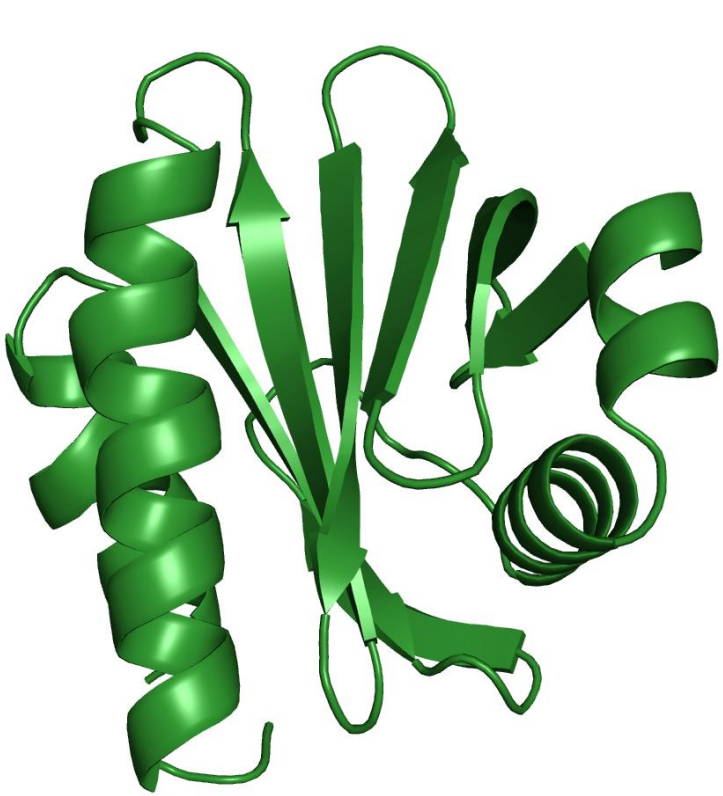
Rosetta Scoring (or Energy) Function

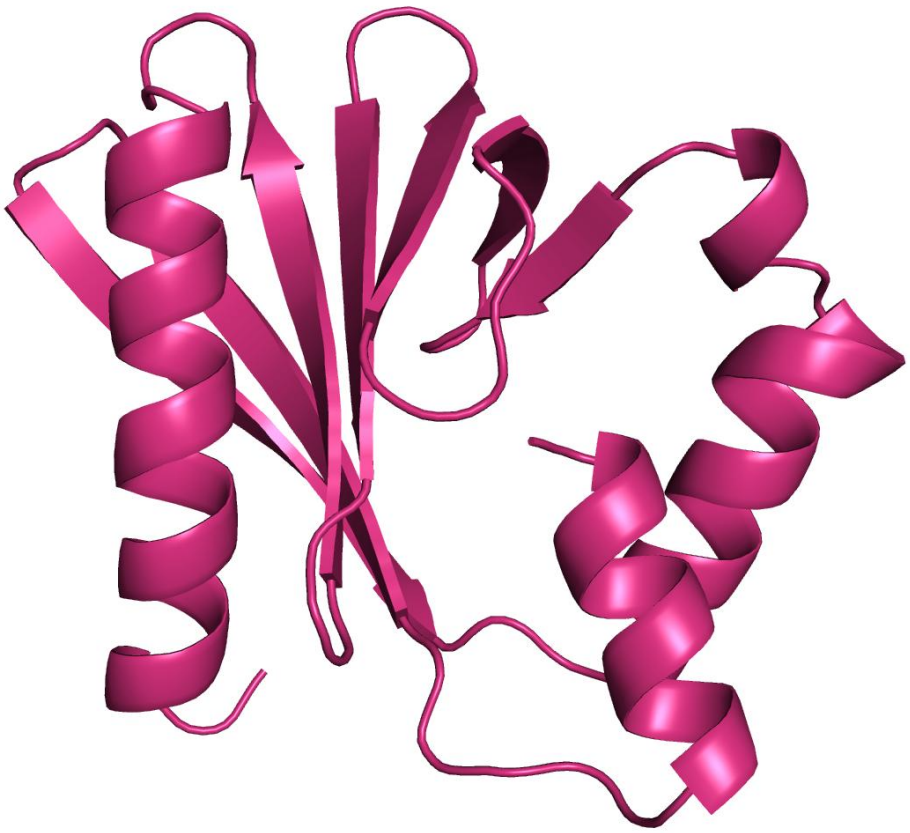
Benjamin K. Mueller
Postdoc, Meiler Lab
Rosetta Workshop
May 2018

