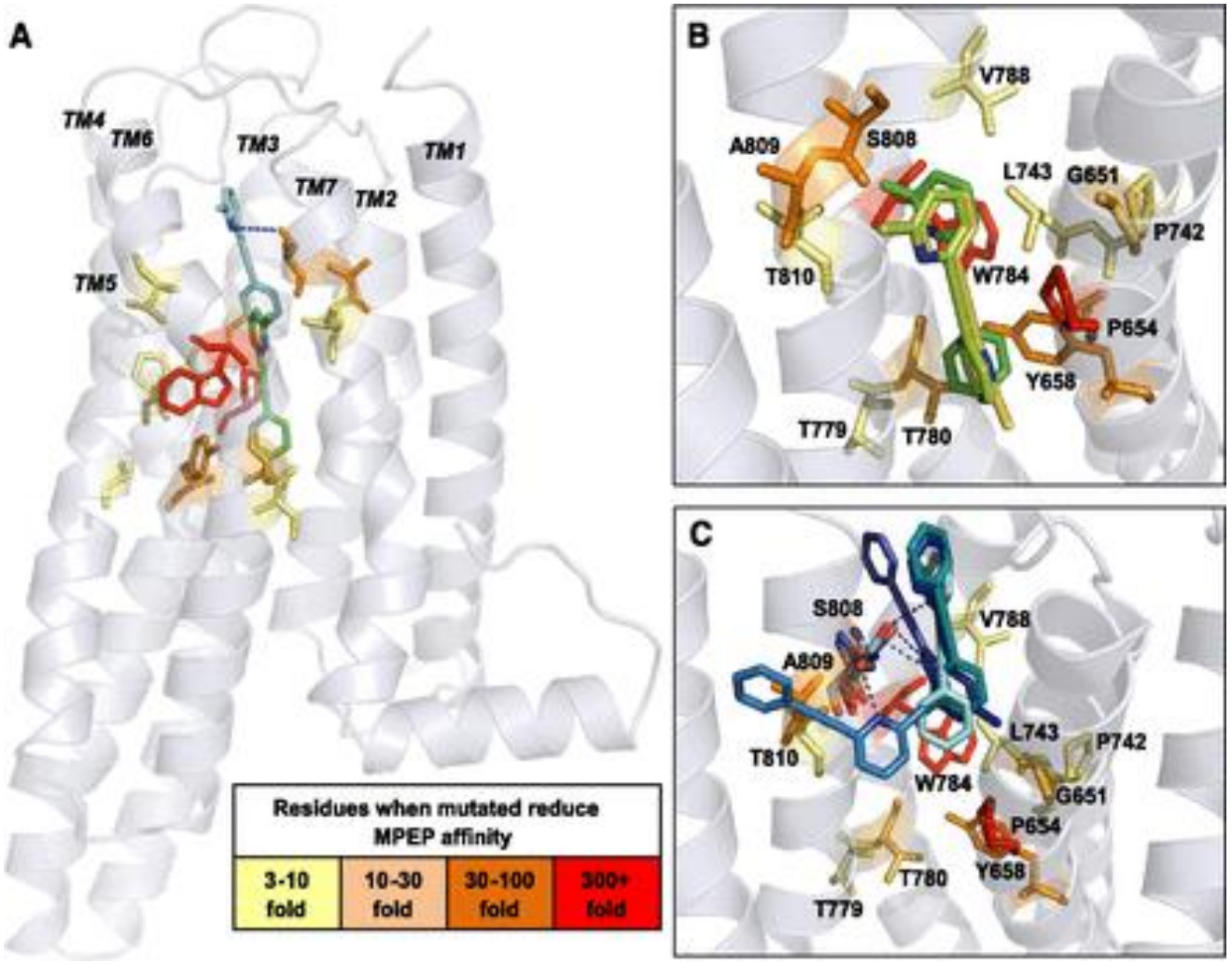
RosettaLigand Docking Tutorial

Rosetta Workshop – March 4, 2025

Chris Jurich
chris.jurich@vanderbilt.edu
Thomas Scott
james.t.scott@vanderbilt.edu
Meiler Lab
Vanderbilt University

Ligand Docking and Rosetta



K. J. Gregory *et al.*, Probing the metabotropic glutamate receptor 5 (mGlu5) positive allosteric modulator (PAM) binding pocket: discovery of point mutations that engender a "molecular switch" in PAM pharmacology., *Molecular pharmacology* **83**, 991–1006 (2013).

Some Quick and Dirty Terminology

- Ligand: small molecule binder (“drug-like”)
- Chain: identifier for protein/ligand molecule
- Low-resolution/Coarse-grained: fast centroid or grid based sampling steps
- High-resolution/Atomistic: full-atom steps with complete Rosetta energy function
- Integer: 0,1,2,3,4…etc.
- Float: 1.2, 2.3, 0.78…etc.
- Boolean: True/False
- String: alpha/numeric/symbol combination

RosettaLigand History

Meiler and Baker 2006

- Protein and ligand ensembles
- Side chain flexibility

Davis and Baker 2009

- Backbone flexibility near binding site

Lemmon and Meiler 2012

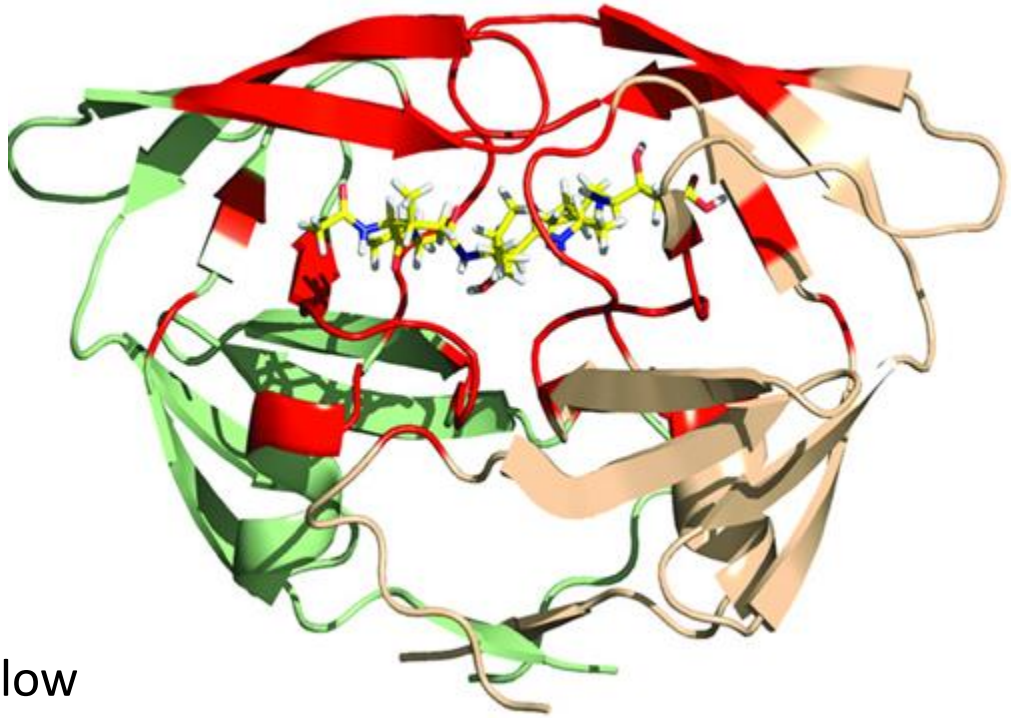
- XML Scripts

Deluca and Meiler 2015

- High throughput screening and improved low resolution sampling

Fu and Meiler 2018

- RosettaLigand Ensemble



HIV-1 PR homodimer (green/wheat) with acetylpepstatin (yellow) in binding site (red)

G. Lemmon et. al. *Chemical biology & drug design* (2012).

When is RosettaLigand likely to work?

Crystal structures (ideally in complex with ligand of similar chemotypes)

Rigid, well-defined binding pocket

Drug-like small molecules capable of making hydrophilic/H-bond interactions

Experimental restraints

Docking (finding binding mode) is much easier than ranking or predicting activity!

Davis, I. W., Raha, K., Head, M. S. & Baker, D. Blind docking of pharmaceutically relevant compounds using RosettaLigand. *Protein Sci.* **18,** 1998–2002 (2009).

What are the alternatives of RosettaLigand

Zero knowledge of drug binding site

Millions of compound docking

Fast screening of small molecules

Machine learning based methods such as Equibind and Diffdock can be used.

