Rosetta Scoring Function (Rosetta Energy Function)

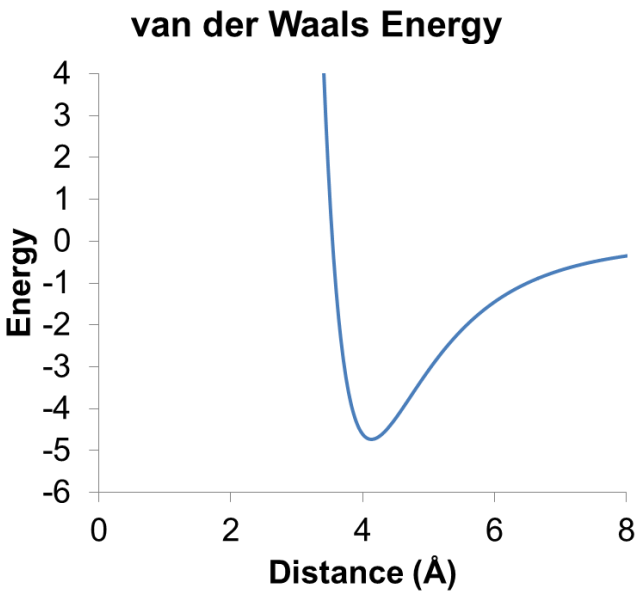
Benjamin K. Mueller
“Borrowed” from: Brian Bender
Rosetta Workshop
April 2017

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

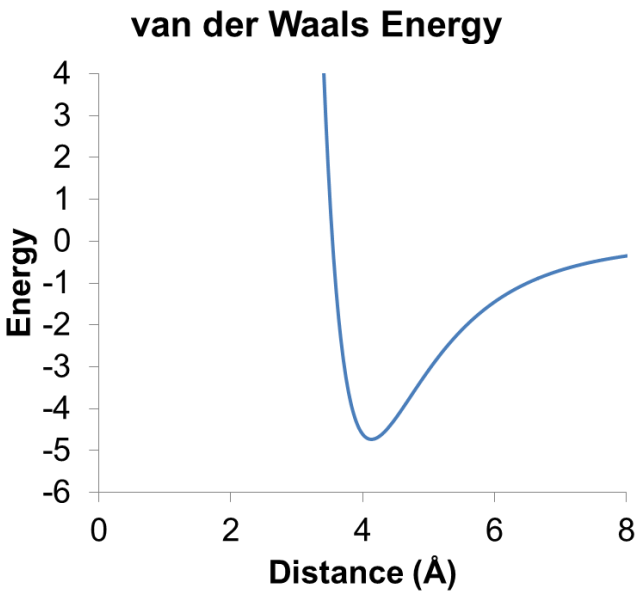
$$\sum_{i<j}\left[\frac{A_{ij}}{r_{ij}^{12}}-\frac{B_{ij}}{r_{ij}^{6}}\right]$$



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

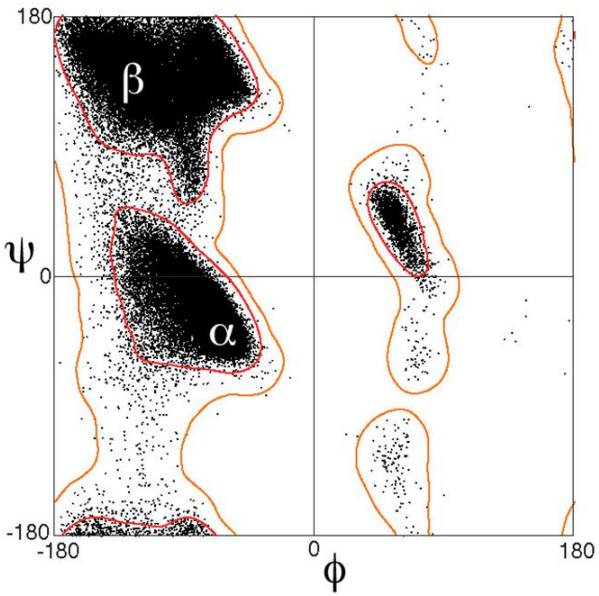
Lennard-Jones Potential

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



Ramachandran Plot

Statistical mining of
Protein Databank (PDB)



