Structure Preparation

Marion Sauer

Vanderbilt University

Rosetta Workshop
May 9, 2018

1. Getting Started
2. Rosetta Minimization
3. Rosetta Relax
4. Adding Minimization or Relax Constraints
5. Summary
6. Questions

Getting Started

I have a structure – now what?

For starters, remove waters, non-canonical amino acids, ligands, or anything else not defined as an "ATOM" or "TER" type in a PDB file.

```
python clean_pdb.py <pdb> <chain ID>
```

Note: script may or may not remove selenomethionines (depending on script version) and removes residues with zero occupancy.

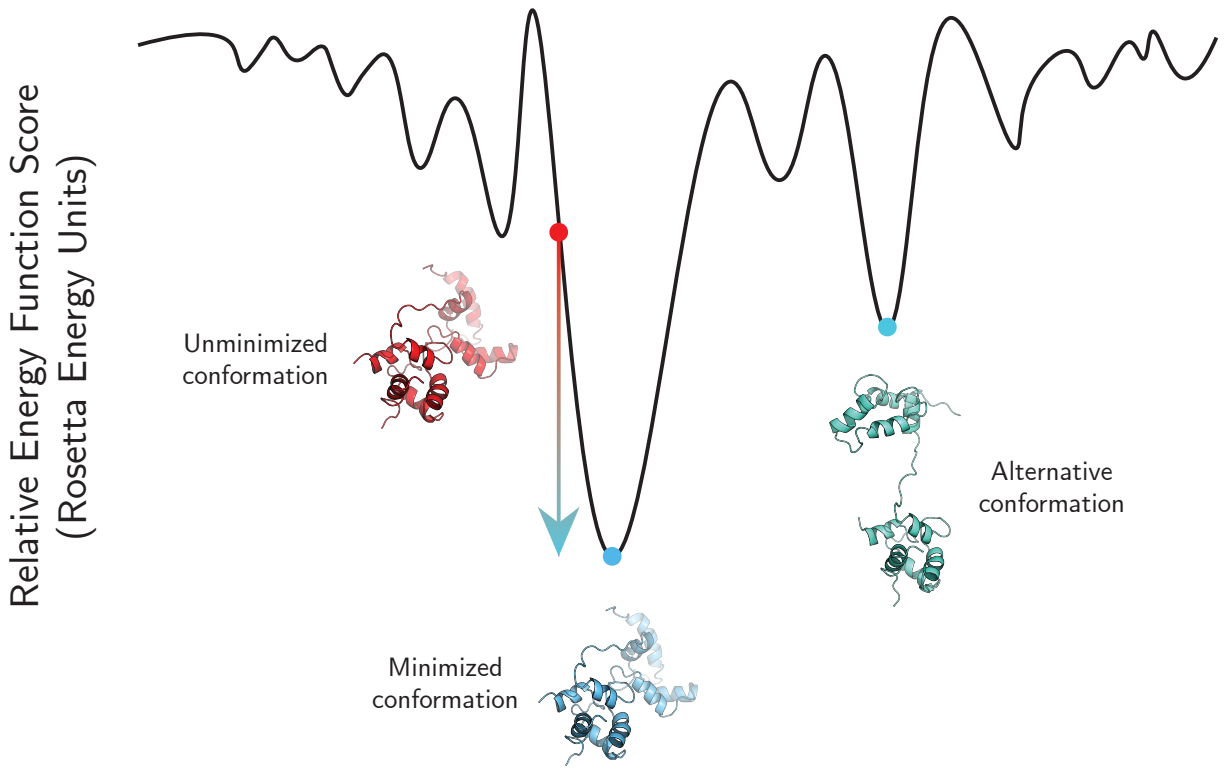
Work around

`-ingore_unrecognized_res` allows you to keep "HETATOM" types and waters

`-ignore_zero_occupancy false` loads residues/atoms that have zero occupancy

Rosetta Minimization

An introduction to Rosetta minimization



Goal

Identify a structure's conformation representative of the nearest local energy minimum using the Rosetta energy score function given the starting conformation and associated energy.