For more information please checkout these paper;

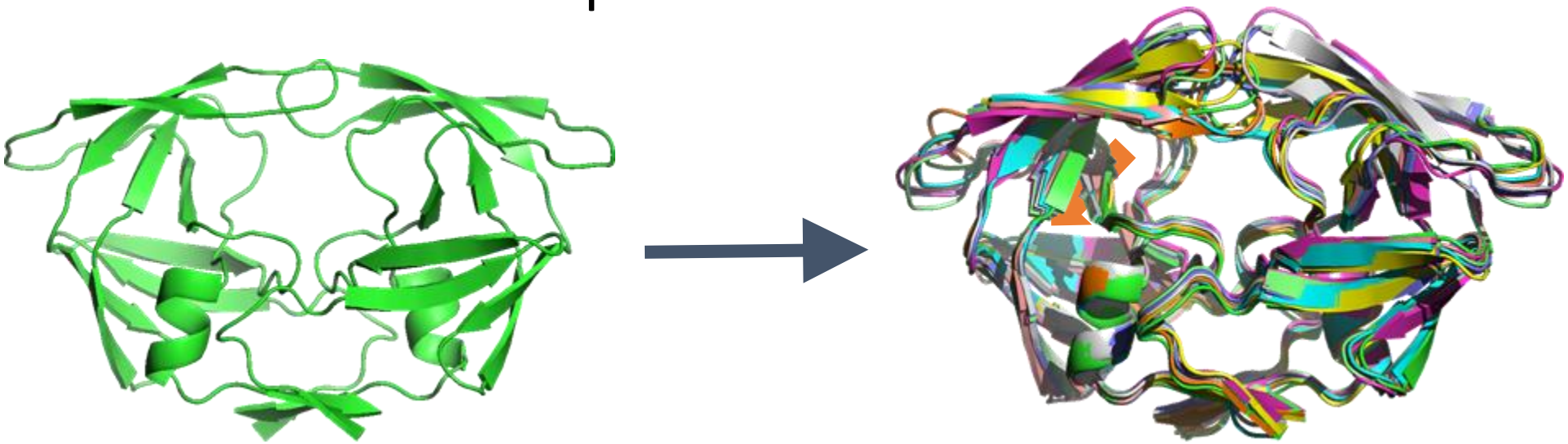
Stärk et. al; EquiBind: Geometric Deep Learning for Drug Binding Structure Prediction, Doi: 10.48550/ARXIV.2202.05146

Corso et. al; DiffDock: Diffusion Steps, Twists, and Turns for Molecular Docking, Doi: https://doi.org/10.48550/arXiv.2210.01776

Protein Model Selection

Protein conformation sources:

- Crystal structures/NMR
- Comparative modeling
- Alphafold, Omegafold for de-novo structure prediction

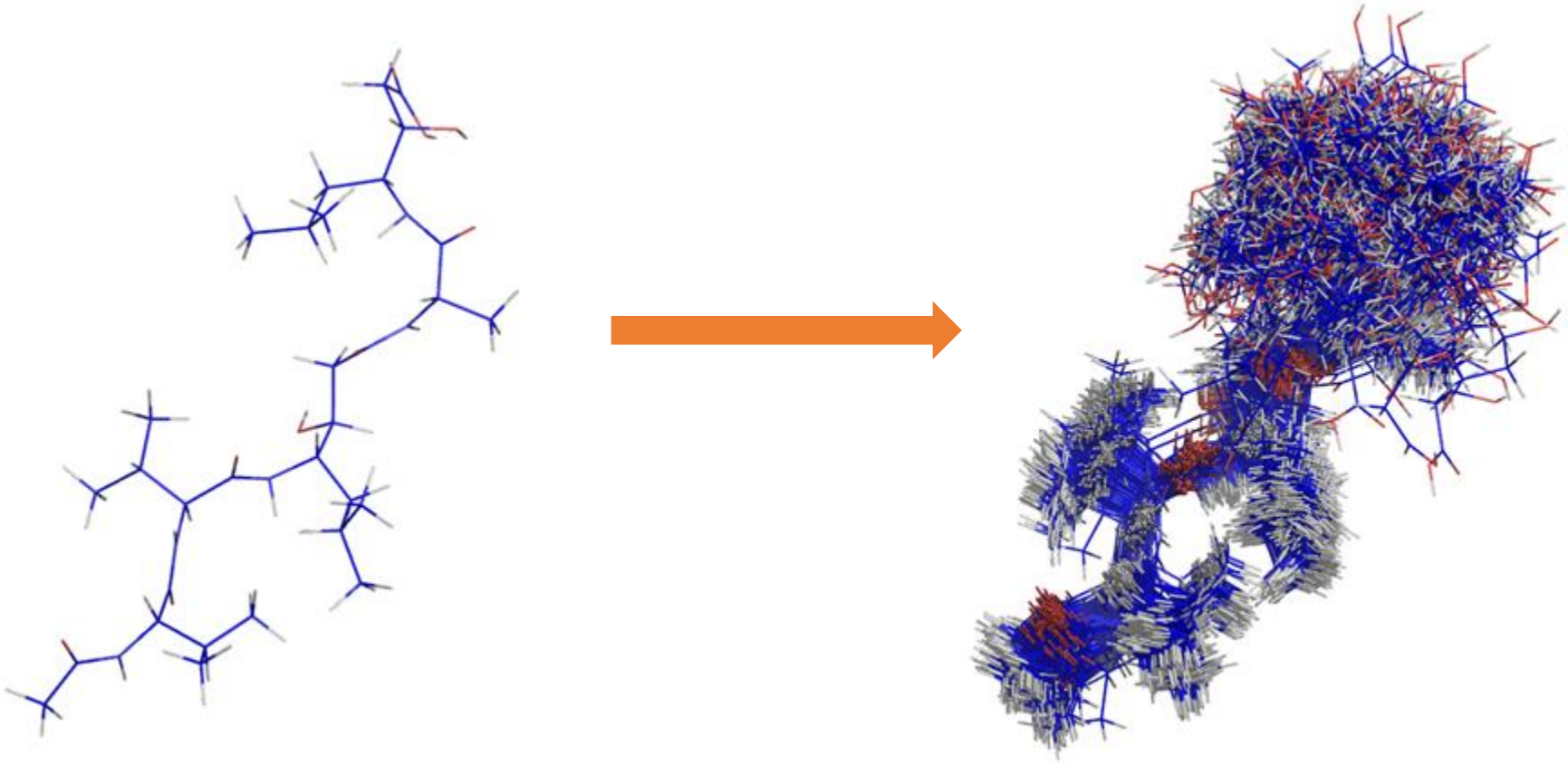


Better to run dock into multiple models

Ligand Conformer Generation

Ligand conformation sources:

- BioChemicalLibrary (BCL) http://www.meilerlab.org/servers/bcl-academic-license
- Commercial software: MOE/OpenEye
- FROG2 Server http://bioserv.rpbs.univ-paris-diderot.fr/services/Frog2/



The Ligand Params File

Residue Parameter Files define:

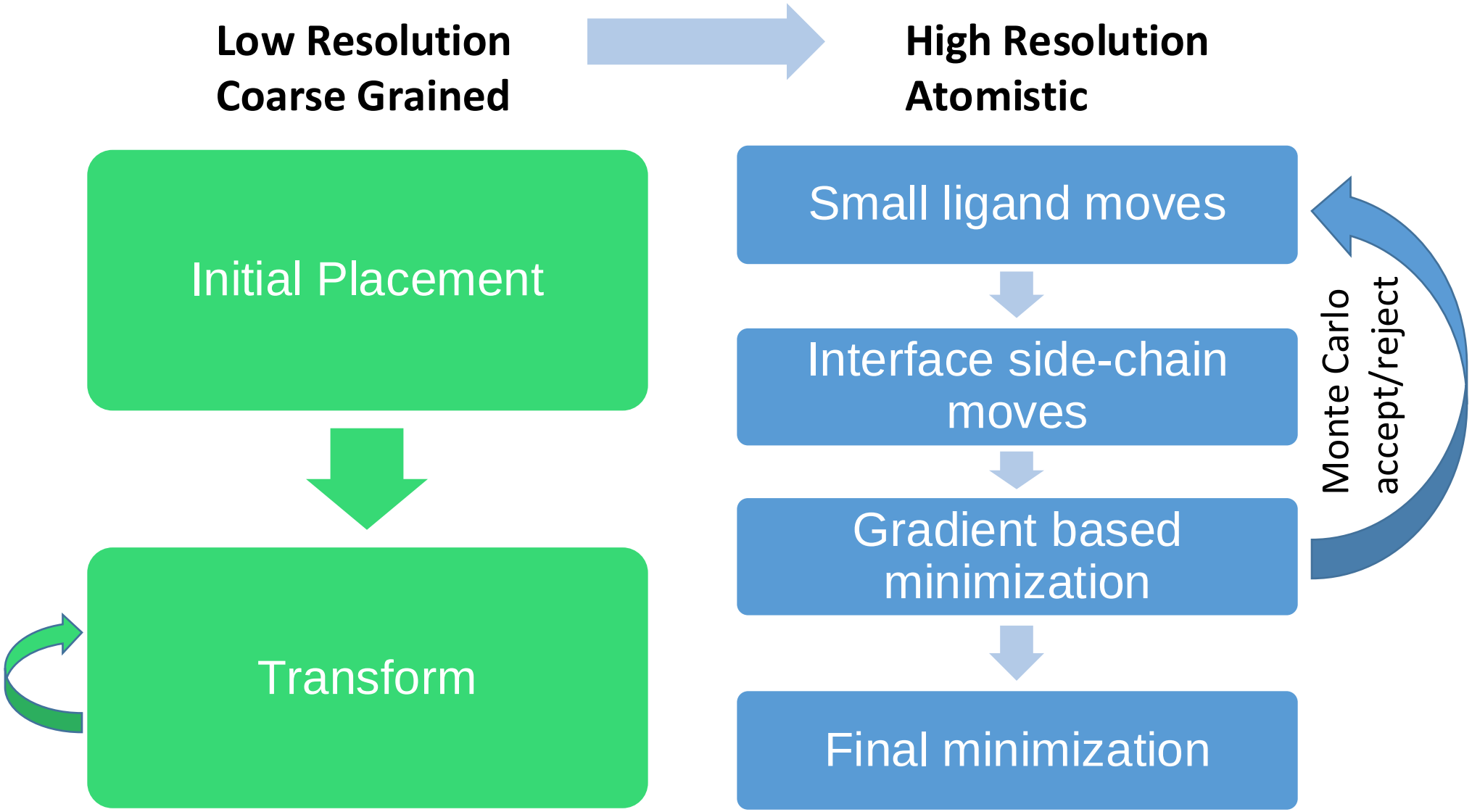
- Residue name and type
- Atom names, types, and partial/total charges
- Bonds, chi angles, and rotamers
- Internal coordinates
- Additional properties
- Points to conformer file

Rosetta Database has existing parameter files for amino acids, but we need to generate custom ones for ligands

Script to generate params file:

$ROSETTA/main/source/scripts/python/public/molfile_to_params.py

RosettaLigand Algorithm Overview



Overview of XML Script

SCOREFXNS
- Soft-rep
- Hard-rep

SCORINGGRIDS

LIGAND_AREAS

INTERFACE_BUILDERS
- Side-chain and backbone

MOVEMAP_BUILDERS
- Docking and Minimization

MOVERS
- StartFrom
- Transform
- HighResDocker
- FinalMinimizer
- InterfaceScoreCalculator

Low Resolution Steps High Resolution Steps Evaluation Steps

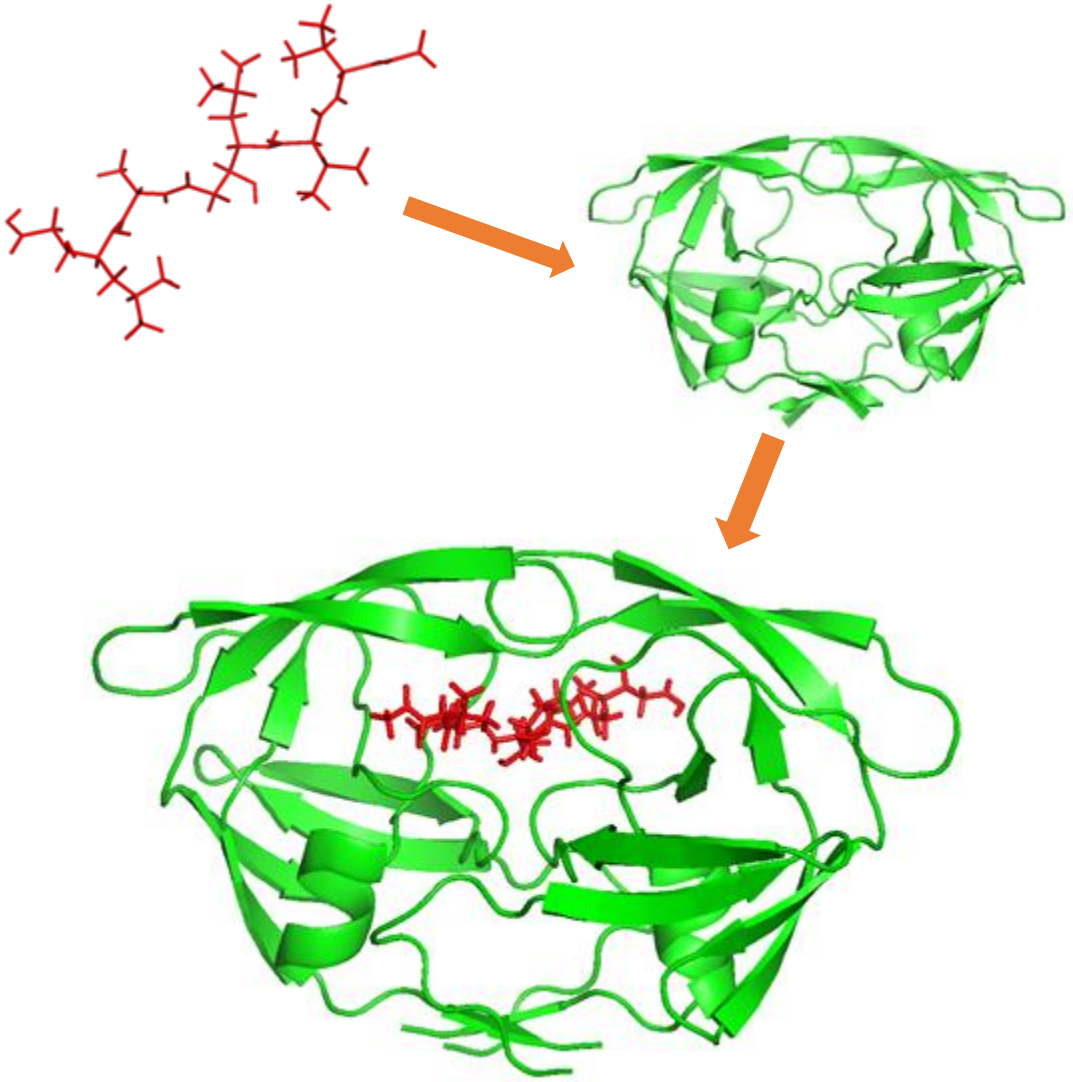
RosettaLigand Algorithm

Initial placement

Transform

Docking Cycles

High res refinement



StartFrom

Can also manually move ligand to the starting position

```
<StartFrom name=(string) chain=(string)/>
        <Coordinates x=(float) y=(float) z=(float)/>
</StartFrom>
```

Places the centroid of the ligand into the starting spot

Optional. Provide a list of possible XYZ starting coordinates. One of these will be chosen at random for initial ligand placement.

