

Rosetta Scoring (or Energy) Function

Benjamin K. Mueller

Postdoc, Meiler Lab

Rosetta Workshop

December 2017

Rosetta Scoring (or Energy) Function

... or how models are evaluated in
Rosetta

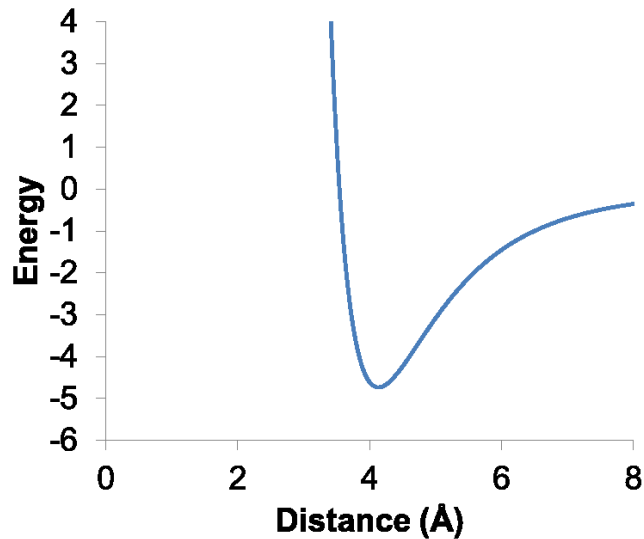
Rosetta Combines Physics-Based and Knowledge-Based Potentials to Build the Energy Function

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

Lennard-Jones Potential

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

van der Waals Energy



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

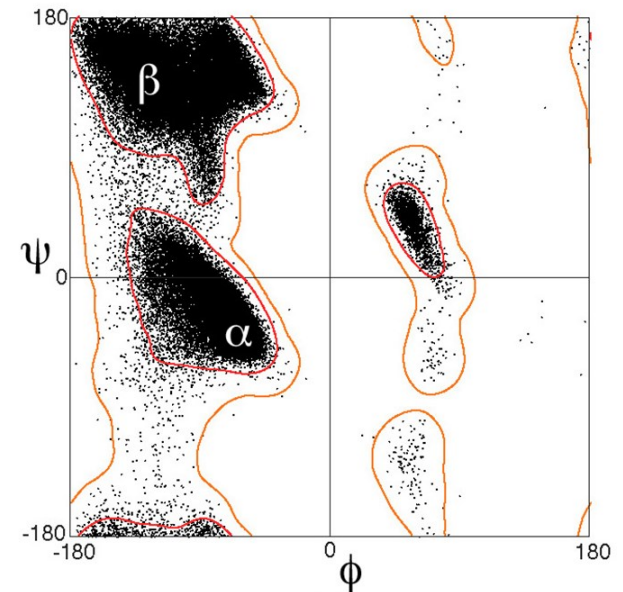
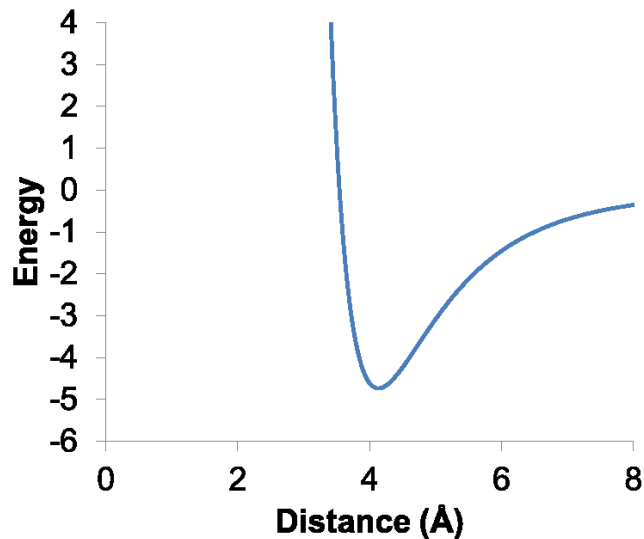
Lennard-Jones Potential