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

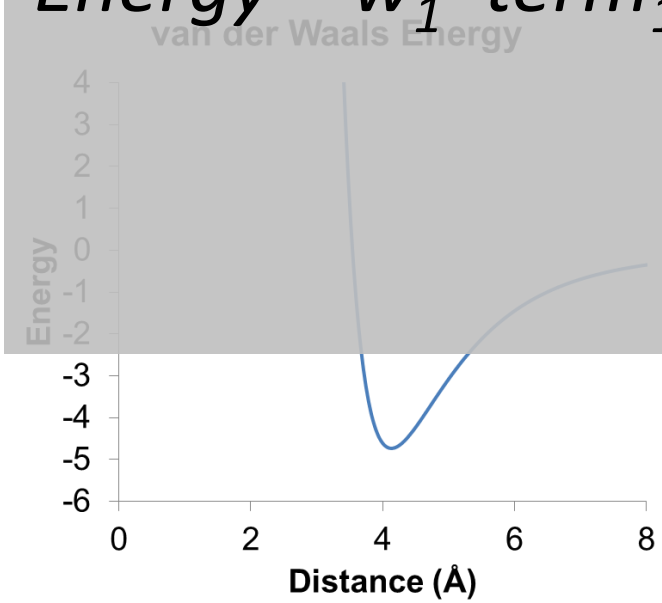
Lennard-Jones Potential

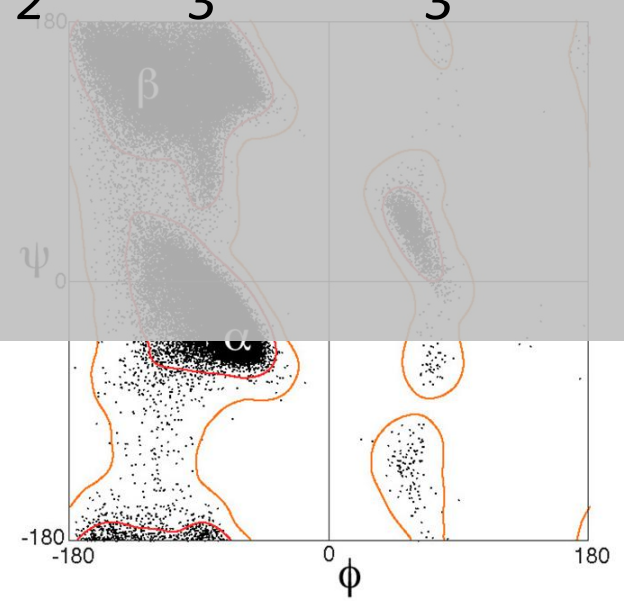
$$\sum_{i<j}\left[\frac{A_{ij}}{r_{ij}^{12}}-\frac{B_{ij}}{r_{ij}^{6}}\right]$$

Ramachandran Plot

Statistical mining of Protein Databank (PDB)

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





A Few things to Remember

$$\Delta E_{\text{total}} = \sum_i w_i E_i(\Theta_i, \text{aa}_i)$$

Energy is given in: Rosetta Energy Units (REUs)

High Energy = Bad
Low Energy = Good

Rosetta is a Residue-Centric Scoring Function

One Body	Two Body	Whole Struct

Rosetta is a Residue-Centric Scoring Function

One Body	Two Body	Whole Struct
Backbone terms • rama • p_aa_pp • omega Sidechain terms • fa_dun • yhh_planarity Reference energies • ref	Lennard-Jones • fa_atr • fa_rep Solvation Energy • fa_sol Hydrogen bonding • Hbond_lr_bb • Hbond_sr_bb • Hbond_bb_sc • Hbond_sc Electrostatics • fa_elec Sidechain-sidechain • fa_pair Disulfides Constraints	Radius of Gyration • rg Contact Order • co 2° Structure Alignment • hs_pair • ss_pair • Sheet