Impacts of minimization on design

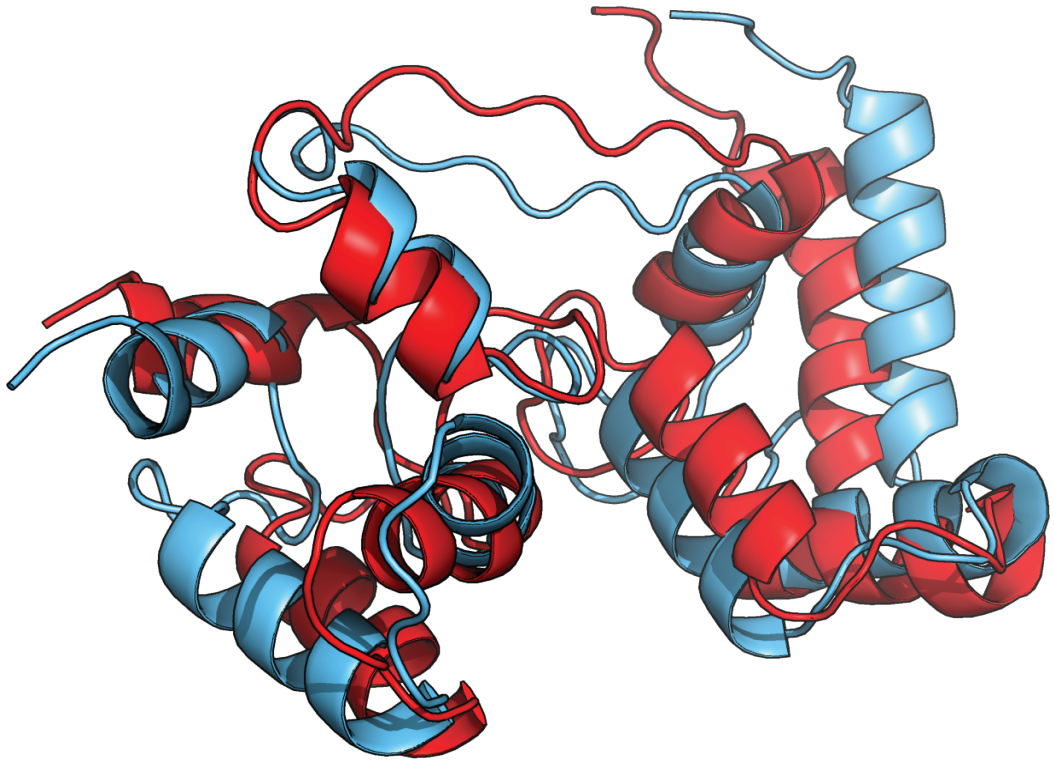


Figure: Minimization alters the input backbone conformation, sometimes dramatically