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

Ramachandran Plot

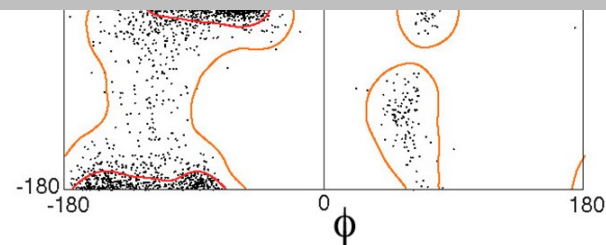
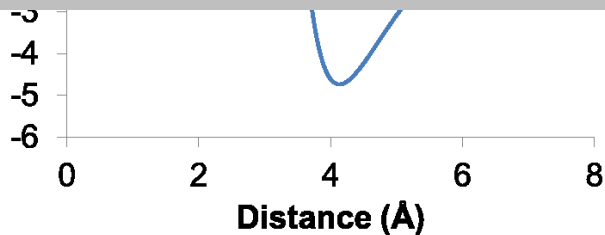
Statistical mining of
Protein Databank (PDB)

van der Waals Energy

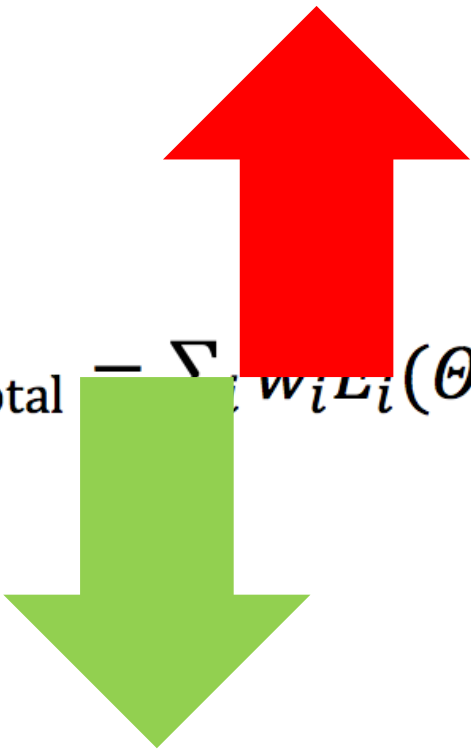


The Score Function is a Weighted Linear Combination of Individual Score Terms

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



Important Note

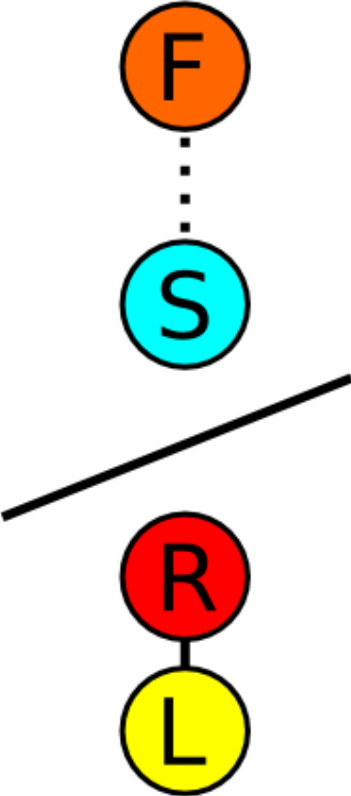
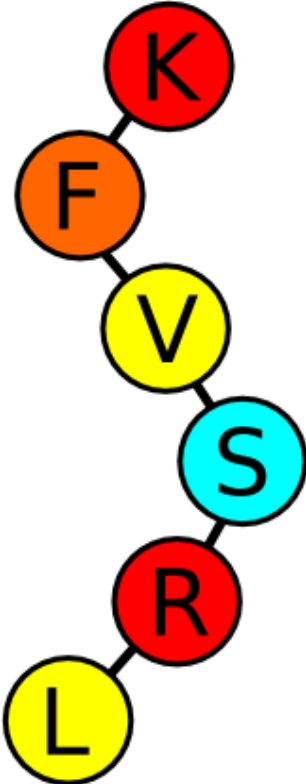

$$\Delta E_{\text{total}} = \sum_i w_i E_i(\theta_i, \mathbf{a}_i)$$

Energy is currently given in:
“kcal/mol”