Rosetta Scoring (or Energy) Function

… or how models are evaluated in Rosetta

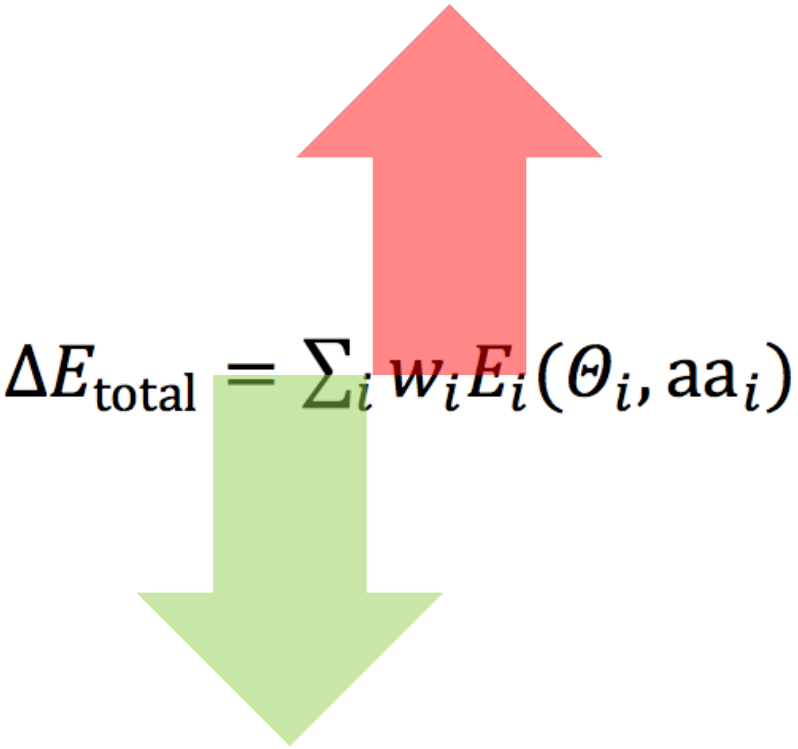
Evaluating Models in Rosetta





An algorithm must assign a quantitative number to a model to decide whether it is “good” or “bad”

Important Note

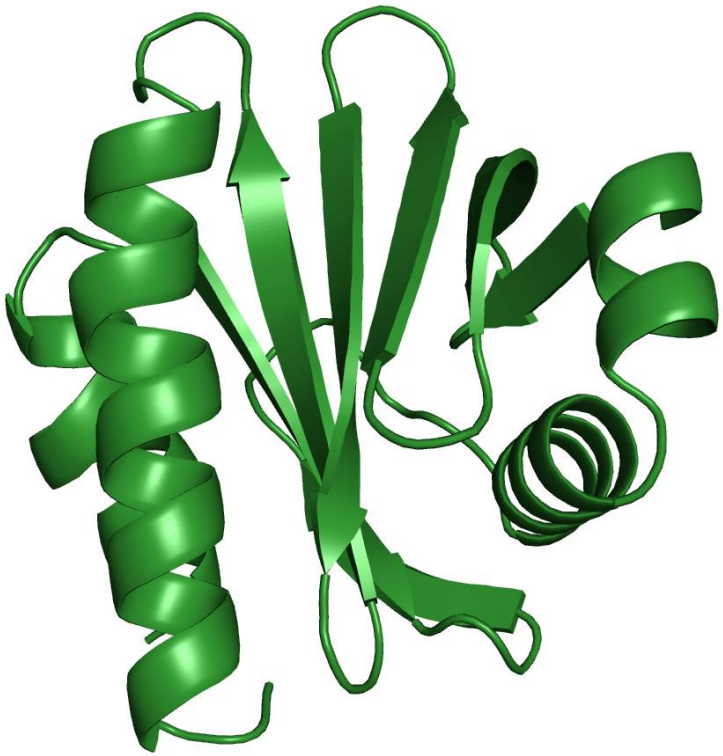


Energy is currently given in: "kcals/mol"

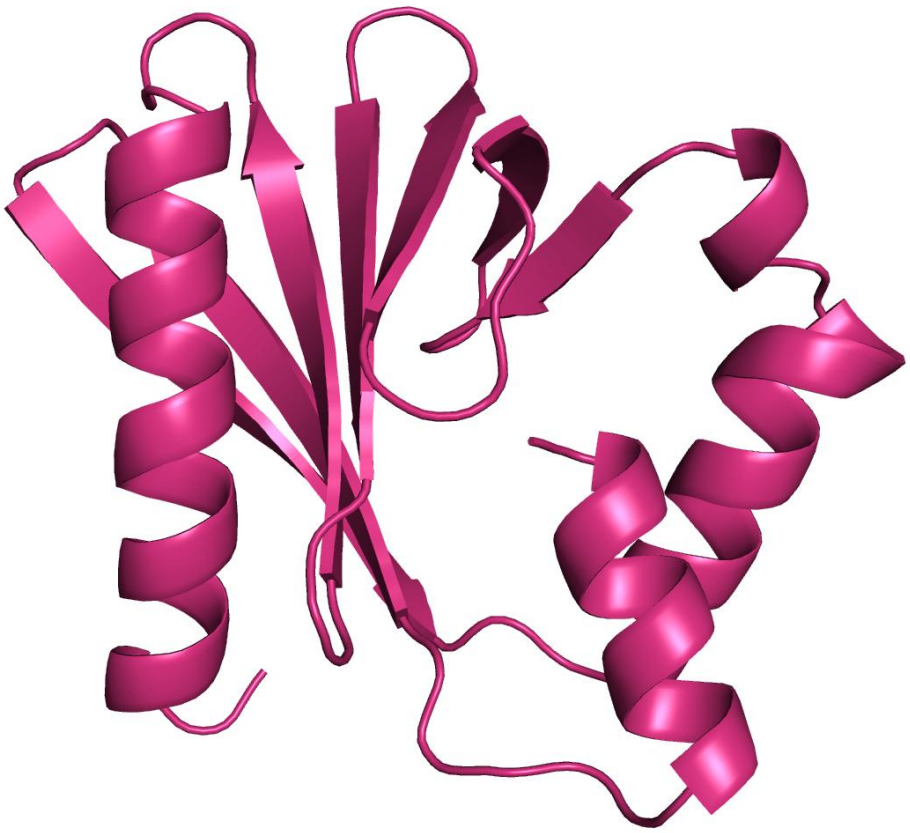
Previously in: "REU"