Impacts of minimization on design

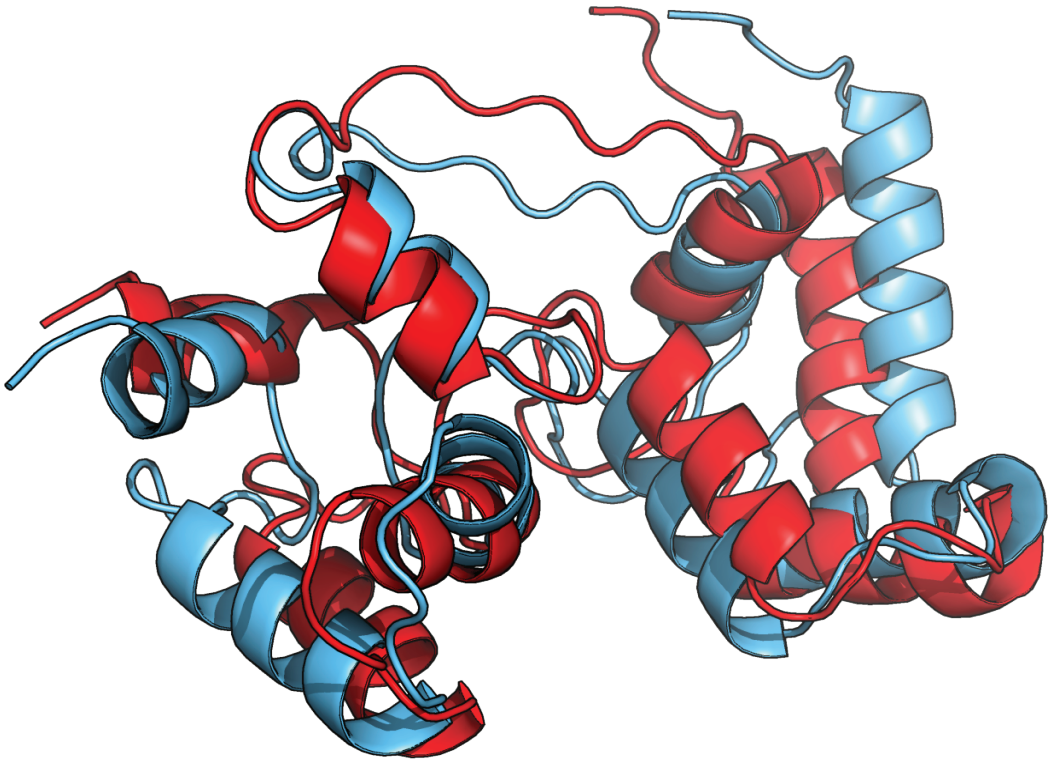


Figure: Design relies on the backbone coordinates/dihedral angles to predict favorable sidechain placement