Previously in: “REU”

High Energy = Bad
Low Energy = Good

Rosetta is a Residue-Centric Scoring Function

One Body	Two Body	Whole Body
		

Rosetta is a Residue-Centric Scoring Function

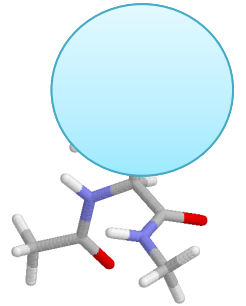
One Body	Two Body	Whole Body
Backbone - p_aa_pp - rama_prepro	Lennard-Jones - fa_atr - fa_rep	Radius of Gyration - rg
Side Chain - fa_dun - yhh_planarity	Solvation - fa_sol Hydrogen Bond - hbond_lr_bb - hbond_sr_bb - hbond_bb_sc - hbond_sc	Contact Order - co Structure Alignment - hs_pair - ss_pair -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

* NOTE: Not all score terms are listed here

Rosetta has 2 score function modes

Low resolution (or centroid) mode:

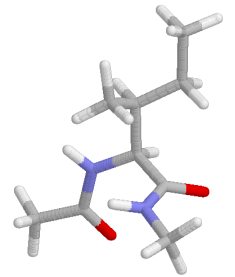
Reduced atom representation (centroid)
Simple energy function
Aggressively search conformational space



[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

^aAU = arbitrary units.

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 4HKX_relax_0001.pdb
label fa_atr fa_rep fa_sol fa_intra_rep fa_intra_sol_xover4 lk_ball_wtd fa_elec pro_close
hbond_sr_bb hbond_lr_bb hbond_bb_sc hbond_sc dslf_fa13 omega fa_dun p_aa_pp yhh_planarity ref
linear_chainbreak overlap_chainbreak rama_prepro total
weights 1 0.55 1 0.005 1 1 1 1.25 1 1 1 1.25 0.4 0.7 0.6 0.625 1 1.33333 1 0.45 NA
pose -657.979 77.1566 380.752 1.3431 23.0965 -14.9312 -203.427 0.3388 -13.103 -58.5413 -27.157
-15.9292 -0.67614 9.43471 140.412 -30.8445 0.11543 61.0868 0 0 2.1637 -326.689
GLN:NtermProteinFull_1 -1.48977 0.42854 1.18113 0.00713 0.28344 -0.19442 -0.50158 0 0 0 0 0 0
0.05724 2.56759 0 0 -1.45095 0 0 0 0.88833
VAL_2 -4.5668 1.06721 0.11949 0.01762 0.0463 -0.38312 -0.4485 0 0 0 0 0 0 -0.07335 0.04122
-0.4061 0 2.64269 0 0 -0.23209 -2.17543
GLN_3 -4.08934 0.14321 3.00406 0.00998 0.31734 -0.15915 -1.28294 0 0 0 0 0 0 0.1129 2.42608
0.06123 0 -1.45095 0 0 -0.15873 -1.06633
LEU_4 -8.04522 1.02226 0.14827 0.01552 0.05329 -0.37278 -0.70349 0 0 0 0 0 0 -0.02067 0.26477
-0.09394 0 1.66147 0 0 0.04147 -6.02905
VAL_5 -4.18566 0.16545 2.78841 0.0159 0.04504 0.07092 -2.12574 0 0 0 0 0 0 0.04194 0.0256
-0.7274 0 2.64269 0 0 -0.20212 -1.44497
GLN_6 -9.48887 0.87951 7.78324 0.01336 0.21414 0.32642 -4.49175 0 0 0 -2.35214 -0.76922 0
0.01077 4.10003 -0.05635 0 -1.45095 0 0 -0.22288 -5.5047
SER_7 -3.10817 0.23469 4.30483 0.00131 0.06508 -0.06925 -1.94284 0 0 0 -1.3938 0 0 0.10072
0.25684 -0.49292 0 -0.28969 0 0 -0.22691 -2.56012
GLY_8 -1.39358 0.15797 1.38988 3e-05 0 -0.12084 0.16418 0 0 0 0 0 0 0.05484 0 -0.26219 0
0.79816 0 0 -0.15431 0.63414
```

Output Score Table

Score Terms:

```
label fa_atr fa_rep fa_sol fa_intra_rep fa_intra_sol_xover4 lk_ball_wtd fa_elec pro_close  
hbond_sr_bb hbond_lr_bb hbond_bb_sc hbond_sc dslf_fa13 omega fa_dun p_aa_pp yhh_planarity ref  
linear_chainbreak overlap_chainbreak rama_prepro total
```

Output Score Table

Weights:

```
label fa_atr fa_rep fa_sol fa_intra_rep fa_intra_sol_xover4 lk_ball_wtd fa_elec pro_close  
hbond_sr_bb hbond_lr_bb hbond_bb_sc hbond_sc dslf_fa13 omega fa_dun p_aa_pp yhh_planarity ref  
linear_chainbreak overlap_chainbreak rama_prepro total  
weights 1 0.55 1 0.005 1 1 1 1.25 1 1 1 1 1.25 0.4 0.7 0.6 0.625 1 1.33333 1 0.45 NA
```

Output Score Table

Scores for individual terms:

label	fa_atr	fa_rep	fa_sol	fa_intra_rep	fa_intra_sol_xover4	lk_ball_wtd	fa_elec	pro_close	hbond_sr_bb	hbond_lr_bb	hbond_bb_sc	hbond_sc	dslf_fa13	omega	fa_dun	p_aa_pp	yhh_planarity	ref	linear_chainbreak	overlap_chainbreak	rama_prepro	total
pose	-657.979	77.1566	380.752	1.3431	23.0965	-14.9312	-203.427	0.3388	-12.102	58.5413	-27.157	-15.9292	-0.67614	9.43471	140.412	-30.8445	0.11543	61.0868	0	0	2.1037	-326.689

Total Score of Protein

Output Score Table

Scores by amino acid position:

```
label fa_atr fa_rep fa_sol fa_intra_rep fa_intra_sol_xover4 lk_ball_wtd fa_elec pro_close  
hbond_sr_bb hbond_lr_bb hbond_bb_sc hbond_sc dslf_fa13 omega fa_dun p_aa_pp yhh_planarity ref  
linear_chainbreak overlap_chainbreak rama_prepro total
```