High Energy = Bad
Low Energy = Good

Evaluating Models in Rosetta

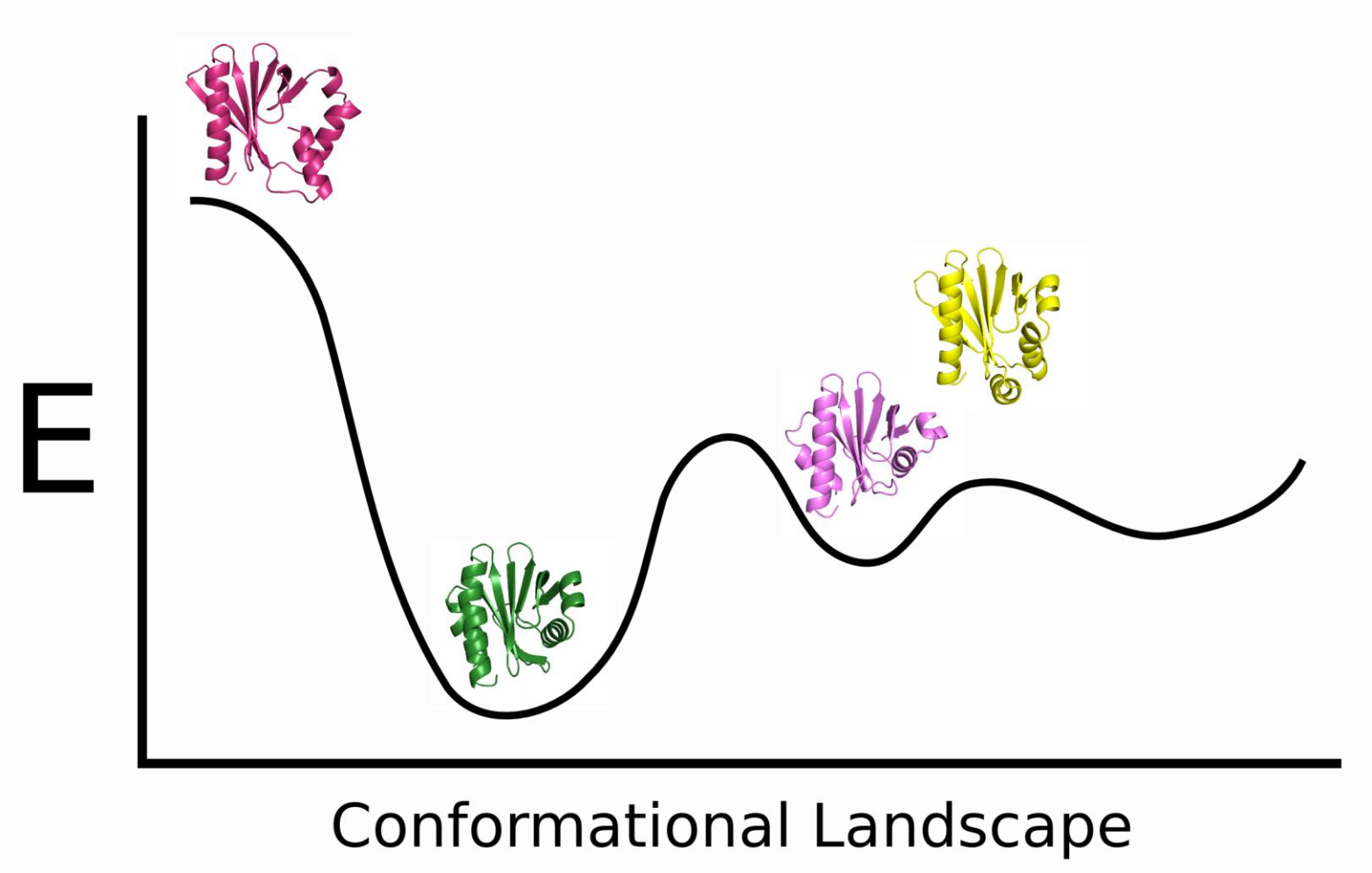


-400 kcals/mol



-350 kcals/mol

Score is Central to Monte Carlo Selection

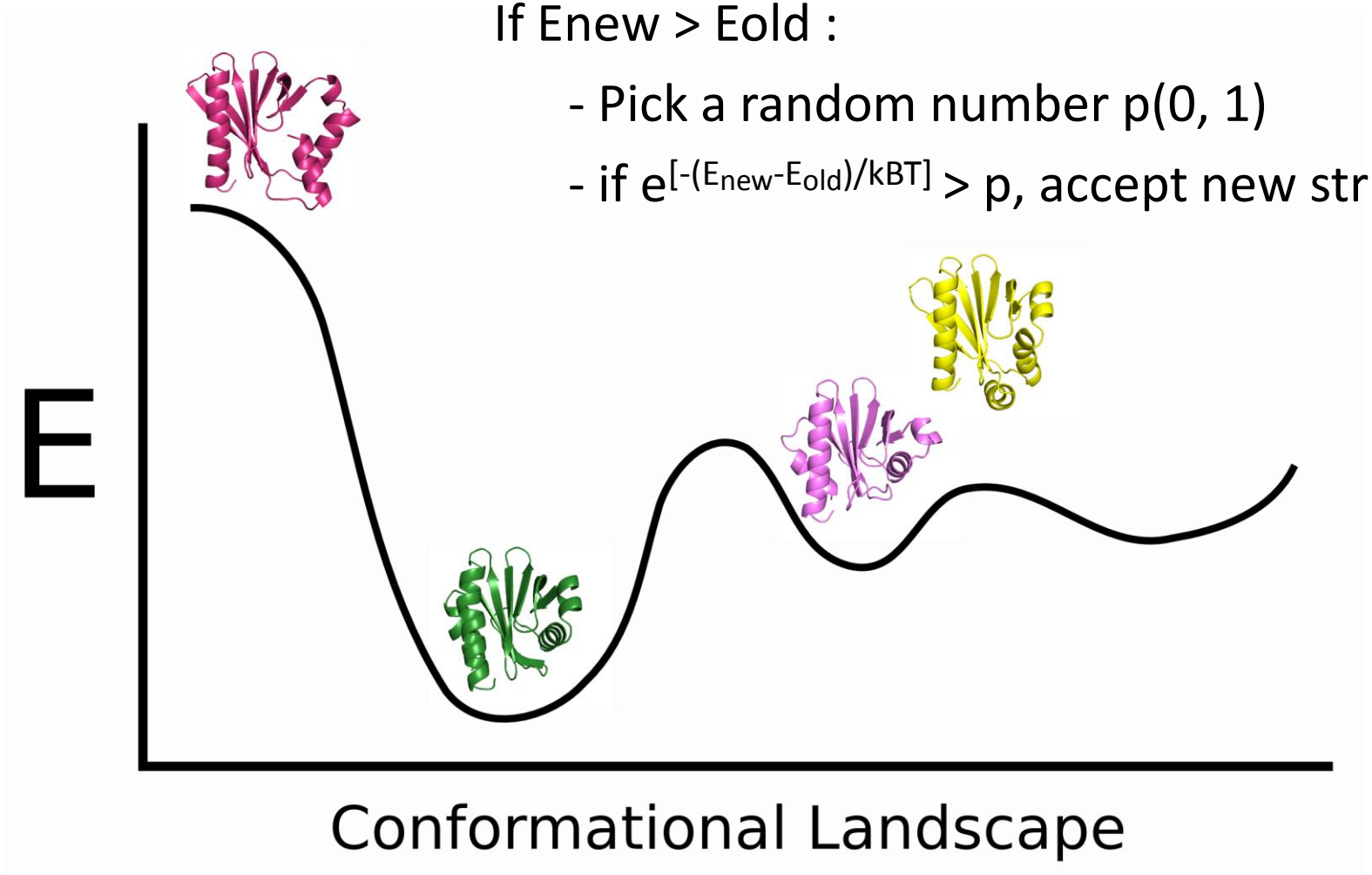


Metropolis Criterion

If Enew < Eold : Accept new structure

If Enew > Eold :

- Pick a random number p(0, 1)
- if $e^{[-(E_{new}-E_{old})/kBT]} > p$, accept new structure

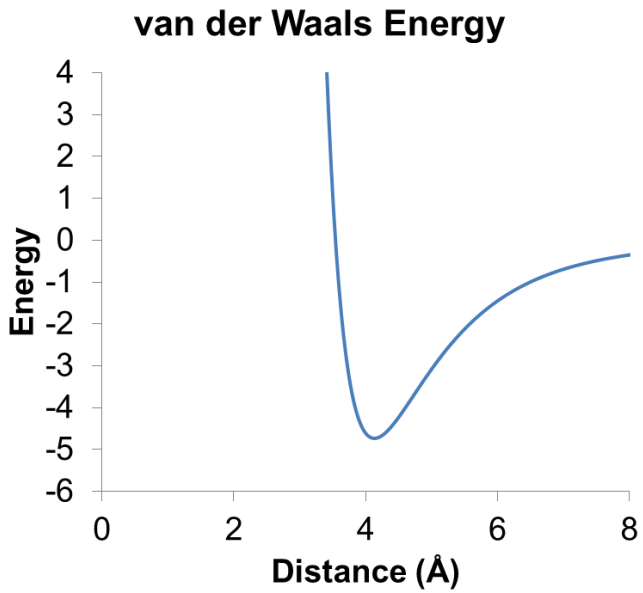