Rosetta Scoring “Resolution”

Low resolution:

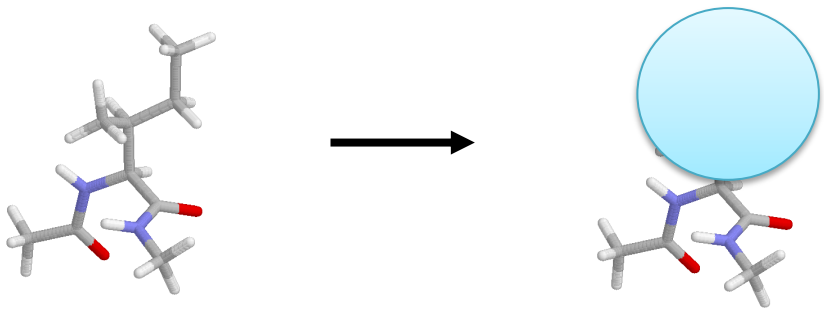
Reduced atom representation (centroid)
Simple energy function
Aggressively search conformational space

High resolution:

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

Slide by: Glenn Butterfoss
–Rosetta Design Group

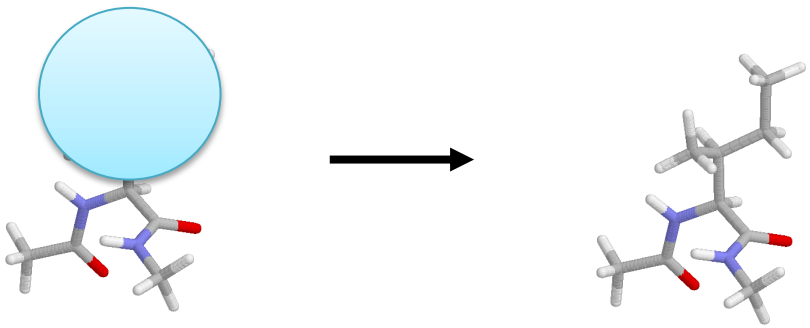
Low Resolution Atom Representation: Centroid



- Backbone full atom
- Side-chain represented as a "super atom" at the center of mass of the full atom sidechain
- Sample large conformational space

