6 flow charts to accompany tutorials

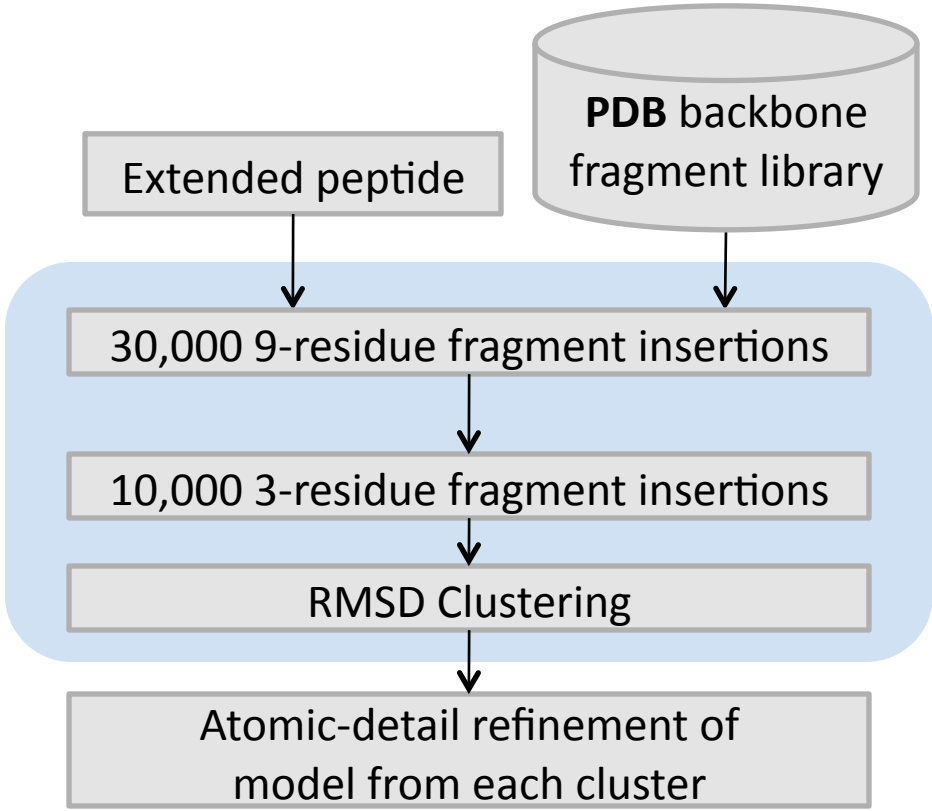
Text on the left is quoted from the Rosetta Review article

Blue boxes represent the protocol. Outside the blue box are inputs and outputs

1. De Novo Protein Folding
2. Loop Modeling
3. Refinement
4. Protein/protein docking
5. Small molecule docking
6. Protein design