Rosetta Combines Physics-Based and Knowledge-Based Potentials to Build the Energy Function

Rosetta Combines Physics-Based and Knowledge-Based Potentials to Build the Energy Function

Lennard-Jones Potential

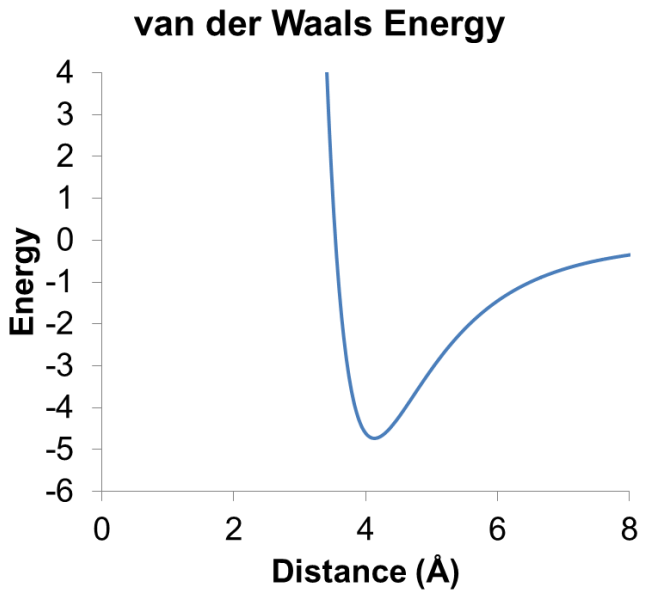
$$\sum_{i<j}\left[\frac{A_{ij}}{r_{ij}^{12}}-\frac{B_{ij}}{r_{ij}^{6}}\right]$$



Rosetta Combines Physics-Based and Knowledge-Based Potentials to Build the Energy Function