`database/chemical/residue_type_sets/centroid`

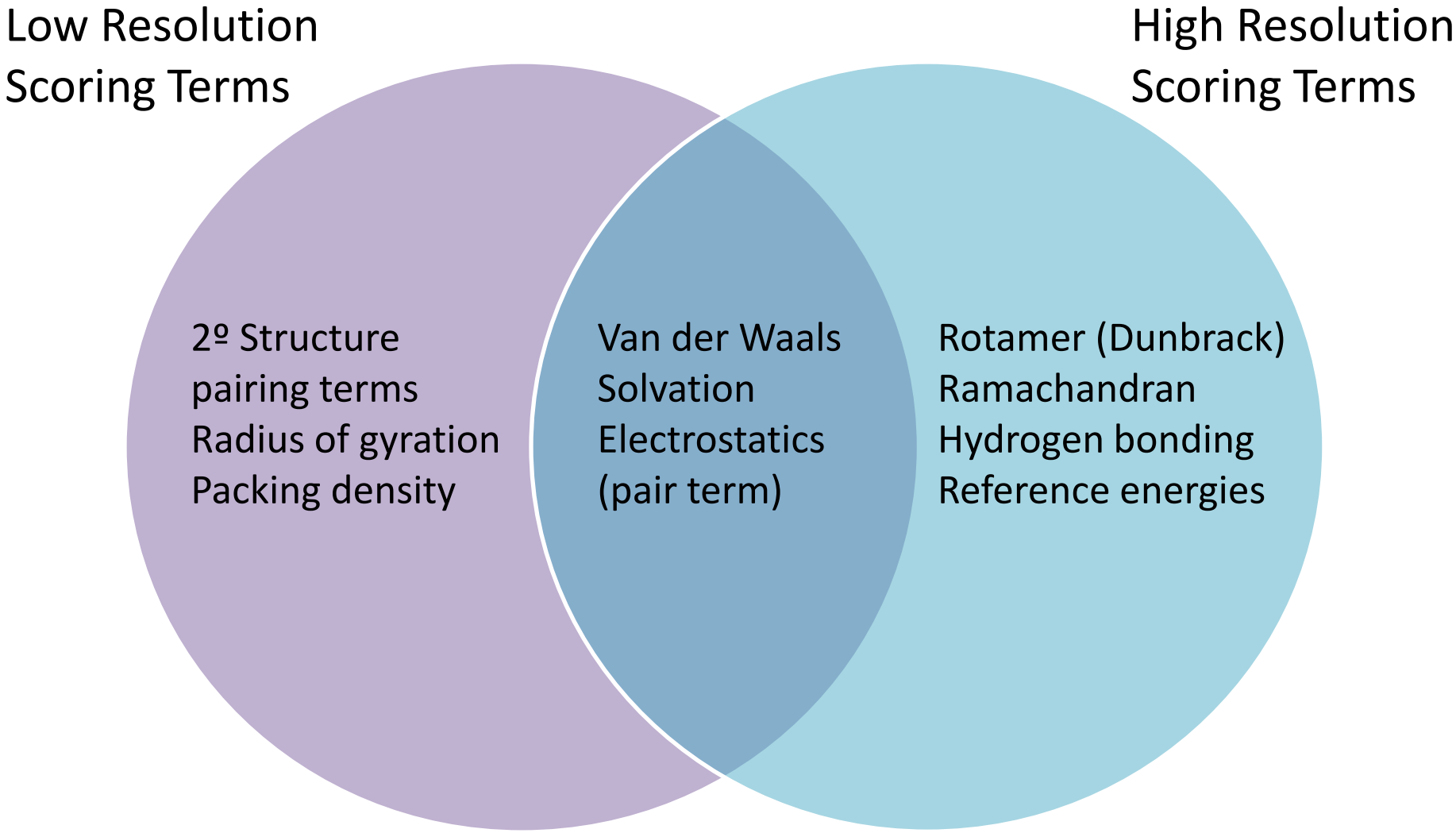
High Resolution
Atom Representation: Full-atom



- All atoms represented
- Side-chain represented as rotamers

`database/chemical/residue_type_sets/fa_standard`

Common Score Types



Breakdown of Score Terms by Class

Table 2. Standard Rosetta Score Function Terms

score term	definition
low-resolution scoring terms	
env	hydrophobicity term for each amino acid
vdw	steric repulsion between two residues
pair	probability of two residues interacting
rg	radius of gyration
cbeta	solvation term based on a number of surrounding residues
hs_pair, ss_pair, and sheet	secondary structure terms
high-resolution scoring terms (talaris2014)	
fa_atr, fa_rep, and fa_intra_rep	decomposed 6−12 Lennard-Jones potential
fa_sol	EEF1 solvation term
pro_close	proline ring closure energy
omega	omega backbone dihedral potential
dslf_fa13	updated disulfide geometry potential
rama	potential of ϕ and ψ angles for each amino acid
p_aa_pp	probability of an amino acid given a set of ϕ and ψ angles
fa_dun	rotamer likelihood
hbond_sr_bb, hbond_lr_bb, hbond_bb_sc, and hbond_sc	combined covalent−electrostatic hydrogen bond potentials for α-helices, β-sheets, side-chain backbone, and side-chain−side-chain interactions, respectively
yhh_planarity	tyrosine hydroxyl out-of-plane penalty
fa_elec	Coulombic electrostatic potential between two residues with a distance-dependent dielectric (deprecates fa_pair)