RosettaLigand Algorithm

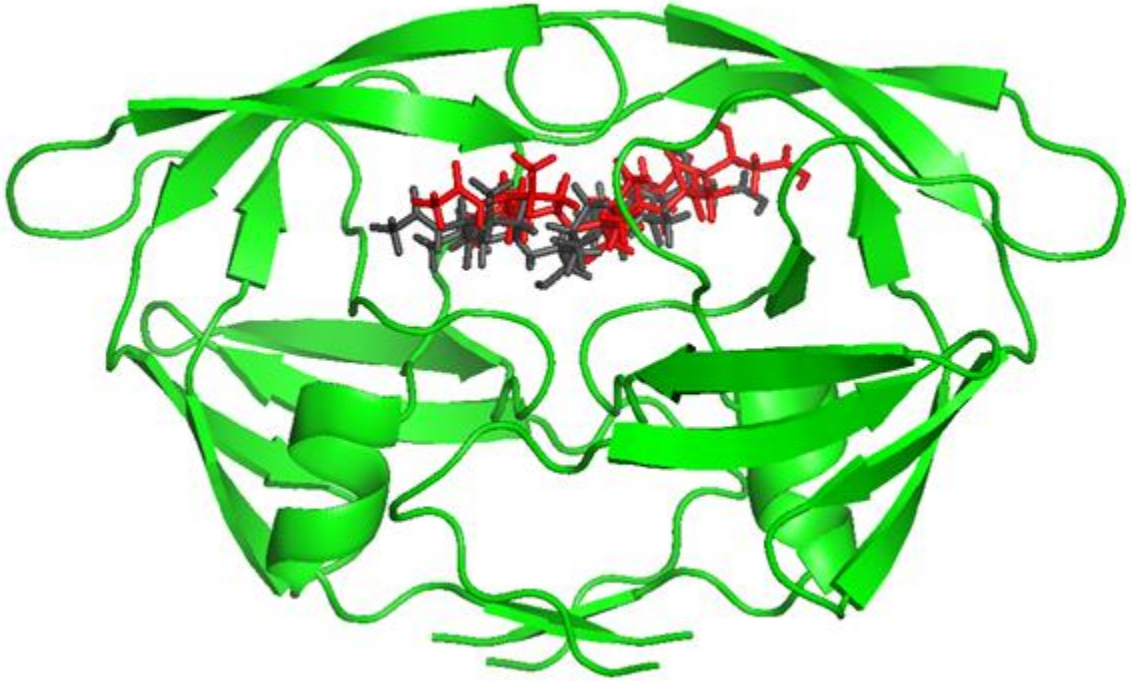
Initial placement

Transform

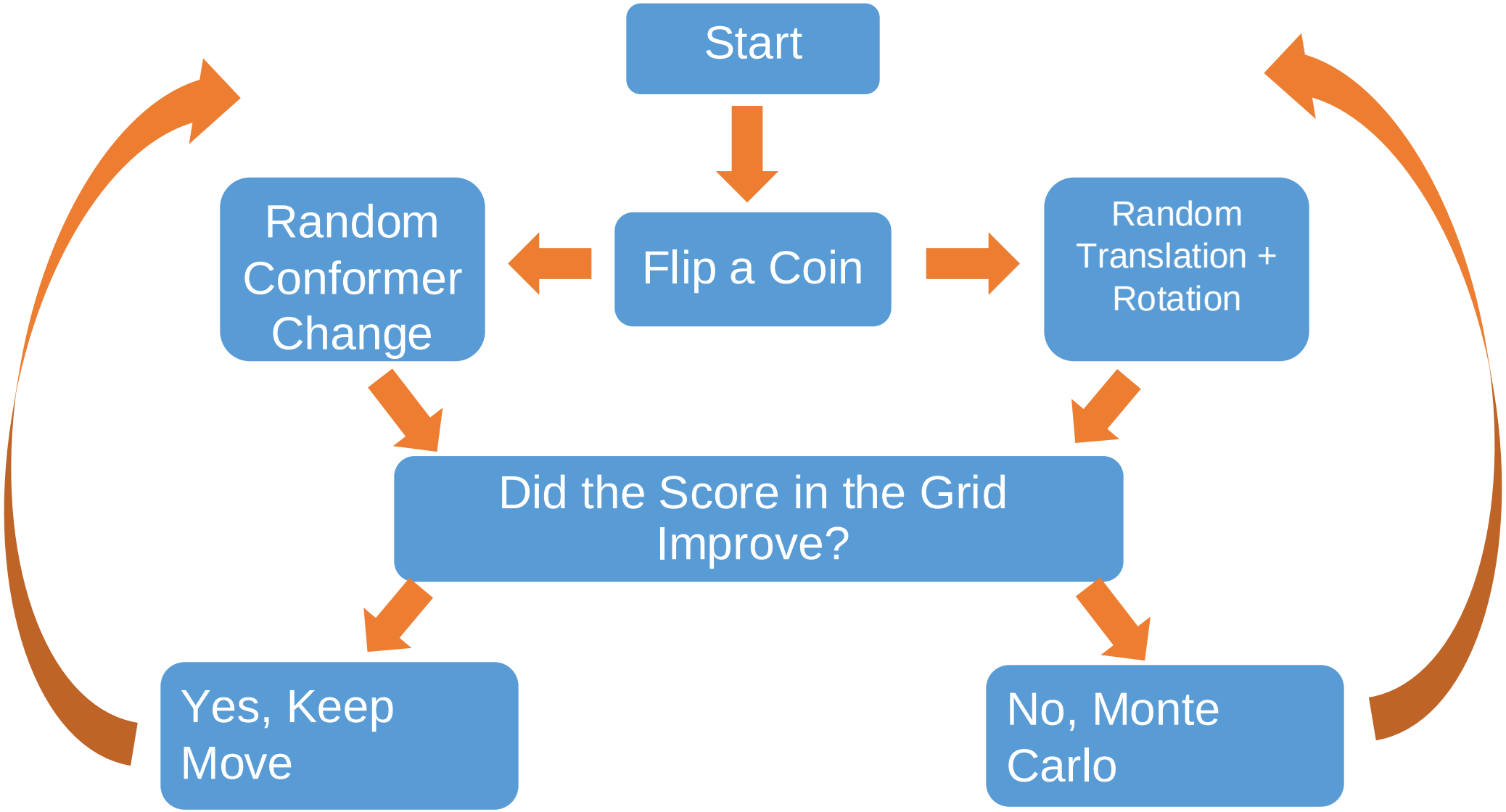
Docking Cycles

High res refinement

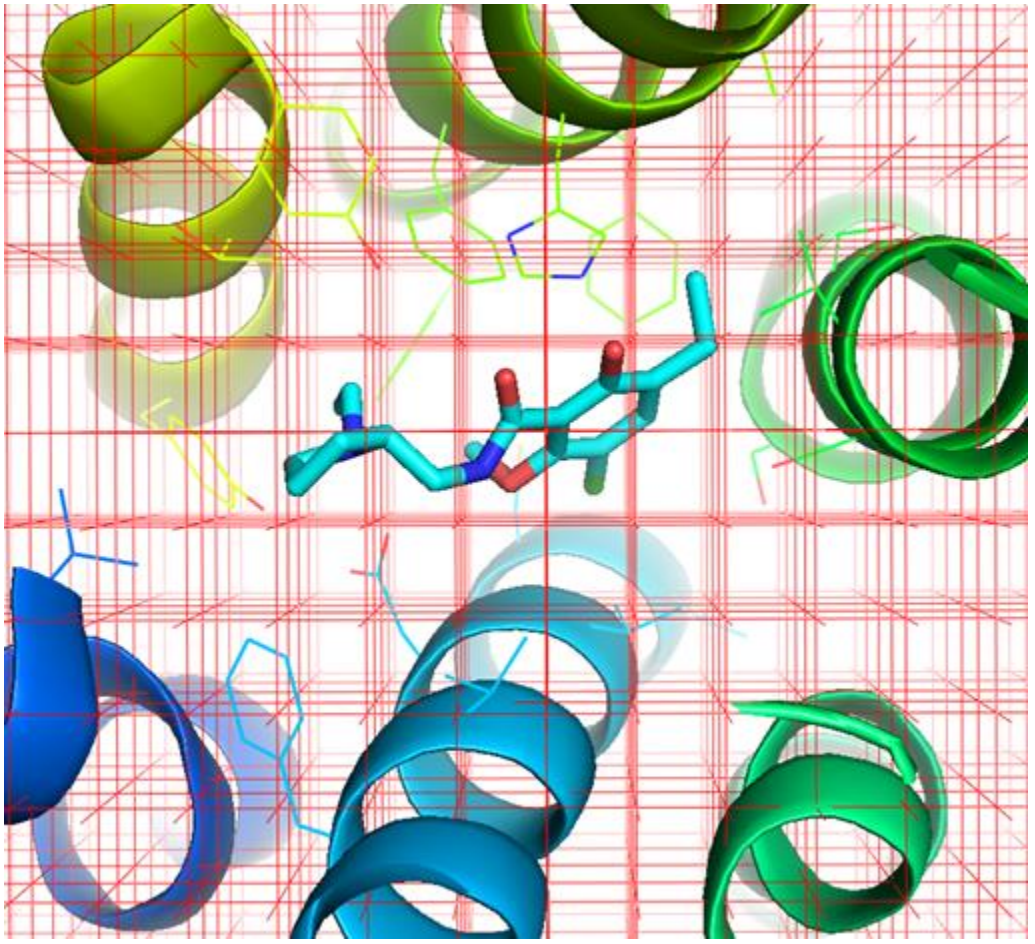
Grid based Monte-Carlo Translation, Rotation, and Conformation Sampling

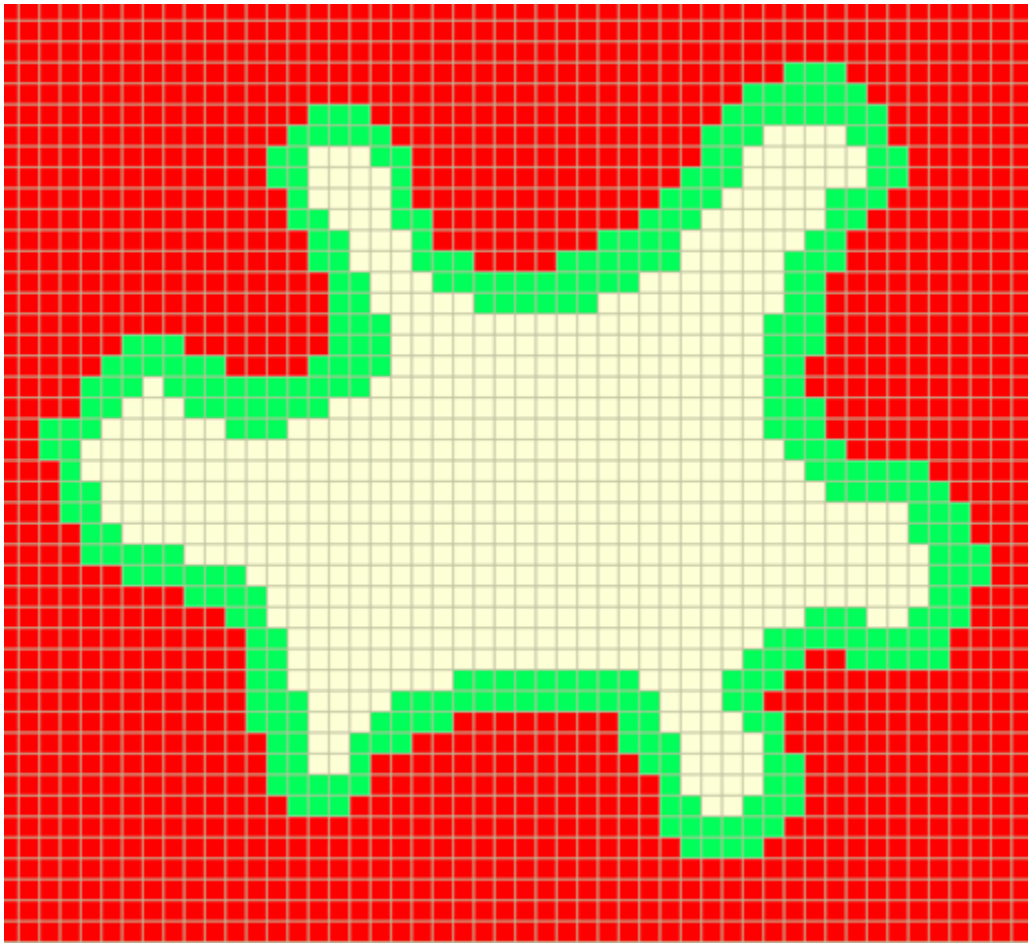


Transform Algorithm



Setting Up The Grid for Transform Step





- Calculates score at each point in grid
- Determines where the ligand can move
  - This only based on sterics

Repulsive

+1 0 -1

Grid Manager

```
<SCORINGGRIDS ligand_chain=(string) width=(float)>
  <classic grid_type="ClassicGrid"weight=(float)/>
  <Another Grid>
</SCORINGGRIDS>
```

Sets up scoring the grid(s) around the ligand

Ligand_chain: Identifies the ligand to build around (Typically X)

Width: Grid width to keep all atoms in

GridType: Type of grid to use, classic grid is shape complementarity

Weight: Scoring weight to use if we have multiple grids

Transform Step

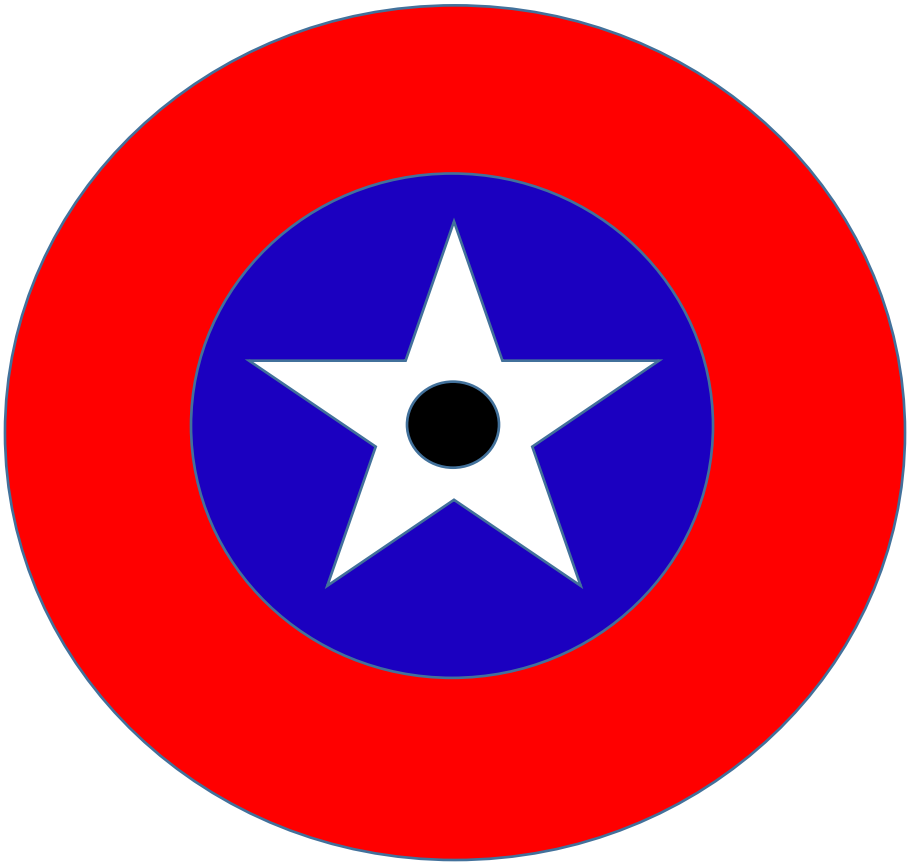
Center: Initial center of ligand that acts as center of grids