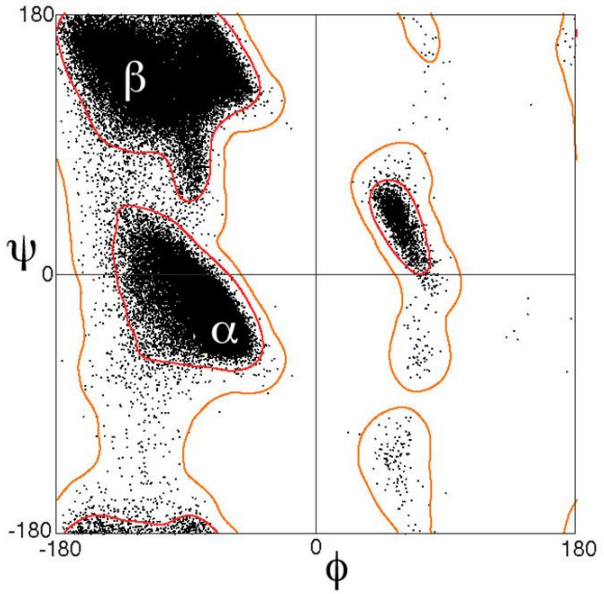
Lennard-Jones Potential

$$\sum_{i<j}\left[\frac{A_{ij}}{r_{ij}^{12}}-\frac{B_{ij}}{r_{ij}^{6}}\right]$$



Ramachandran Plot

Statistical mining of
Protein Databank (PDB)



The Score Function is a Weighted Linear Combination of Individual Score Terms

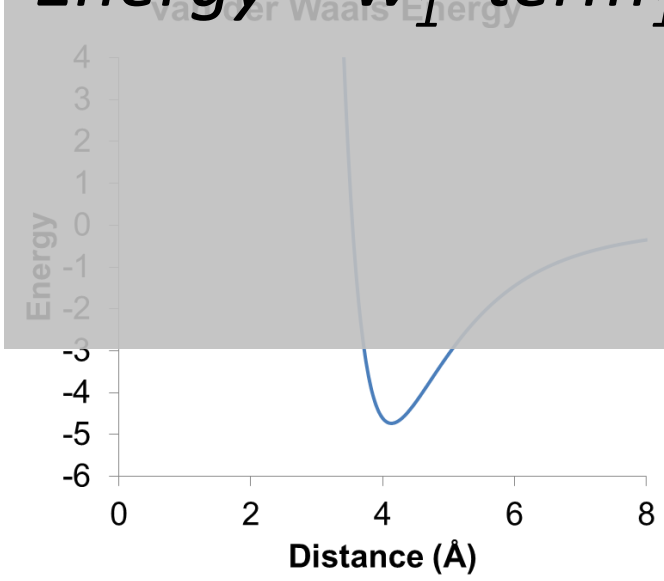
Lennard-Jones Potential

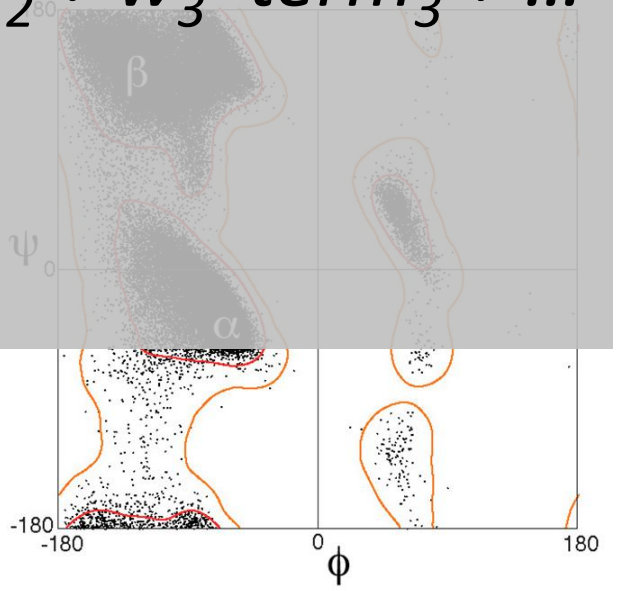
$$\sum_{i<j}\left[\frac{A_{ij}}{r_{ij}^{12}}-\frac{B_{ij}}{r_{ij}^{6}}\right]$$