Impacts of minimization on design

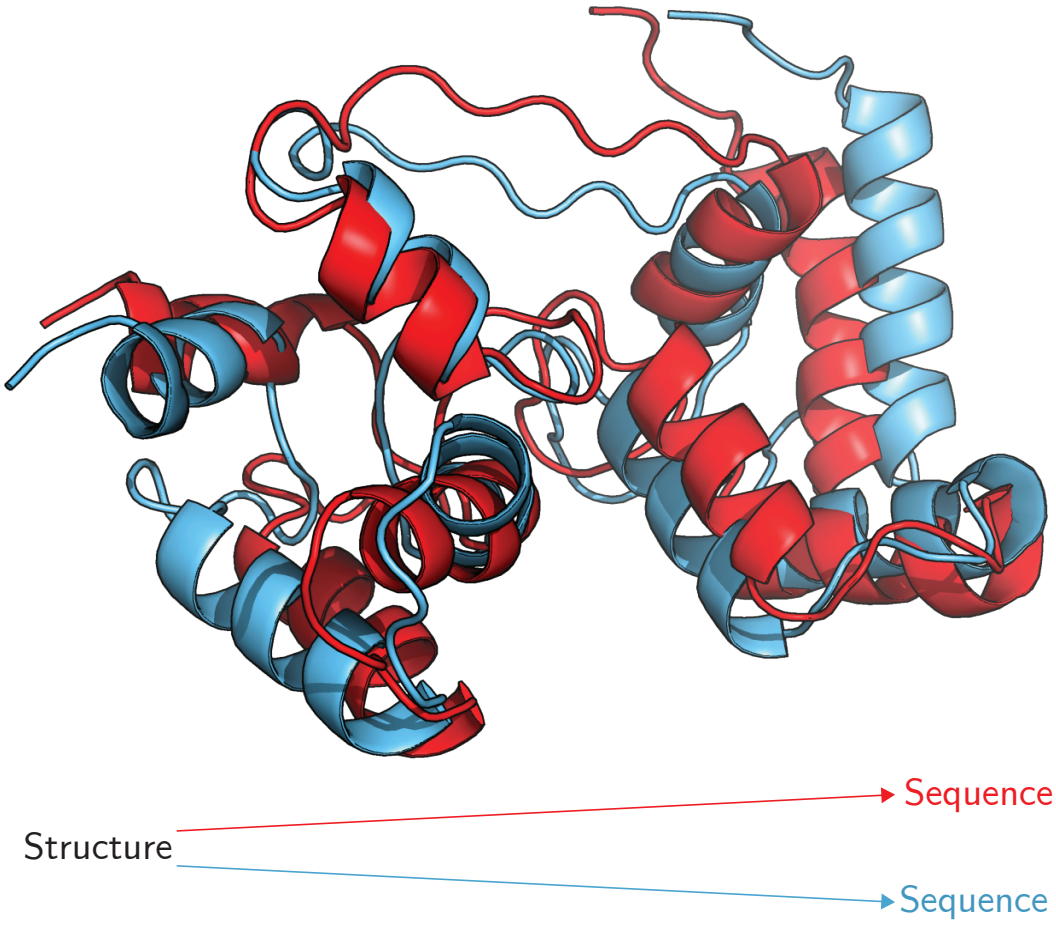


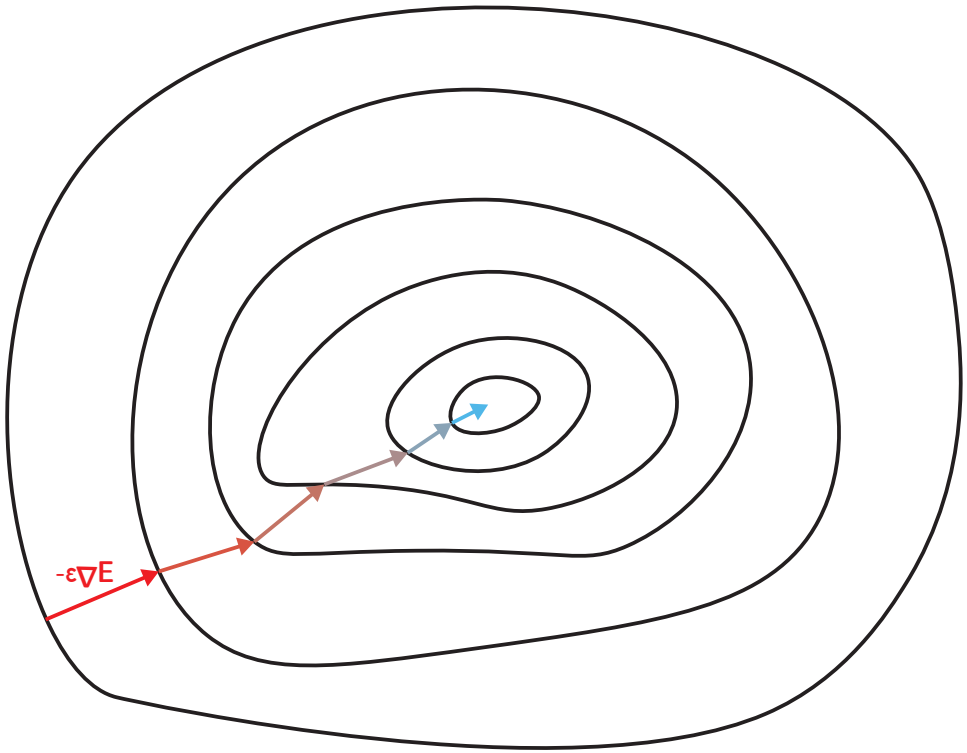
Figure: Altering the template backbone can, and most likely will, alter the predicted sequence tolerance of designs

Gradient-descent minimization

Calculate the overall gradient vector $\bigtriangledown E$ to get

$$\bigtriangledown E = \frac{dE}{dx_1}, \frac{dE}{dx_2}, \ldots \frac{dE}{dx_N}$$