talaris2014 is the Default Score Function – for now…

```
METHOD_WEIGHTS ref 0.773742 0.443793 -1.63002 -1.96094 0.61937 0.173326
0.388298 1.0806 -0.358574 0.761128 0.249477 -1.19118 -0.250485 -1.51717 -0.32436
0.165383 0.20134 0.979644 1.23413 0.162496
fa_atr 1
fa_rep 0.55
fa_sol 0.9375
fa_intra_rep 0.005
fa_elec 0.875
pro_close 1.25
hbond_sr_bb 1.17
hbond_lr_bb 1.17
hbond_bb_sc 1.17
hbond_sc 1.1
dslf_fa13 1.25
rama 0.25
omega 0.625
fa_dun 0.7
p_aa_pp 0.4
yhh_planarity 0.625
ref 1
```

Score Function =
List of score terms and corresponding weight

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 design_3UKM.pdb
label fa_atr fa_rep fa_intra_rep fa_mbenv fa_mbsolv pro_close fa_pair
hbond_sr_bb hbond_lr_bb hbond_bb_sc hbond_sc dslf_ss_dst dslf_cs_ang
dslf_ss_dih dslf_ca_dih rama omega fa_dun p_aa_pp ref Menv_smooth total
weights 0.8 0.44 0.004 0.3 0.35 1 0.49 1.17 1.17 2.34 2.2 0.5 2 5 5 0.2
0.5 0.56 0.32 1 0.5 NA
pose -2065.11 325.373 4.42479 124.101 312.059 0.66056 -18.5298 -257.704
-8.99415 -52.1492 -79.8909 -1.76827 8.00442 0.30851 3.12137 -15.362
46.2914 253.303 -32.8605 -42.76 -193.732 -1691.22
LEU_p:NtermProteinFull_1 -4.60762 0.57643 0.01812 4.16856 0.56856 0 0 0
0 0 0 0 0 0 0 0 0.02487 1.03873 0 -0.2 -1.74471 -0.15705
```

Output Score Table

Score Terms:

```
# All scores below are weighted scores, not raw scores.
#BEGIN_POSE_ENERGIES_TABLE design_3UKM.pdb
label fa_atr fa_rep fa_intra_rep fa_mbenv fa_mbsolv pro_close fa_pair
hbond_sr_bb hbond_lr_bb hbond_bb_sc hbond_sc dslf_ss_dst dslf_cs_ang
dslf_ss_dih dslf_ca_dih rama omega fa_dun p_aa_pp ref Menv_smooth total
weights 0.8 0.44 0.004 0.3 0.35 1 0.49 1.17 1.17 2.34 2.2 0.5 2 5 5 0.2
0.5 0.56 0.32 1 0.5 NA
pose -2065.11 325.373 4.42479 124.101 312.059 0.66056 -18.5298 -257.704
-8.99415 -52.1492 -79.8909 -1.76827 8.00442 0.30851 3.12137 -15.362
46.2914 253.303 -32.8605 -42.76 -193.732 -1691.22
LEU_p:NtermProteinFull_1 -4.60762 0.57643 0.01812 4.16856 0.56856 0 0 0
0 0 0 0 0 0 0 0 0 0.02487 1.03873 0 -0.2 -1.74471 -0.15705
```

Output Score Table

Weights:

```
# All scores below are weighted scores, not raw scores.
#BEGIN_POSE_ENERGIES_TABLE design_3UKM.pdb
label fa_atr fa_rep fa_intra_rep fa_mbenv fa_mbsolv pro_close fa_pair
hbond_sr_bb hbond_lr_bb hbond_bb_sc hbond_sc dslf_ss_dst dslf_cs_ang
dslf_ss_dih dslf_ca_dih rama omega fa_dun p_aa_pp ref Menv_smooth total
weights 0.8 0.44 0.004 0.3 0.35 1 0.49 1.17 1.17 2.34 2.2 0.5 2 5 5 0.2
0.5 0.56 0.32 1 0.5 NA
pose -2065.11 325.373 4.42479 124.101 312.059 0.66056 -18.5298 -257.704
-8.99415 -52.1492 -79.8909 -1.76827 8.00442 0.30851 3.12137 -15.362
46.2914 253.303 -32.8605 -42.76 -193.732 -1691.22
LEU_p:NtermProteinFull_1 -4.60762 0.57643 0.01812 4.16856 0.56856 0 0 0
0 0 0 0 0 0 0 0 0 0.02487 1.03873 0 -0.2 -1.74471 -0.15705
```

Output Score Table

Scores for individual terms:

```
# All scores below are weighted scores, not raw scores.
#BEGIN_POSE_ENERGIES_TABLE design_3UKM.pdb
label fa_atr fa_rep fa_intra_rep fa_mbenv fa_mbsolv pro_close fa_pair
hbond_sr_bb hbond_lr_bb hbond_bb_sc hbond_sc dslf_ss_dst dslf_cs_ang
dslf_ss_dih dslf_ca_dih rama omega fa_dun p_aa_pp ref Menv_smooth total
weights 0.8 0.44 0.004 0.3 0.35 1 0.49 1.17 1.17 2.34 2.2 0.5 2 5 5 0.2
0.5 0.56 0.32 1 0.5 NA
pose -2065.11 325.373 4.42479 124.101 312.059 0.66056 -18.5298 -257.704
-8.99415 -52.1492 -79.8909 -1.76827 8.00442 0.30851 3.12137 -15.362
46.2914 253.303 -32.8605 -42.76 -193.732 -1691.22
LEU_p:NtermProteinFull_1 -4.60762 0.57643 0.01812 4.16856 0.56856 0 0 0
0 0 0 0 0 0 0 0 0 0.02487 1.03873 0 -0.2 1.74471 0.15705
```

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 design_3UKM.pdb
label fa_atr fa_rep fa_intra_rep fa_mbenv fa_mbsolv pro_close fa_pair
hbond_sr_bb hbond_lr_bb hbond_bb_sc hbond_sc dslf_ss_dst dslf_cs_ang
dslf_ss_dih dslf_ca_dih rama omega fa_dun p_aa_pp ref Menv_smooth total
weights 0.8 0.44 0.004 0.3 0.35 1 0.49 1.17 1.17 2.34 2.2 0.5 2 5 5 0.2
0.5 0.56 0.32 1 0.5 NA
pose -2065.11 325.373 4.42479 124.101 312.059 0.66056 -18.5298 -257.704
-8.99415 -52.1492 -79.8909 -1.76827 8.00442 0.30851 3.12137 -15.362
46.2914 253.303 -32.8605 -42.76 -193.732 -1691.22
LEU_p:NtermProteinFull_1 -4.60762 0.57643 0.01812 4.16856 0.56856 0 0 0
0 0 0 0 0 0 0 0 0 0.02487 1.03873 0 -0.2 -1.74471 -0.15705
```