Ramachandran Plot

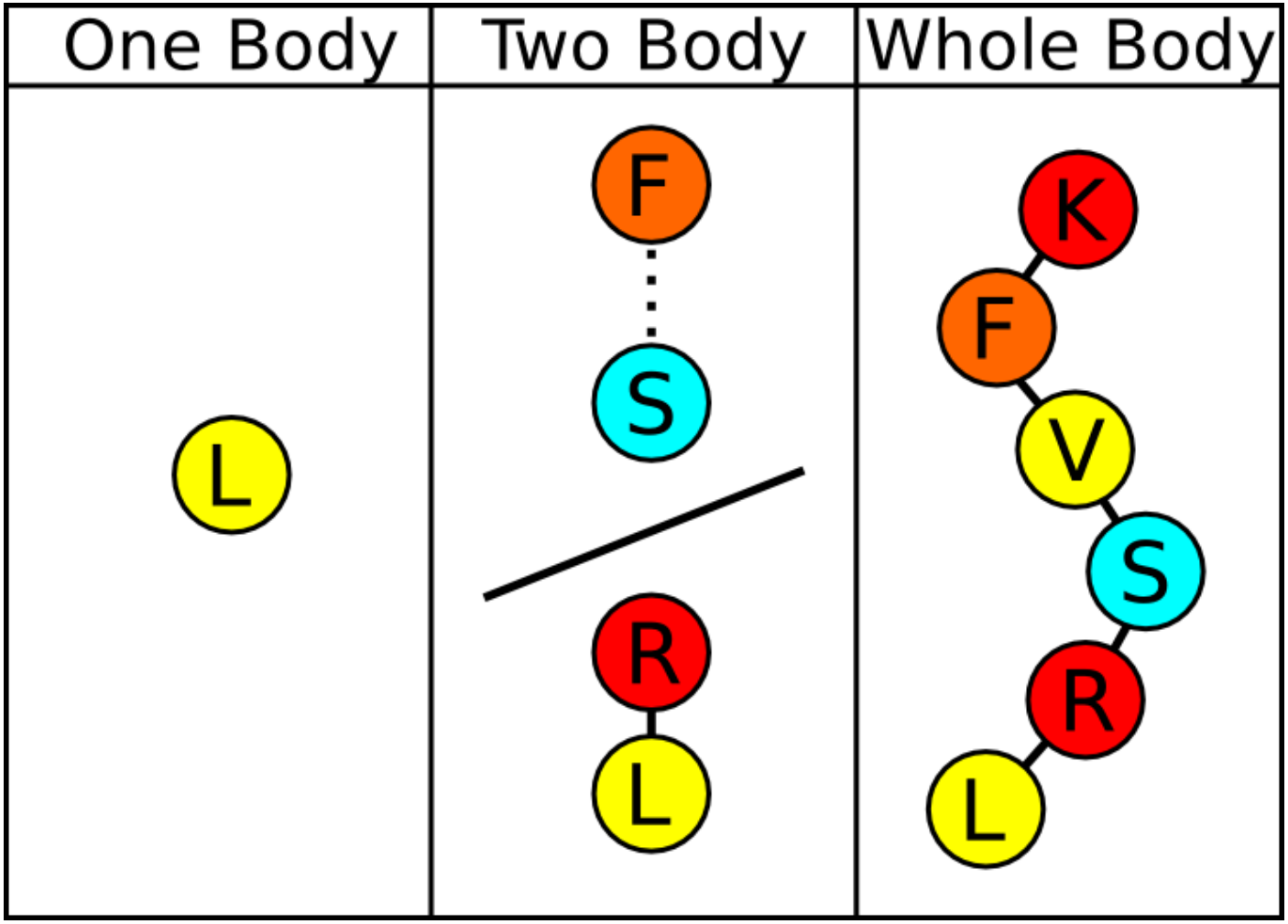
Statistical mining of Protein Databank (PDB)

$$Energy = w_1*term_1 + w_2*term_2 + w_3*term_3 + \ldots$$





Rosetta is a Residue-Centric Scoring Function



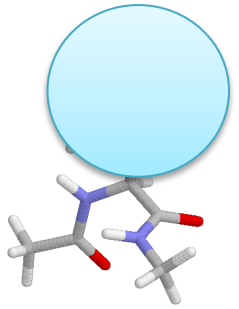
Rosetta is a Residue-Centric Scoring Function

One Body	Two Body	Whole Body
Backbone - p_aa_pp - rama_prepro	Lennard-Jones - fa_atr - fa_rep	Radius of Gyration - rg
Side Chain - fa_dun - yhh_planarity	Solvation - fa_sol	Contact Order - co
Reference - ref	Hydrogen Bond - hbond_lr_bb - hbond_sr_bb - hbond_bb_sc - hbond_sc	Structure Alignment - hs_pair - ss_pair - sheet

* NOTE: Not all score terms are listed here

Rosetta has 2 score function modes

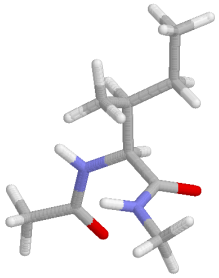
Low resolution (or centroid) mode:



Reduced atom representation (centroid)
Simple energy function
Aggressively search conformational space

`database/chemical/residue_type_sets/centroid`

High resolution (or full atom) mode:



Full atom (FA)
More sophisticated energy function
"Local" search of conformational (and sequence) space

`database/chemical/residue_type_sets/fa_standard`

Breakdown of Full Atom Score Terms – REF2015

term	description	weight	units	ref(s)
fa_atr	attractive energy between two atoms on different residues separated by a distance d	1.0	kcal/mol	5, 6
fa_rep	repulsive energy between two atoms on different residues separated by a distance d	0.55	kcal/mol	5, 6
fa_intra_rep	repulsive energy between two atoms on the same residue separated by a distance d	0.005	kcal/mol	5, 6
fa_sol	Gaussian exclusion implicit solvation energy between protein atoms in different residues	1.0	kcal/mol	36
lk_ball_wtd	orientation-dependent solvation of polar atoms assuming ideal water geometry	1.0	kcal/mol	50, 71
fa_intra_sol	Gaussian exclusion implicit solvation energy between protein atoms in the same residue	1.0	kcal/mol	36
fa_elec	energy of interaction between two nonbonded charged atoms separated by a distance d	1.0	kcal/mol	50
hbond_lr_bb	energy of short-range hydrogen bonds	1.0	kcal/mol	38, 49
hbond_sr_bb	energy of long-range hydrogen bonds	1.0	kcal/mol	38, 49
hbond_bb_sc	energy of backbone–side-chain hydrogen bonds	1.0	kcal/mol	38, 49
hbond_sc	energy of side-chain–side-chain hydrogen bonds	1.0	kcal/mol	38, 49
dslf_fa13	energy of disulfide bridges	1.25	kcal/mol	49
rama_prepro	probability of backbone ϕ, ψ angles given the amino acid type	(0.45 kcal/mol)/kT	kT	50, 51
p_aa_pp	probability of amino acid identity given backbone ϕ, ψ angles	(0.4 kcal/mol)/kT	kT	51
fa_dun	probability that a chosen rotamer is native-like given backbone ϕ, ψ angles	(0.7 kcal/mol)/kT	kT	52
omega	backbone-dependent penalty for cis ω dihedrals that deviate from 0° and trans ω dihedrals that deviate from 180°	(0.6 kcal/mol)/AU	AU[a]	72
pro_close	penalty for an open proline ring and proline ω bonding energy	(1.25 kcal/mol)/AU	AU	51
yhh_planarity	sinusoidal penalty for nonplanar tyrosine χ_3 dihedral angle	(0.625 kcal/mol)/AU	AU	49
ref	reference energies for amino acid types	(1.0 kcal/mol)/AU	AU	1, 51

[a]AU = arbitrary units.