Box size: The center of ligand cannot move outside this box

Grid Width: No ligand atoms are allowed outside this box

Translation: Random 3D translation between 0 and move distance (Gaussian distribution)

Rotation: Random 3D rotation between 0 and angle (Gaussian distribution)



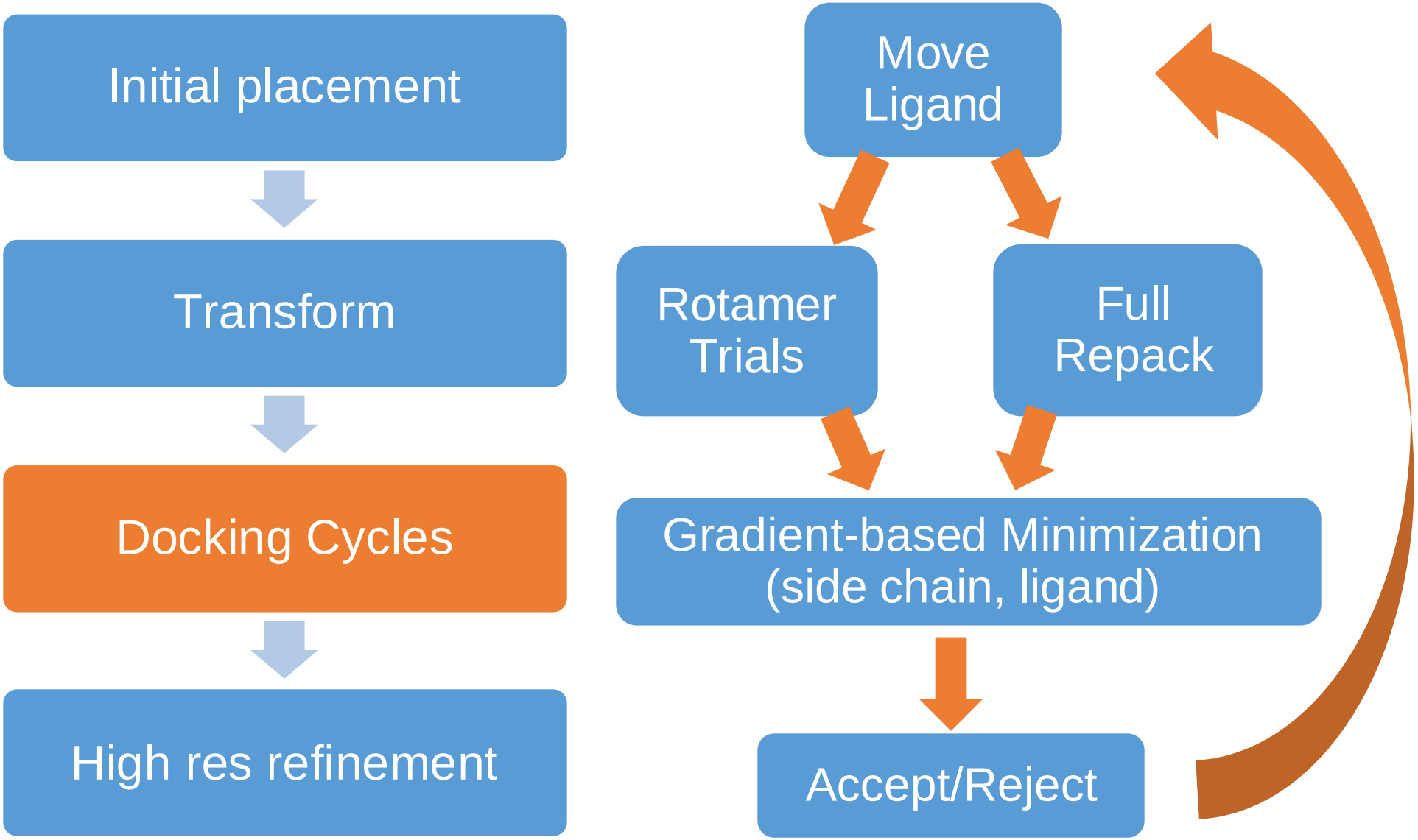
Transform

<Transform name=(string) chain=(string) box_size=(float) move_distance=(float) angle=(float) cycles=(int) repeats=(int) temperature=(float)/>

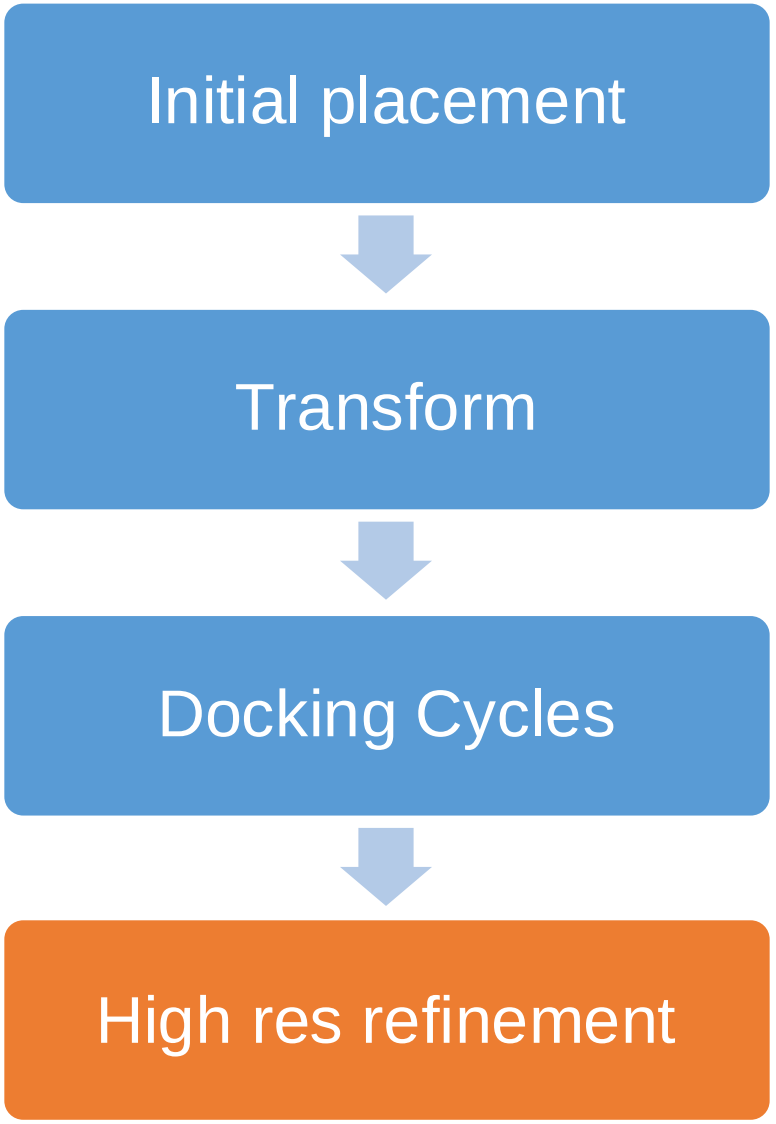
Monte Carlo transformation of ligand in grid

- Chain: Ligand chain to move
- Box_size: Distance from starting position to restrain ligand center
- Move_distance: Max translational distance of a single move
- Angle: Max angular rotation of a single move
- Cycles: Number of Monte Carlo steps to perform
- Repeats: Number of times to perform cycles from starting position
  (EX: 3 repeats of 500 cycles means 1500 total steps)
- Temperature: Factor that controls determining how often an unfavorable move is accepted by Monte Carlo

RosettaLigand Algorithm



RosettaLigand Algorithm



Gradient Based Minimization of side-chain and backbone torsion angles

HighResDocker

<HighResDocker name=(string) cycles=(int)
repack_every_Nth=(int) scorefxn=(string)
movemap_builder=(string) resfile=(string)/>

Atomistic Monte Carlo refinement of ligand and sidechain conformations

cycles: number of rotamer trials or repack steps (coupled w/small ligand movements).

repack_every_Nth: full repacks are interspersed with rotamer trials

scorefxn: Use a soft repulsive score here.

movemap_builder: movemap defined in the LIGAND AREA section

resfile: Specify what rotamer exchanges are possible (including design). These are restricted to the defined interface region. Can be in options instead

FinalMinimizer

<FinalMinimizer name=(string) scorefxn=(string)
movemap_builder=(string)>

Minimize the structure of the docked protein/ligand complex

scorefxn: Use a hard repulsive to remove clashes

movemap_builder: movemap defined in the LIGAND AREA section

Move Maps, Interfaces, and Ligand Areas

LIGAND_AREAS
- cutoff: maximum distance between C-beta atom and ligand to still be part of interface
- add_nbr_radius: accounts for ligand atom size when computing cutoff distance
- all_atom_mode: uses all ligand atoms for identifying nearby residues