```
VAL_2 -4.5668 1.06721 0.11949 0.01762 0.0463 -0.38312 -0.4485 0 0 0 0 0 0 -0.07335 0.04122  
-0.4061 0 2.64269 0 0 -0.23209 -2.17543
```

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 ...

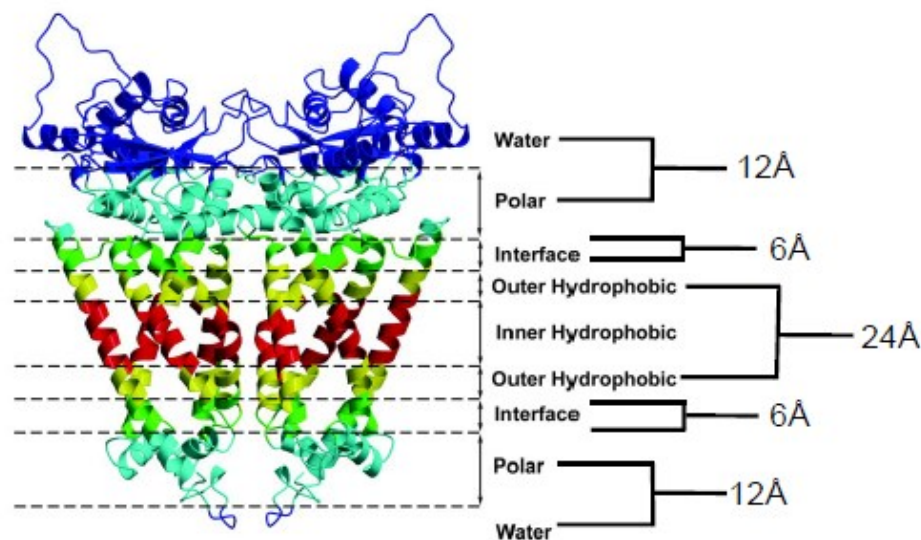
Interface with Rosetta using constraint files

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

Command line flags

-score:weights <filename>

-score:set_weight <scoreterm₁> <wt₁> <scoreterm₂> <wt₂>

-score:patch <patchfile>

Patchfile example

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

Modifying Scorefunctions in Rosetta

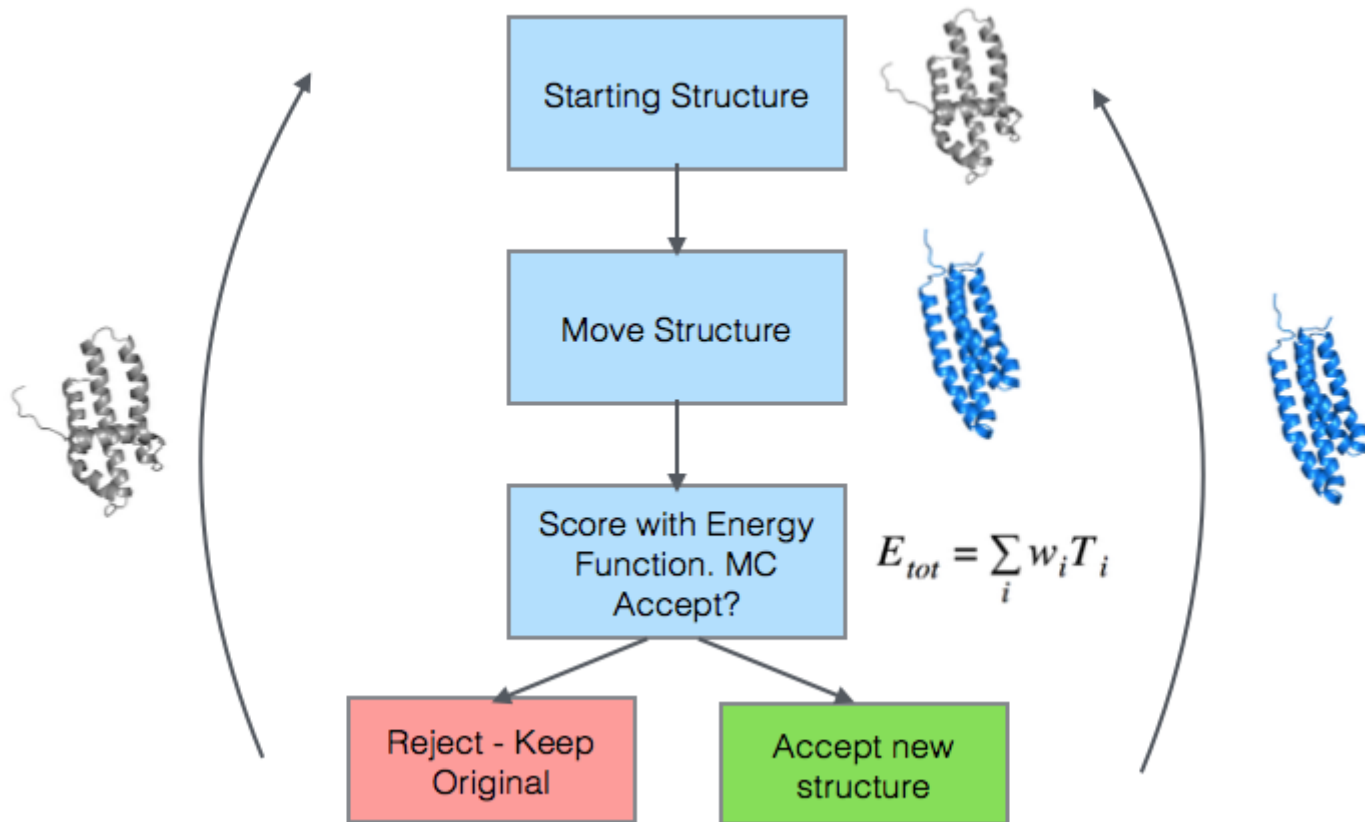
```
<ROSETTASCRIPTS>
  <SCOREFXNS>
    <ScoreFunction name="ligand_soft_rep" weights="ligand_soft_rep">
      <Reweight scoretype="fa_elec" weight="0.42"/>
    </ScoreFunction>
    <ScoreFunction name="hard_rep" weights="ligandprime">
    </ScoreFunction>
  </SCOREFXNS>
  <OUTPUT scorefxn="hard_rep"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