Alford, et. al JCTC 2017

What you'll actually see in your output…

Output Score Table

Found in output score table (score.sc) and at the end of every output pdb (S_0001.pdb)

```
# All scores below are weighted scores, not raw scores.
#BEGIN_POSE_ENERGIES_TABLE dock_01_girk_1212_jon_ML297_0500.pdb
label fa_atr fa_rep fa_sol fa_intra_rep fa_elec total
weights 0.8 0.4 0.6 0.004 0.42 NA
pose -5400.73 638.748 2428.12 13.5145 -302.036 -4267.34
GLN:NtermProteinFull_1 -1.02843 0.12113 0.83426 0.00595 -0.05882 0.54065
ARG_2 -5.46203 0.44148 4.02307 0.01303 -1.13134 -3.04715
PHE_3 -5.25196 0.29047 1.51945 0.02354 -0.38757 -1.99927
VAL_4 -4.7212 0.38219 2.30519 0.01101 -0.33608 -4.69107
ASP_5 -6.18239 0.68119 4.3013 0.02667 -1.21805 -4.77894
LYS_6 -3.70129 0.51352 2.90682 0.00832 -0.90685 -3.25917
ASN_7 -2.97985 0.42136 1.9979 0.00362 -0.17245 -2.08876
GLY_8 -2.75353 0.28458 1.44949 1e-05 -0.20107 -4.00172
ARG:CtermProteinFull_1294 -2.34298 0.24806 1.73197 0.01216 -0.06632 -0.34395
```

Output Score Table

Score Terms:

```
# All scores below are weighted scores, not raw scores.
#BEGIN_POSE_ENERGIES_TABLE dock_01_girk_1212_jon_ML297_0500.pdb
label fa_atr fa_rep fa_sol fa_intra_rep fa_elec total
weights 0.8 0.4 0.6 0.004 0.42 NA
pose -5400.73 638.748 2428.12 13.5145 -302.036 -4267.34
GLN:NtermProteinFull_1 -1.02843 0.12113 0.83426 0.00595 -0.05882 0.54065
ARG_2 -5.46203 0.44148 4.02307 0.01303 -1.13134 -3.04715
PHE_3 -5.25196 0.29047 1.51945 0.02354 -0.38757 -1.99927
VAL_4 -4.7212 0.38219 2.30519 0.01101 -0.33608 -4.69107
ASP_5 -6.18239 0.68119 4.3013 0.02667 -1.21805 -4.77894
LYS_6 -3.70129 0.51352 2.90682 0.00832 -0.90685 -3.25917
ASN_7 -2.97985 0.42136 1.9979 0.00362 -0.17245 -2.08876
GLY_8 -2.75353 0.28458 1.44949 1e-05 -0.20107 -4.00172
ARG:CtermProteinFull_1294 -2.34298 0.24806 1.73197 0.01216 -0.06632 -0.34395
```

Output Score Table

Weights:

```
# All scores below are weighted scores, not raw scores.
#BEGIN_POSE_ENERGIES_TABLE dock_01_girk_1212_jon_ML297_0500.pdb
label fa_atr fa_rep fa_sol fa_intra_rep fa_elec total
weights 0.8 0.4 0.6 0.004 0.42 NA
pose -5400.73 638.748 2428.12 13.5145 -302.036 -4267.34
GLN:NtermProteinFull_1 -1.02843 0.12113 0.83426 0.00595 -0.05882 0.54065
ARG_2 -5.46203 0.44148 4.02307 0.01303 -1.13134 -3.04715
PHE_3 -5.25196 0.29047 1.51945 0.02354 -0.38757 -1.99927
VAL_4 -4.7212 0.38219 2.30519 0.01101 -0.33608 -4.69107
ASP_5 -6.18239 0.68119 4.3013 0.02667 -1.21805 -4.77894
LYS_6 -3.70129 0.51352 2.90682 0.00832 -0.90685 -3.25917
ASN_7 -2.97985 0.42136 1.9979 0.00362 -0.17245 -2.08876
GLY_8 -2.75353 0.28458 1.44949 1e-05 -0.20107 -4.00172
ARG:CtermProteinFull_1294 -2.34298 0.24806 1.73197 0.01216 -0.06632 -0.34395
```

Output Score Table

Scores for individual terms:

```
# All scores below are weighted scores, not raw scores.
#BEGIN_POSE_ENERGIES_TABLE dock_01_girk_1212_jon_ML297_0500.pdb
label fa_atr fa_rep fa_sol fa_intra_rep fa_elec total
weights 0.8 0.4 0.6 0.004 0.42 NA
pose -5400.73 638.748 2428.12 13.5145 -302.030 -4267.34
GLN:NtermProteinFull_1 -1.02843 0.12113 0.83426 0.00595 -0.05882 0.54065
ARG_2 -5.46203 0.44148 4.02307 0.01303 -1.15157 -3.04715
PHE_3 -5.25196 0.29047 1.51945 0.02354 -0.38757 -1.99927
VAL_4 -4.7212 0.38219 2.30519 0.01101 -0.33608 -4.69107
ASP_5 -6.18239 0.68119 4.3013 0.02667 -1.21805 -4.77894
LYS_6 -3.70129 0.51352 2.90682 0.00832 -0.90685 -3.25917
ASN_7 -2.97985 0.42136 1.9979 0.00362 -0.17245 -2.08876
GLY_8 -2.75353 0.28458 1.44949 1e-05 -0.20107 -4.00172
ARG:CtermProteinFull_1294 -2.34298 0.24806 1.73197 0.01216 -0.06632 -0.34395
```