INTERFACE_BUILDERS
- extension_window: include “near-interface” residues for additional flexibility during backbone minimization

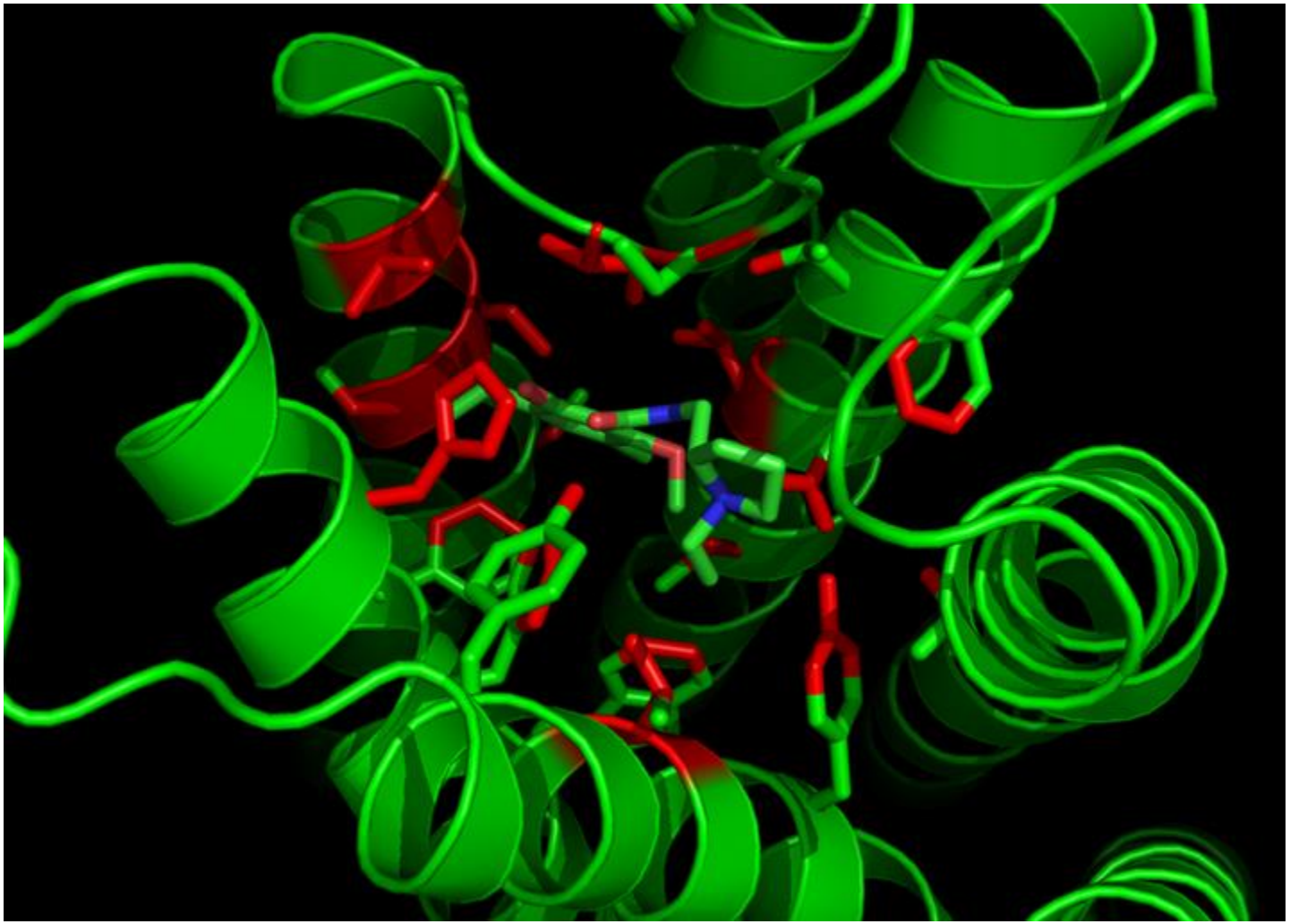
MOVEMAP_BUILDERS
- sc_interface: residues selected for repacking and rotamer trials
- bb_interface: residues selected for backbone ϕ/ψ adjustments

Move Maps, Interfaces, and Ligand Areas

```
<LIGAND_AREAS>
        <inhibitor_dock_sc chain="X" cutoff="6.0" add_nbr_radius="true" all_atom_mode="false"/>
        <inhibitor_final_sc chain="X" cutoff="6.0" add_nbr_radius="true" all_atom_mode="false"/>
        <inhibitor_final_bb chain="X" cutoff="7.0" add_nbr_radius="false" all_atom_mode="true"
                Calpha_restraints="0.3"/>
    </LIGAND_AREAS>
    <INTERFACE_BUILDERS>
        <side_chain_for_docking ligand_areas="inhibitor_dock_sc"/>
        <side_chain_for_final ligand_areas="inhibitor_final_sc"/>
        <backbone ligand_areas="inhibitor_final_bb" extension_window="3"/>
    </INTERFACE_BUILDERS>
    <MOVEMAP_BUILDERS>
        <docking sc_interface="side_chain_for_docking" minimize_water="false"/>
        <final sc_interface="side_chain_for_final" bb_interface=backbone minimize_water="false"/>
    </MOVEMAP_BUILDERS>
```

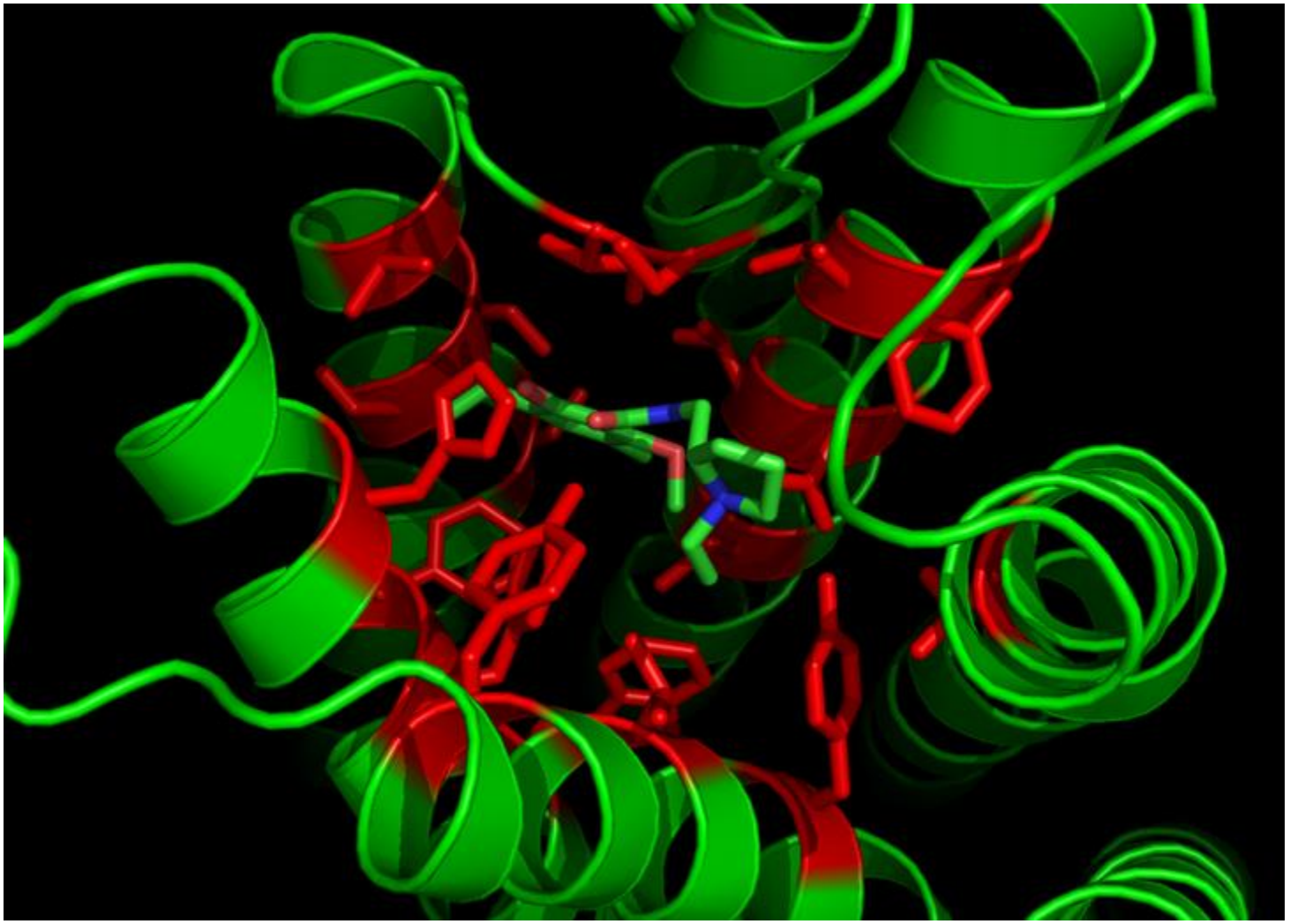
Move Maps, Interfaces, and Ligand Areas

Ligand area: identifies parameters for each ligand and identifies the atoms that are within a cutoff of the ligand