More score term information

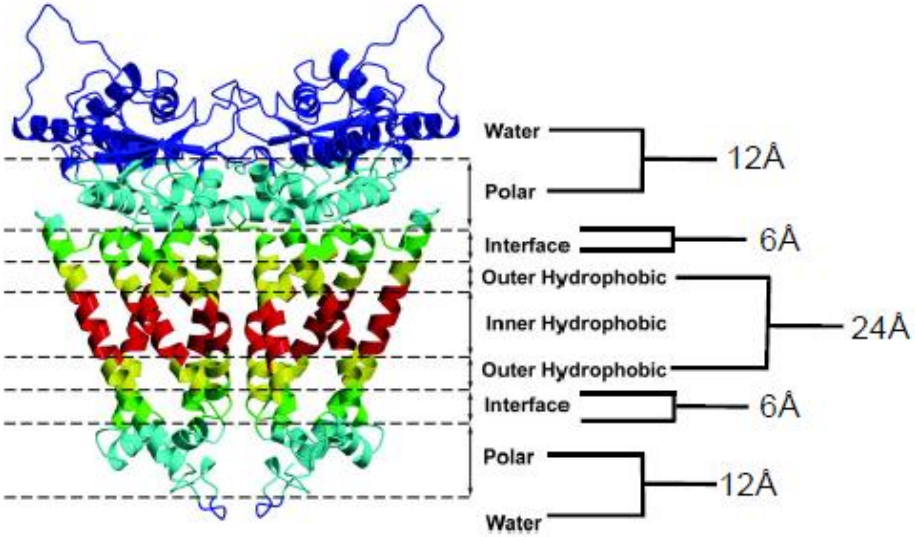
Constraints (actually “restraints”)

- 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 Talaris2014
- To turn on score term, set weight to non-zero value

