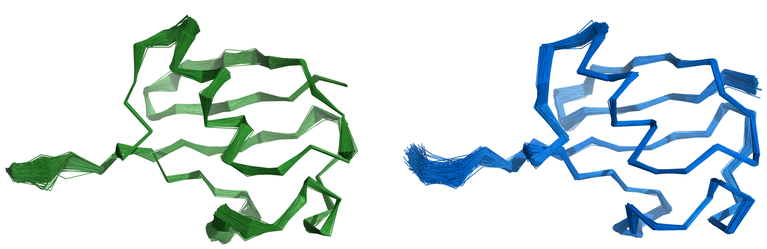


Figure 1: Left: An input ensemble for ubiquitin is generated using Rosetta's backrub protocol and used as input for fragment mixing. Right: The output ensemble generated by fragment mixing. Note that fragment mixing is capable of moving the N and C-terminal ends of the model (unlike backrub).

Sampling benefits

- These fragment insertions allow for **coupled backbone/side chain** sampling, on top of the coupled backbone sampling provided by standard fragments
- Fragments can be mapped by residue number instead of sequence, allowing fragment mixing to also **couple sampling of structure and sequence space**

Fragment mixing may increase sampling of biologically relevant alternative conformations

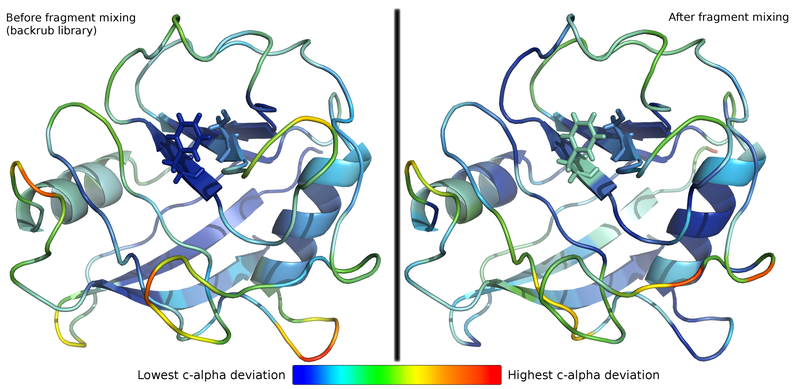


Figure 2: The average c-α deviation of each residue is colored onto the crystal structure of cyclophilin-A (3K0M). The c-α deviation is calculated for each member of the Rosetta output ensemble relative to the 3K0M structure, and colored according to the relative minimum and maximum deviation for that ensemble. **Left:** Backrub generates an ensemble with the most flexibility in loop regions **Right:** Fragment mixing (using the left backrub ensemble as the fragment pool) flexibility in loop regions differs from backrub flexibility. Additionally, residues 99 and 113 (shown in sticks), known to be important in the dynamics of this enzyme's functions, show increased backbone flexibility, indicating that fragment mixing may increase the sampling of biologically relevant conformations over the amount present in the input ensemble.

Design Benchmark

- Benchmark challenge: Use a set of 10 designs (4 representative examples are shown) that have crystal structures available for the starting scaffold and for the final design to test the ability of fragment mixing to predict the structure of the designed proteins after mutations are applied.
- Predicting the design structure is challenging because the mutations are produced by both computation and directed evolution, and are scattered throughout the protein

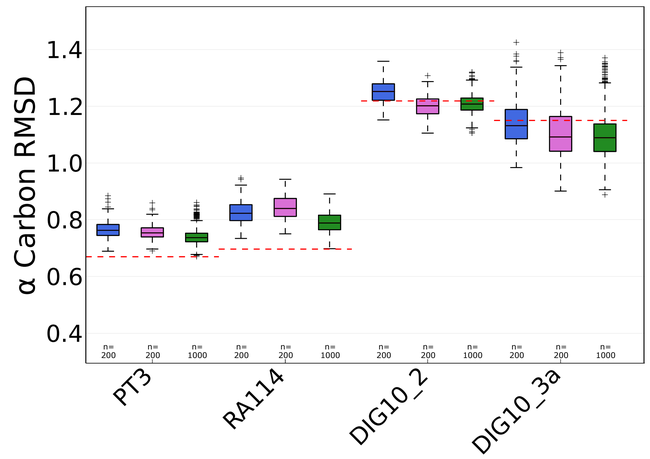


Figure 3: The RMSD of each α-carbon in a predicted structure compared to its known position is calculated and averaged. The distribution of this mean α-carbon RMSD for all predicted structures is shown in boxplots. Fragment mixing usually generates better predictions than minimization alone, but assuming no change from the starting scaffold structure (red dashed line) is also likely to be a good prediction. Boxplot colors: Backrub ensemble Cartesian minimized ensemble (control) Fragment mixing output ensemble Dashed line — scaffold starting structure

Potential applications for fragment mixing

- Sample combinations of mutations that were not previously scored together. Increase the chance of finding good combinations by bringing along some of the mutation's local environment in fragment form.
  - **Fragment mixing as a design tool**
- Generate structural ensembles using alternative Rosetta methods, and use these ensembles as input fragments.
  - **Fragment mixing as an ensemble generator/refiner**
- Generate ensembles outside of Rosetta, using experimental or alternate computational techniques, and refine/generate ensembles in the context of Rosetta's score function
  - **Fragment mixing as a novel inputter for Rosetta**

Acknowledgments

Funding provided by UCSF Discovery Fellowship, ARCS, SynBERC, and the Biological and Medical Informatics Graduate Program at UCSF. Many thanks to Tanja and my labmates for helpful discussion, and to the broader Rosetta development community for continuing inspiration and development.

Poster copy and contact info at:
http://www.kylebarlow.com