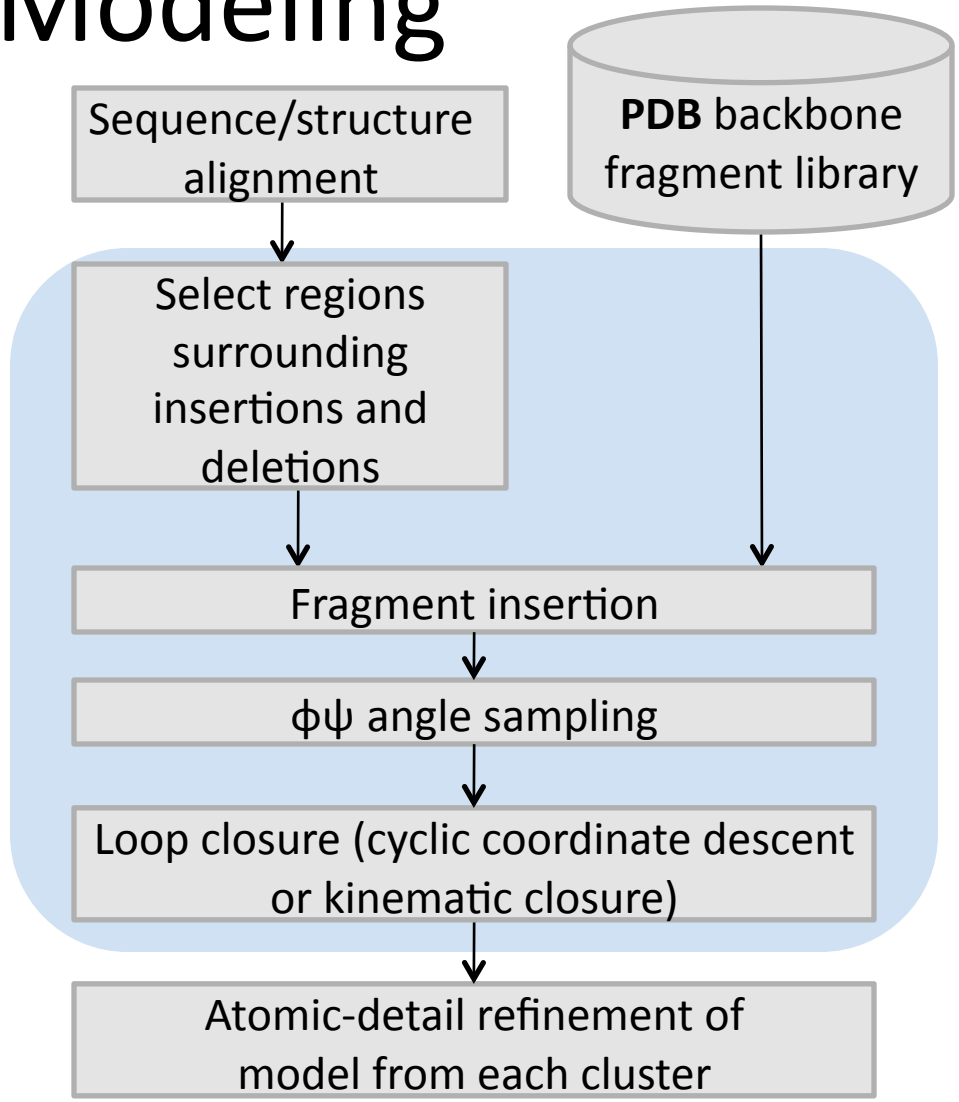
de novo Protein Folding

Rosetta begins with an extended peptide chain. Insertion of backbone fragments rapidly "folds" the protein using the low resolution energy function and sampling approaches. Rosetta attempts approximately 30,000 nine residue fragment insertions followed by a further 10,000 three residue fragment insertions to generate a protein model [6]. Usually 20,000-50,000 models are folded for each individual protein [15]. The resulting models can either undergo atomic-detail refinement or if computational expense is an issue, clustering based on C_α-RMSD [16, 17] can reduce the number of models before performing refinement.