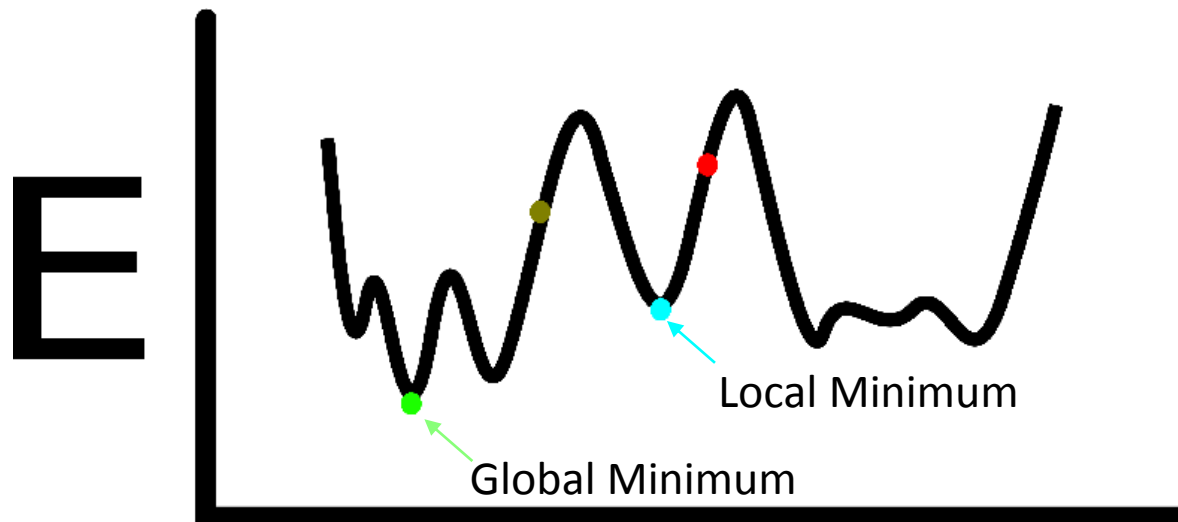
Scoring in a larger context

Score is Central to Monte Carlo Selection



Score is Central to Monte Carlo Selection

- Criteria for accepting structures (Metropolis Criterion):
- If $E_{\text{new}} < E_{\text{old}}$: Accept new structure
- If $E_{\text{new}} > E_{\text{old}}$:
 - Pick a random number $p(0, 1)$
 - if $e^{[-(E_{\text{new}} - E_{\text{old}})/kBT]} > p$, accept new structure



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