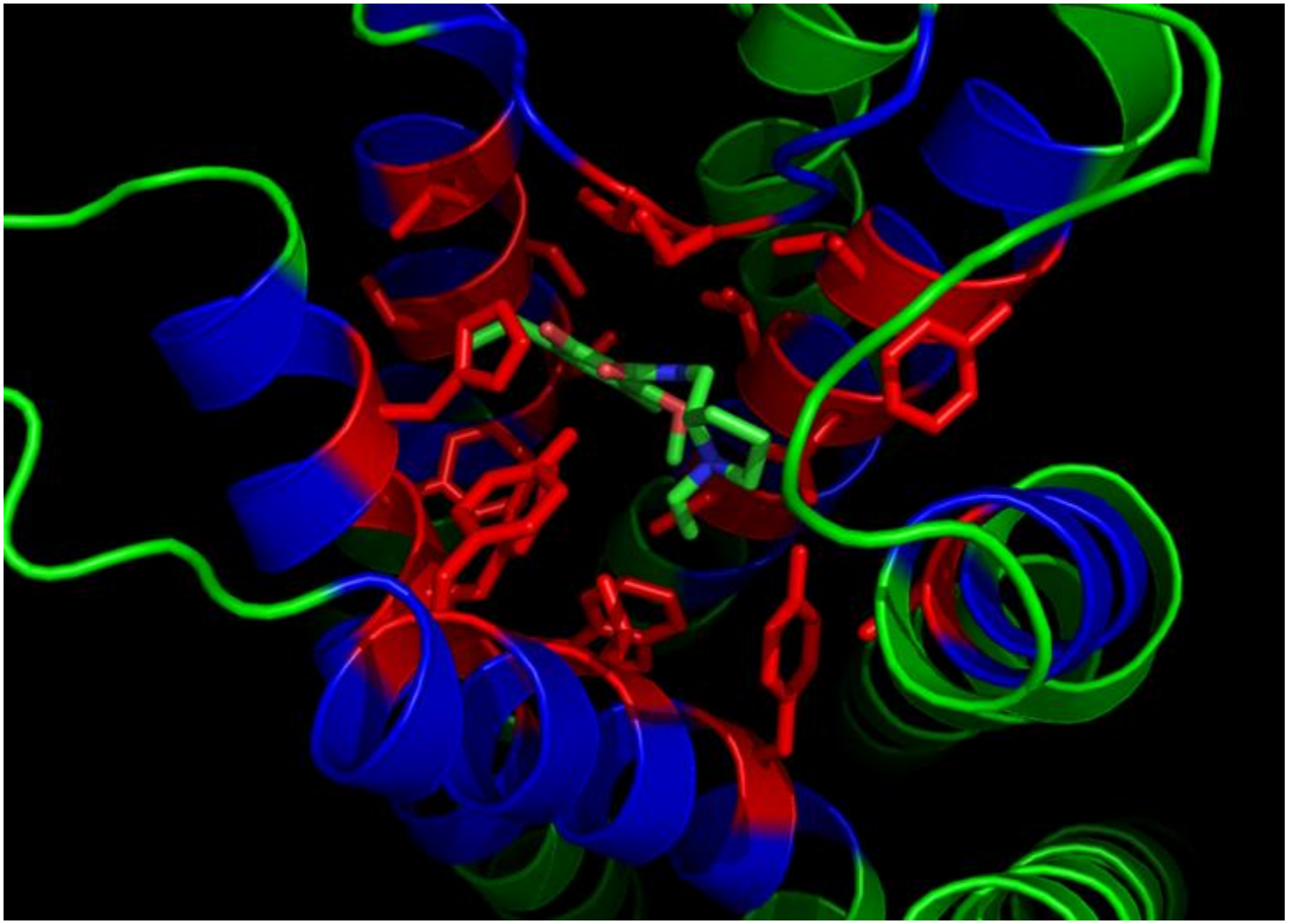
Move Maps, Interfaces, and Ligand Areas

Interface Builder: Extends those atoms to complete the residues



Move Maps, Interfaces, and Ligand Areas

Interface Builder: Extends to include surrounding residues for backbone flexibility



Move Maps, Interfaces, and Ligand Areas

Move map: identifies which residues are allowed to have backbone and side-chain movement. Typically, we have one for docking and one for minimization

Residue	Side-chain Movement	Backbone Movement
VAL 55	No	Yes
VAL 56	Yes	Yes
TYR 57	No	Yes
LEU 58	No	No

Note that these XML scripts defining Move Maps, Interfaces, Ligand Areas, etc. are set when you are running a dock. These have been benchmarked by developers so you won't be changing these.

InterfaceScoreCalculator

<InterfaceScoreCalculator name=(string)
chains=(comma separated chars) scorefxn=(string)
native=(string)/>

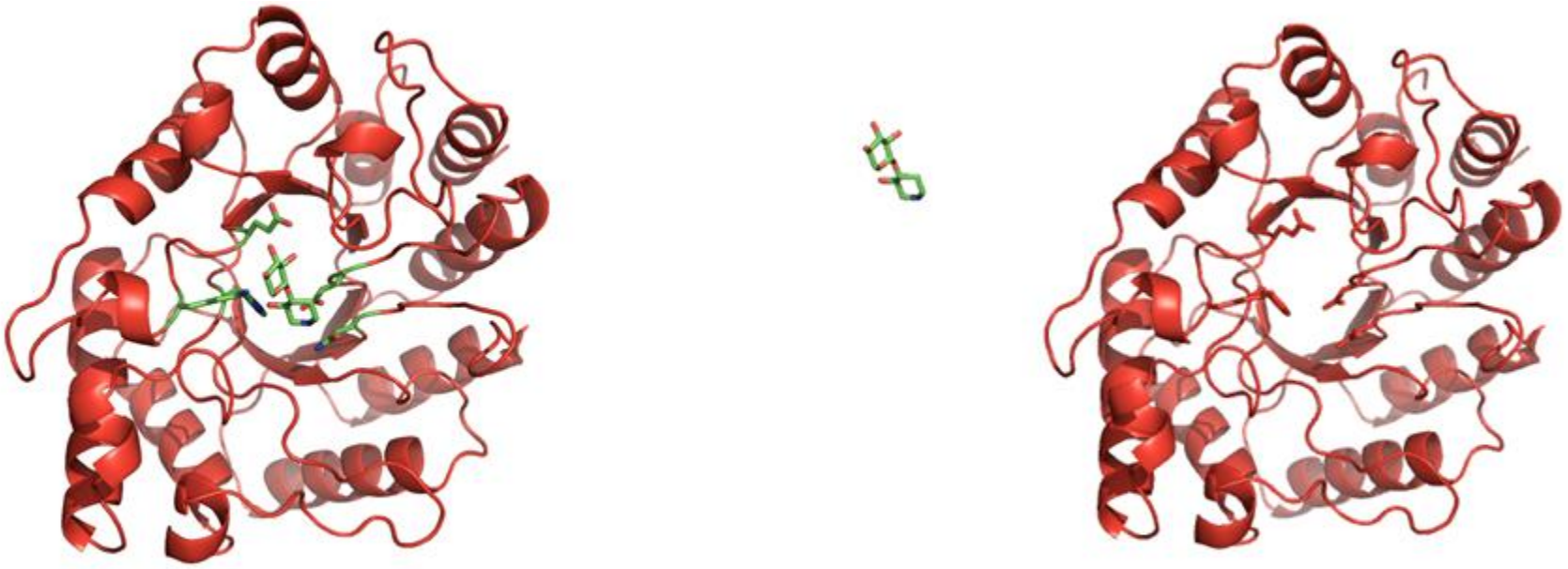
Calculates ligand-specific scores based on bound complex vs separated complex

chains: ligand chains to perform calculation for

scorefxn: Use a hard repulsive to identify clashes

native: native or “correct” structure for rms comparison if available

InterfaceScoreCalculator



interface_delta_X = Score(Bound) - Score(Unbound)

(Note: sidechains do NOT repack here)

FAQs

- How many output poses should I make?
  Depends.
  - Size of active site
  - Ligand--size, rotatable bonds, etc.

- High-throughput screening?
  - Rosetta/tools/hts_tools/
  - (Deluca et al.)

- How long do docking runs usually take?
  - Not long compared to other types of runs

References

Methodology

- **Fu D., Meiler. J. RosettaLigandEnsemble: A Small-Molecule Ensemble-Driven Docking Approach. ACS Omega, 3(4): 3655-3664 (2018).**
- **DeLuca, S., Khar, K. & Meiler, J. Fully Flexible Docking of Medium Sized Ligand Libraries with RosettaLigand. *PLoS One* 10, e0132508 (2015).**
- **S. Combs *et al.*, Small-molecule ligand docking into comparative models with Rosetta, *Nature Protocols* 8, 1277–1298 (2013).**
- G. Lemmon, J. Meiler, R. Baron, Ed. RosettaLigand docking with flexible XML protocols, *Methods in Mol Biol* **819**, 143–155 (2012).
- I. W. Davis, D. Baker, RosettaLigand docking with full ligand and receptor flexibility., *Journal of molecular biology* **385**, 381–92 (2009).
- K.W. Kaufmann *et al.,* Small Molecule Rotamers Enable Simultaneous Optimization of Small Molecule and Protein Degrees of Freedom in ROSETTALIGAND Docking, German Conference on Bioinformatics, pp. 148–157 (2008)
- J. Meiler, D. Baker, ROSETTALIGAND : Protein – Small Molecule Docking with Full Side-Chain Flexibility, *Proteins* **548**, 538–548 (2006).