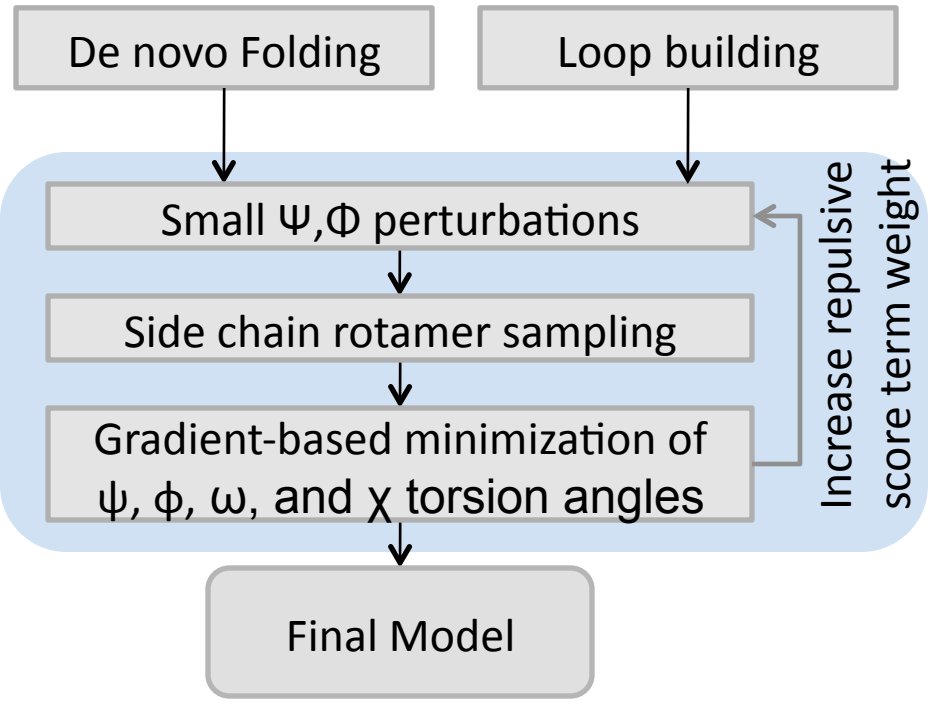
Loop Modeling

Comparative modeling in Rosetta starts with the alignment of a target sequence to a template protein, using sequence-sequence or sequence-structure alignment tools (as described by Raman et al [18]). The quality of the alignment determines the aggressiveness of the sampling in Rosetta [18]. In case of high sequence homology (sequence identity larger than 50%), the protein backbone is only rebuilt in regions surrounding insertions and deletions in the sequence alignment [18, 19]. Consequently, instead of starting from the extended chain, Rosetta starts with the template structure and builds in missing loops using fragment insertion of randomization of phi, psi angles followed by one of the loop closure algorithms such as cyclic coordinate descent or kinematic closure [20-22]. In the case of medium to low sequence identity between template and target, Raman et al. applied a more aggressive iterative stochastic rebuild and refine protocol that allowed the complete rebuilding of large regions of the protein including entire secondary structure elements in some cases



Refinement

After constructing a protein backbone via *de novo* protein folding or comparative modeling, the model enters atom-detail refinement [15, 23, 24]. During the iterative relaxation protocol phi, psi angles of the backbone are perturbed slightly while maintaining the overall global conformation of the protein. The side chains of the protein are adjusted using a simulated annealing Monte Carlo Metropolis search of the rotamer space. Finally, gradient minimization is applied to all torsional degrees of freedom (phi, psi, omega, and chi). The repulsive portion of van der Waals potential is increased incrementally, moving the structure to the nearest local minimum.

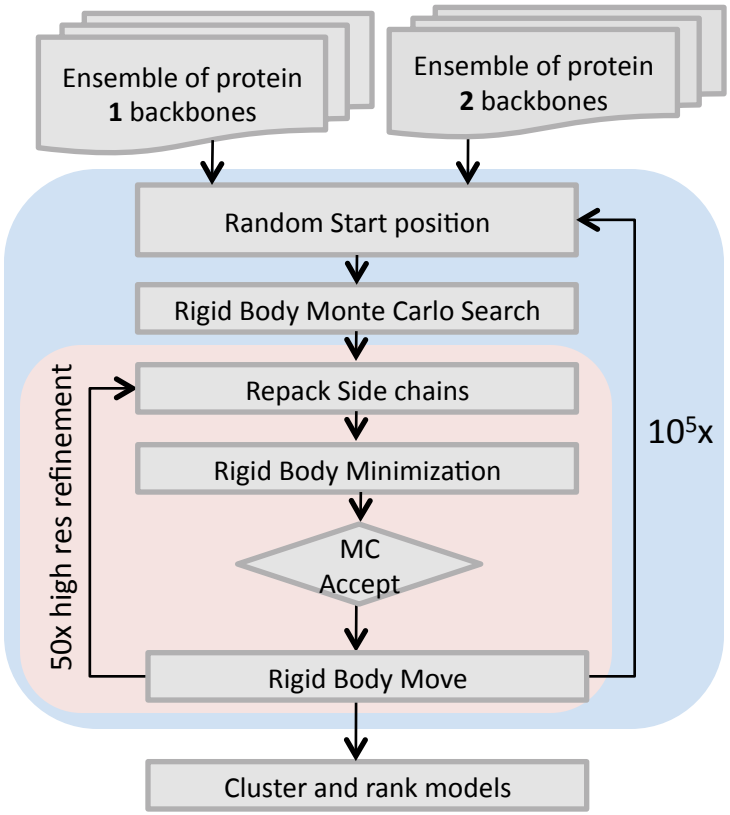


Protein/Protein docking

RosettaDock employs first a low-resolution rigid-body docking. The second high-resolution refinement stage provides for side-chain conformational sampling and backbone relaxation.