where $x_1 \ldots x_N$ are the movable degrees of freedom



Minimization in practice

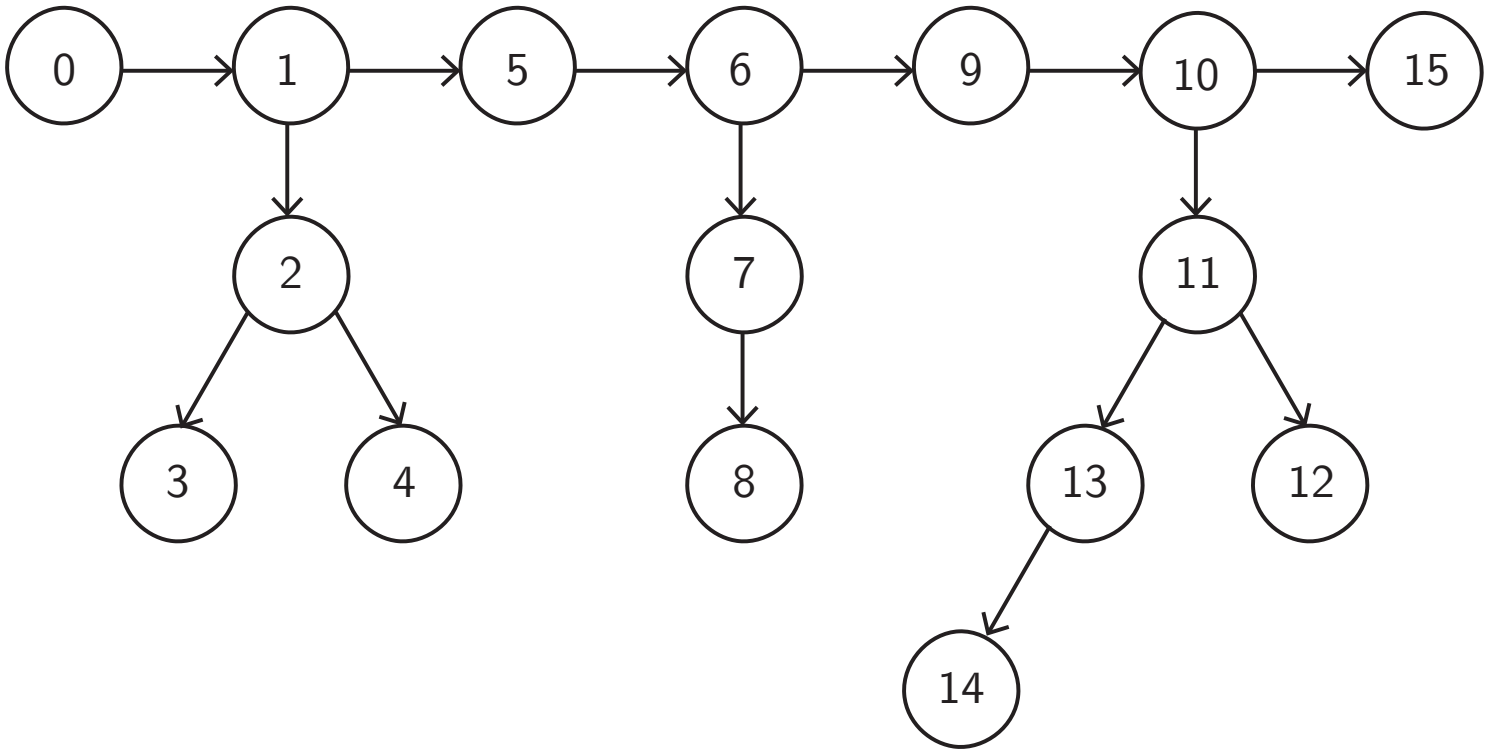


Figure: Minimization follows an order of hierarchy defined by the Rosetta foldtree. The degrees of freedom are defined by whether you use torsion space (dihedral angles), cartesian space (atom coordinates), or dualspace (a combination of the two). Users may control what degrees of freedom are allowed to change with a movemap.

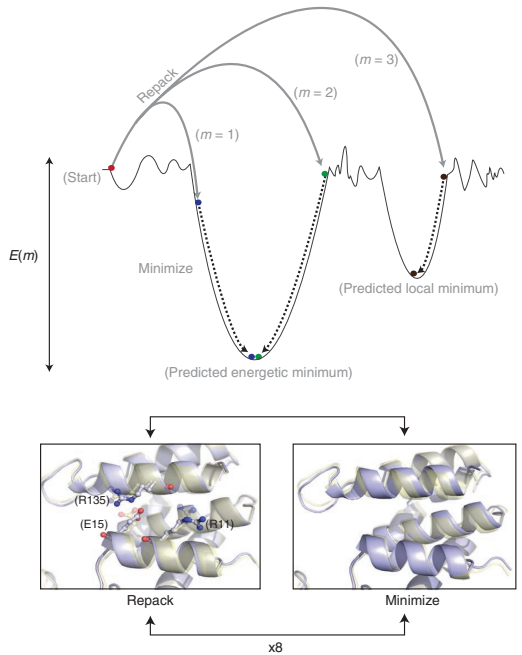
Different flavors of minimization

Rosetta minimization is based off the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method

- default: `lbfgs_armijo_nonmonotone` – best performance for large proteins
- small systems (*e.g.* small peptides): `dfpmin_armijo_nonmonotone`
- debugging: `linmin_iterated` – very slow but more accurate

Rosetta Relax

Relax is more than just minimization



[1]

Figure: FastRelax modes work by running sidechain repack and minimization cycles, ramping up or down the `fa_rep` weight of the forcefield.