References

Applications

- K. J. Gregory *et al.*, Probing the metabotropic glutamate receptor 5 (mGlu5) positive allosteric modulator (PAM) binding pocket: discovery of point mutations that engender a "molecular switch" in PAM pharmacology., *Molecular pharmacology* **83**, 991–1006 (2013).
- B. Allison *et al.*, Computational design of protein-small molecule interfaces., *Journal of structural biology* (2013), doi:10.1016/j.jsb.2013.08.003.
- G. Lemmon, J. Meiler, Towards ligand docking including explicit interface water molecules., *PloS one* **8**, e67536 (2013).
- G. Lemmon, K. Kaufmann, J. Meiler, Prediction of HIV-1 protease/inhibitor affinity using RosettaLigand., *Chemical biology & drug design* **79**, 888–96 (2012).
- K. W. Kaufmann, J. Meiler, Using RosettaLigand for small molecule docking into comparative models., *PloS one* **7** (2012), doi:10.1371/journal.pone.0050769.
- K. W. Kaufmann *et al.*, Structural determinants of species-selective substrate recognition in human and Drosophila serotonin transporters revealed through computational docking studies., *Proteins* **74**, 630–42 (2009).
- Regan, J. et al. Structure-activity relationships of the p38alpha MAP Kinase Inhibitor 1-(5-tert-Butyl-2-p-tolyl-2H-pyrazol-3-yl)-3-[4-(2-morpholin-4-yl-ethoxy)naphthalen-1-yl]urea (BIRB 796). *Journal of Medicinal Chemistry* **46**, 4676-4686 (2003).
- https://www.bindingdb.org/validation_sets/index.jsp

Today's Tutorial

Three possible tutorials:

1. Standard ligand docking

- I would highly suggest doing this first to get familiar with how to work with small-molecules in Rosetta

2. RosettaLigandEnsemble (Fu 2018)

- Simultaneous docking of similar ligands

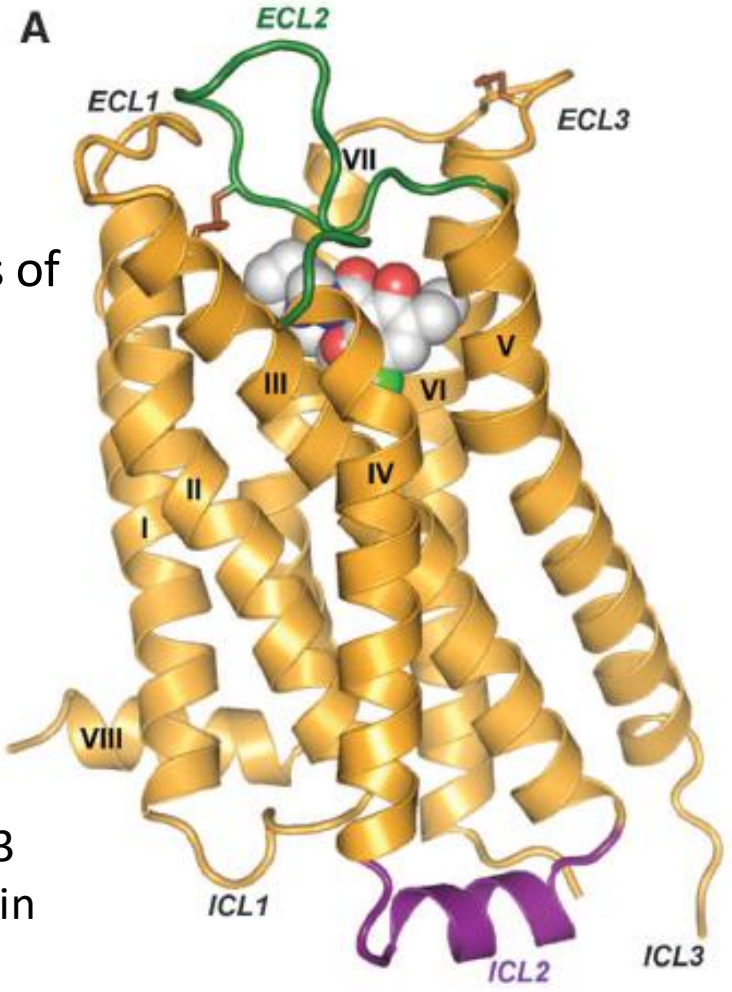
1. Standard protein-ligand docking

- Docking of eticlopride (antagonist) to Dopamine Receptor Subtype 3

Crystal Structure
Chien, E. Y. T. et al. Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. Science 330, 1091–5 (2010)

Rosetta Assessment
Nguyen, E. D. et al. Assessment and challenges of ligand docking into comparative models of g-protein coupled receptors. PLoS One 8, (2013)

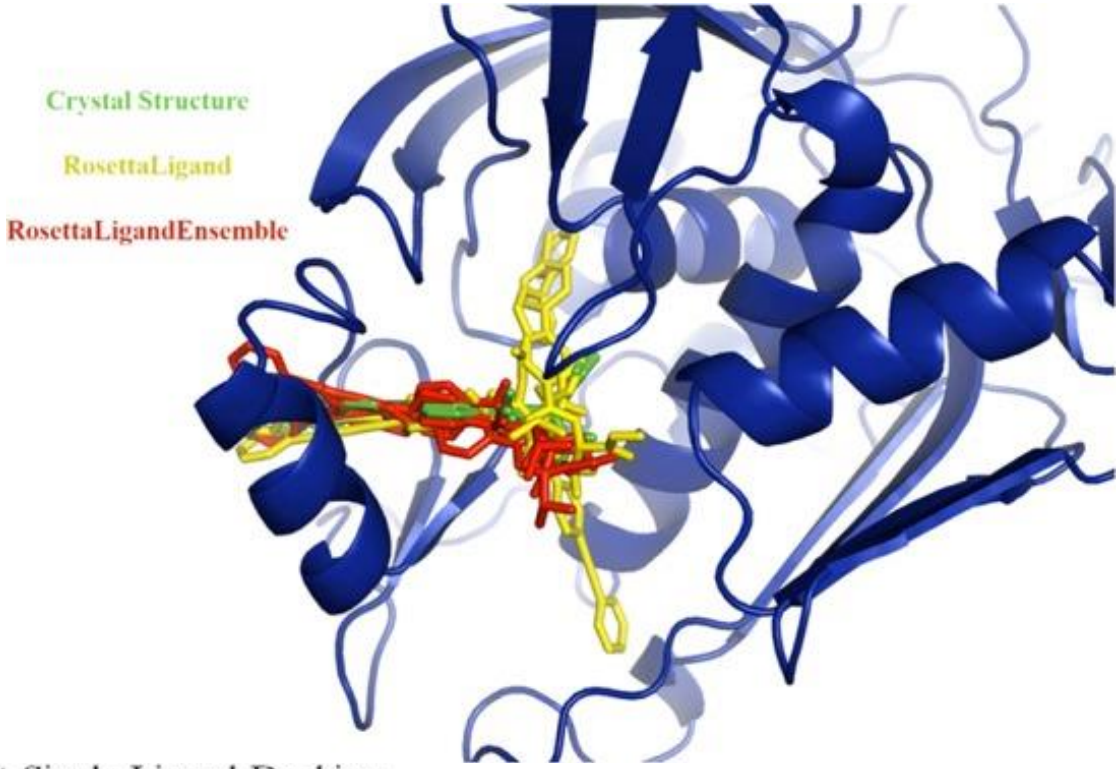
Community Assessment
Kufareva, I. et al. Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment. Structure 19, 1108–1126 (2011).

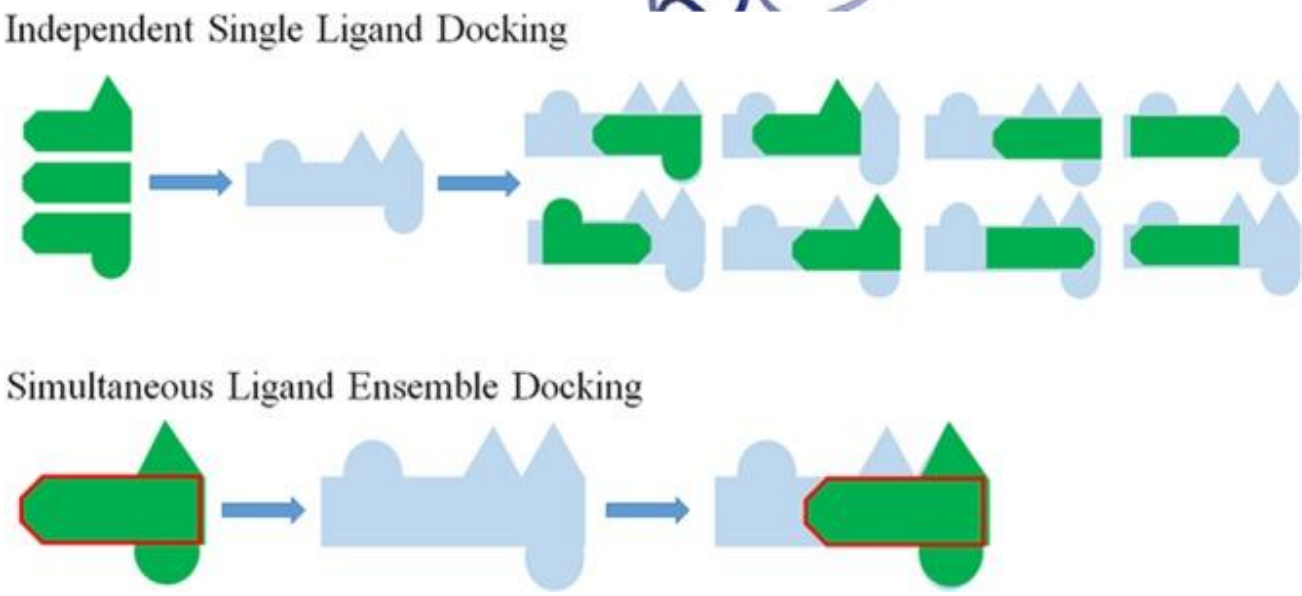


Eticlopride, a D2/D3 selective antagonist in complex with D3 Receptor

2. RosettaLigand Ensemble

- Docking multiple similar ligand simultaneously improves structure prediction in most test cases
- *This works because we assume that similar ligands bind in a similar fashion.*
- Generally used in tandem with SAR studies





Fu, D. Meiler J. (2018)

Let’s get started!

~/rosetta_workshop/ligand_docking/

Start with 1_vanilla_docking

Time permitting, then work on 2_Ensemble_docking.