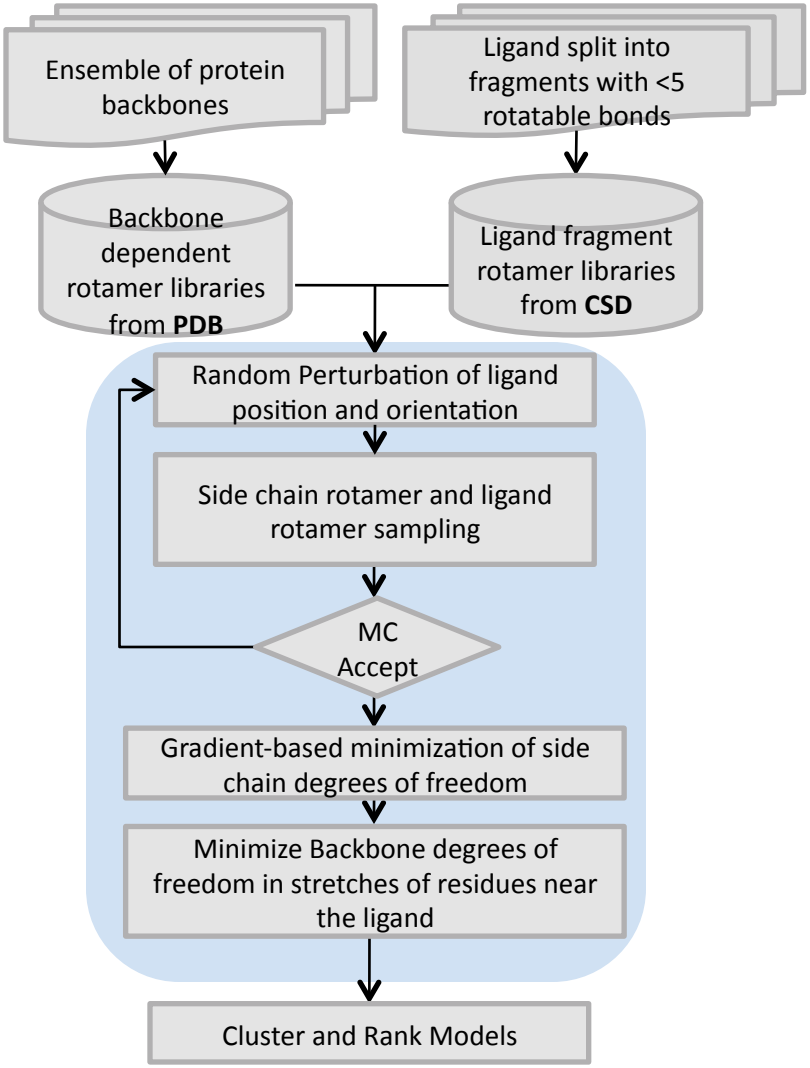
The RosettaDock algorithm begins with random reorientation of both proteins [46]. Next one protein slides into contact with the other. The following low resolution docking conformational search involves 500 Monte Carlo rigid body movements. These moves rotate and translate one protein around the surface of the other with movements chosen from a Gaussian distribution centered at 5° degrees and 0.7 Å. Each conformation is scored using the low-resolution energy function based on residue pair interaction statistics, residue environment statistics, and van der Waals attractive and repulsive terms. In this low resolution step, side-chains are represented by their centroids.

Next, 50 cycles of high resolution refinement at atomic detail are performed. Each cycle consists of a 0.1 Å random rigid-body translation, Monte Carlo based side-chain rotamer sampling (packing), and gradient-based rigid-body minimization to find a local energy minimum. Finally backbone flexibility is introduced around the protein interface.



Protein/small molecule docking

The RosettaLigand algorithm is a modification of the RosettaDock algorithm. First, a ligand conformer is chosen randomly from a user provided ligand conformational ensemble. Second, the ligand is moved to a user defined putative binding site. A low-resolution shape-complementarity search translates and rotates the ligand optimizing attractive and repulsive score terms. In the high-resolution phase cycles of Monte Carlo minimization perturb the ligand pose and optimize amino acid side-chain rotamers and ligand conformers. Lastly all torsion degrees of freedom in ligand and protein undergo gradient minimization and the model is output.



Protein Design

The RosettaDesign algorithm [60] energetically optimizes both the structure and sequence of a protein. A Monte Carlo simulated annealing search is used to sample the sequence space. Every amino acid is considered at each position in the sequence, and rotamers are constrained to the Dunbrack Library [61].