Standard relax cycle script file

```
repeat 5
ramp_repack_min 0.020 0.01
ramp_repack_min 0.250 0.01
ramp_repack_min 0.550 0.01
ramp_repack_min 1     0.00001
accept_to_best
endrepeat
```

Parameters may be changed, but with extreme caution

[1] Combs, S.A., DeLuca, S.L., DeLuca, S.H., Lemmon, G.H., Nannemann, D.P., Nguyen, E.D., Willis, J.R., Sheehan, J.H., Meiler, J. (2013) Small-molecule ligand docking into comparative models in Rosetta. DOI: 10.1038/nprot.2013.074

Adding Minimization or Relax Constraints

Relax with all-heavy-atom constraints

How do you know if FastRelax moved the backbone too excessively?

- `-constrain_relax_to_start_coords` – discourages backbone movement away from starting coordinates by adding backbone coordinate constraints
- `-relax:constrain_relax_to_native_coords` – uses model passed to `-in:file:native` for backbone coordinate contraints
- `-relax:coord_constrain_sidechains` – also adds side chain coordinate constraints; requires one of the two previous flags
- `-constraints:cst_fa_file your_structure_cs.cst` – add custom constraints
- `-relax:script Rosetta/main/source/src/apps/public/relax_w_allatom_cst/always_constrained_relax_script` – forces constraints to stay on during the entire run

Creating a movemap to restrict movement

Each line in a movemap file identifies a jump, residue or residue range, and the allowed degrees of freedom as follows:

```
RESIDUE <#> <BB/CHI/BBCHI/NO>        # a single residue <#> followed by a single option
RESIDUE <#> <#> <BB/CHI/BBCHI/NO>    # a range of residues from <#1> to <#2>
                                     #     followed by a single option
JUMP <#> <YES/NO>                    # a jump <#> followed by a single option
```

For example,

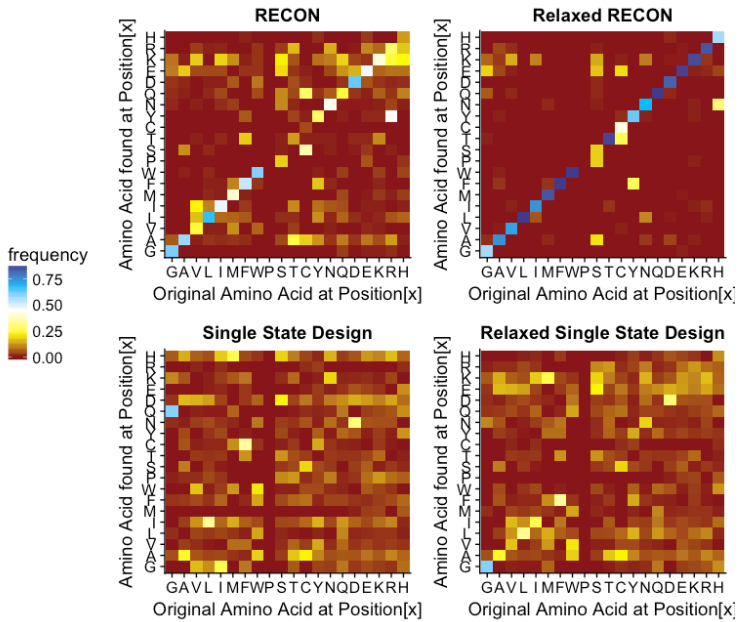
```
RESIDUE 28 BB         # allows backbone movements at residue 28
RESIDUE 32 48 BBCHI   # allows backbone and sidechain chi movements from residues 32 - 48
JUMP 1 YES            # allows rigid-body movents between the structures separated by jump 1
```

Summary

Consequences of altering a structure

Minimization benchmarks show that the energy differences result from improvements in the `fa_dun` and `fa_atr` terms

- Resolution of starting template is important - lower template resolution results in greater sampling away from the native sequence
- Greater minimization/relaxation results in more conservative design sampling
- Idealizing the starting template for the Rosetta scoring function introduces sampling bias



Questions