Total Score of Protein

Output Score Table

Scores by amino acid position:

```
# All scores below are weighted scores, not raw scores.
#BEGIN_POSE_ENERGIES_TABLE dock_01_girk_1212_jon_ML297_0500.pdb
label fa_atr fa_rep fa_sol fa_intra_rep fa_elec total
weights 0.8 0.4 0.6 0.004 0.42 NA
pose -5400.73 638.748 2428.12 13.5145 -302.036 -4267.34
GLN:NtermProteinFull_1 -1.02843 0.12113 0.83426 0.00595 -0.05882 0.54065
ARG_2 -5.46203 0.44148 4.02307 0.01303 -1.13134 -3.04715
PHE_3 -5.25196 0.29047 1.51945 0.02354 -0.38757 -1.99927
VAL_4 -4.7212 0.38219 2.30519 0.01101 -0.33608 -4.69107
ASP_5 -6.18239 0.68119 4.3013 0.02667 -1.21805 -4.77894
LYS_6 -3.70129 0.51352 2.90682 0.00832 -0.90685 -3.25917
ASN_7 -2.97985 0.42136 1.9979 0.00362 -0.17245 -2.08876
GLY_8 -2.75353 0.28458 1.44949 1e-05 -0.20107 -4.00172
ARG:CtermProteinFull_1294 -2.34298 0.24806 1.73197 0.01216 -0.06632 -0.34395
```

More score term information

Constraints (actually “restraints”)

Supplements energy function with additional information

- Commonly from experimental information

Various types:

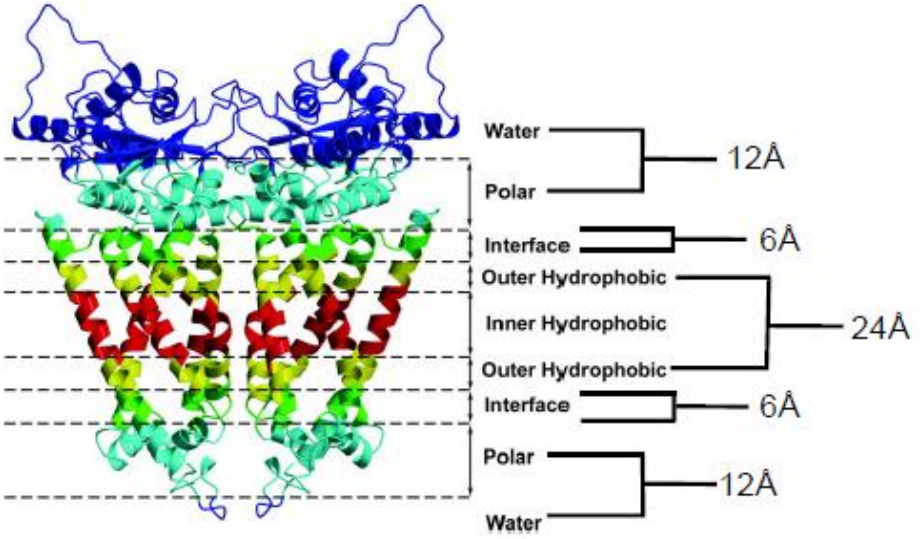
atom_pair_constraint, dihedral_constraint, angle_constraint, coordinate_constraint, residue_type_constraint …

Interface with Rosetta using constraint files

Additional score terms

Membrane terms:

- fa_mbsolv
- fa_mbenv



Yarov-Yarovoy, Schonbrun, and Baker 2006

Over 100+ score terms

Most are turned off (weight is set to 0), not in REF2015

To turn on score term, set weight to non-zero value

Modifying Scorefunctions in Rosetta

Modifying using the command line

1 - score:weights <filename>

2 - score:set_weight <scoreterm$_1$> <wt$_1$> <scoreterm$_2$> <wt$_2$>

3 - score:patch <patchfile>

Patchfile example

```
fa_atr = 0.423
fa_rep = 0.100
```

Modifying Scorefunctions in Rosetta

```
<ROSETTASCRIPTS>
    <SCOREFXNS>
        <ScoreFunction name="ligand_soft_rep" weights="ligand_soft_rep">
            <Reweight scoretype="fa_elec" weight="0.42"/>
        </ScoreFunction>
        <ScoreFunction name="hard_rep" weights="ligandprime">
        </ScoreFunction>
    </SCOREFXNS>
    <OUTPUT scorefxn="hard_rep" />
</ROSETTASCRIPTS>
```

XML script options

- Weights filename or path to file
- Reweight specific terms as needed (as in patch file)
- Must include top-level output tag to ensure proper scoring in output files

References:

REF2015 References:

Alford RF, et. al, **The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design**. *Journal of Chemical Theory and Computation,* **2017.** *13* (6), 3031-3048

Park H, et. al **Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules** *Journal of Chemical Theory and Computation,* **2016.** *12* (12), 6201-6212

Old Scorefxn (Talaris) References:

O'Meara MJ, et. al, **A Combined Covalent-Electrostatic Model of Hydrogen Bonding Improves Structure Prediction with Rosetta.** *Journal of Chemical Theory and Computation*, **2015**.

Leaver-Fay A, et. al **Scientific benchmarks for guiding macromolecular energy function improvement.** *Methods in enzymology*, **2013**. 523: p. 109.

Useful Links

https://www.rosettacommons.org/docs/latest/rosetta_basics/scoring/score-types

https://www.rosettacommons.org/docs/latest/rosetta_basics/scoring/scoring-explained