Modifying Scorefunctions in Rosetta

Command line flags

-score:weights <filename>

-score:set_weight <scoreterm$_1$> <weight$_1$> <scoreterm$_2$> <weight$_2$>

-score:patch <patchfile>

Patchfile example

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

Modifying Scorefunctions in Rosetta

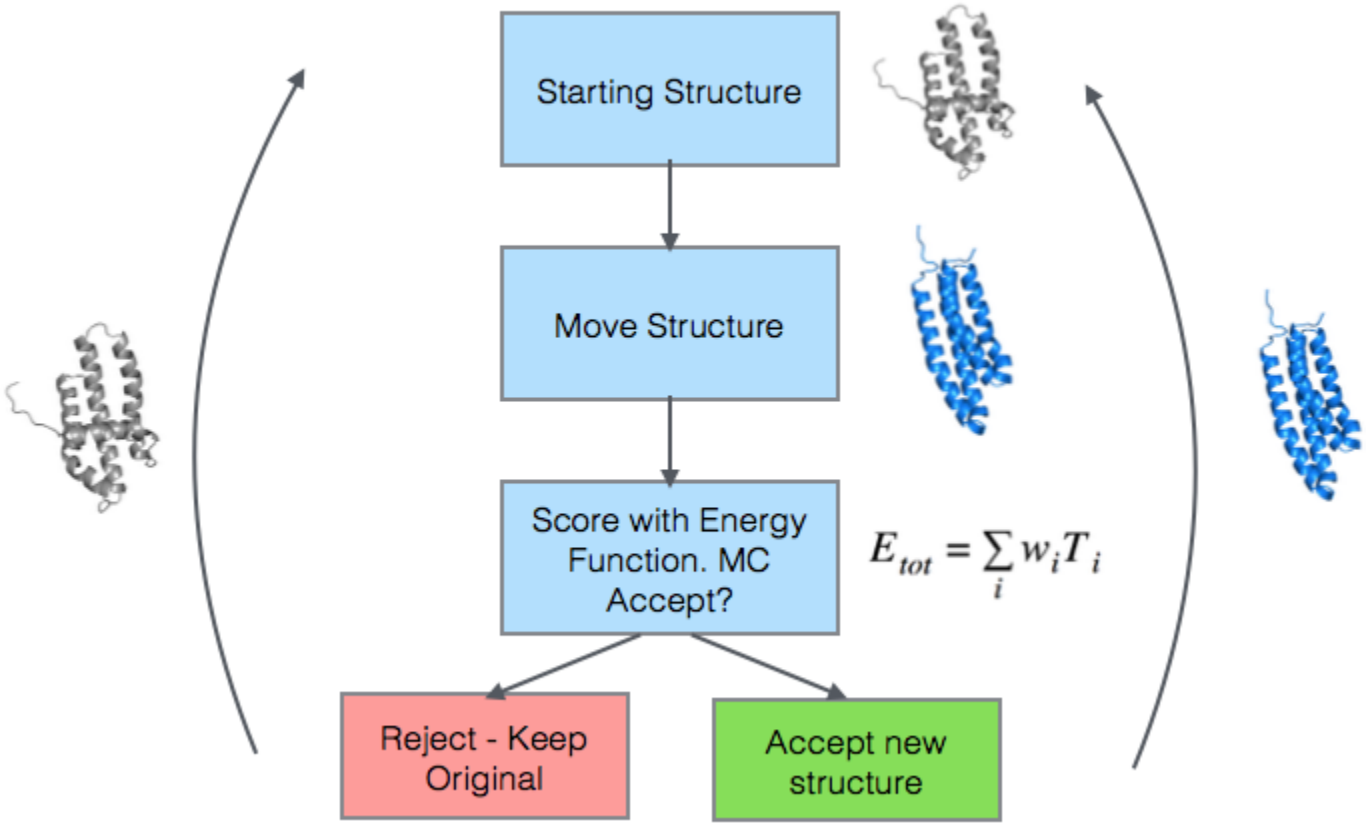
```
<ROSETTASCRIPTS>
    <SCOREFXNS>
        <ScoreFunction name="ligand_soft_rep" weights="ligand_soft_rep">
            <Reweight scoretype="fa_elec" weight="0.42"/>
        </ScoreFunction>
        <ScoreFunction name="hard_rep" weights="ligandprime">
        </ScoreFunction>
    </SCOREFXNS>
   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

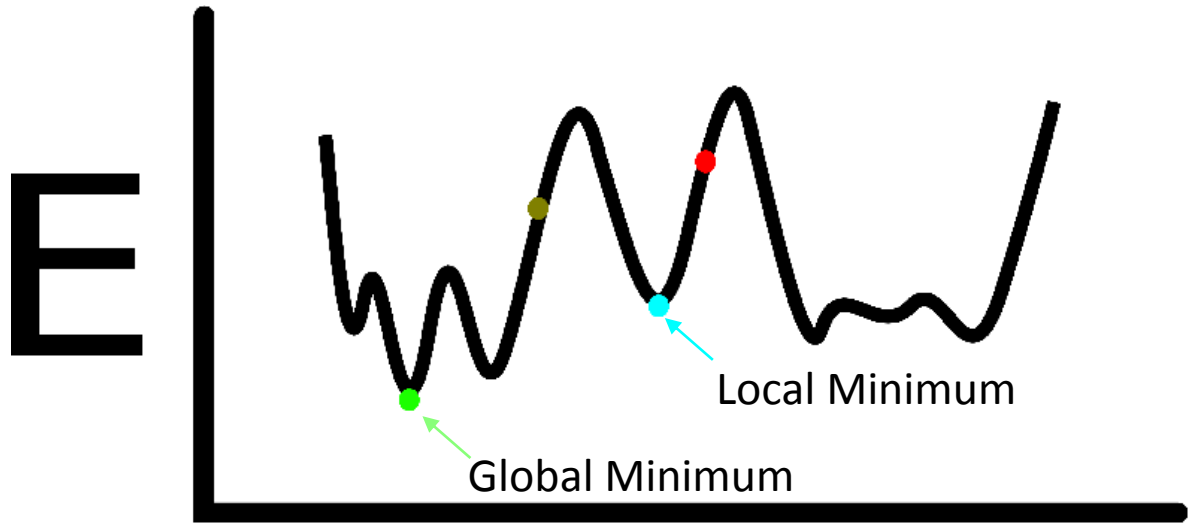


Score is Central to Monte Carlo Selection

- Criteria for accepting structures (Metropolis Criterion):
- If 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



References:

“Talaris 2014” 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.

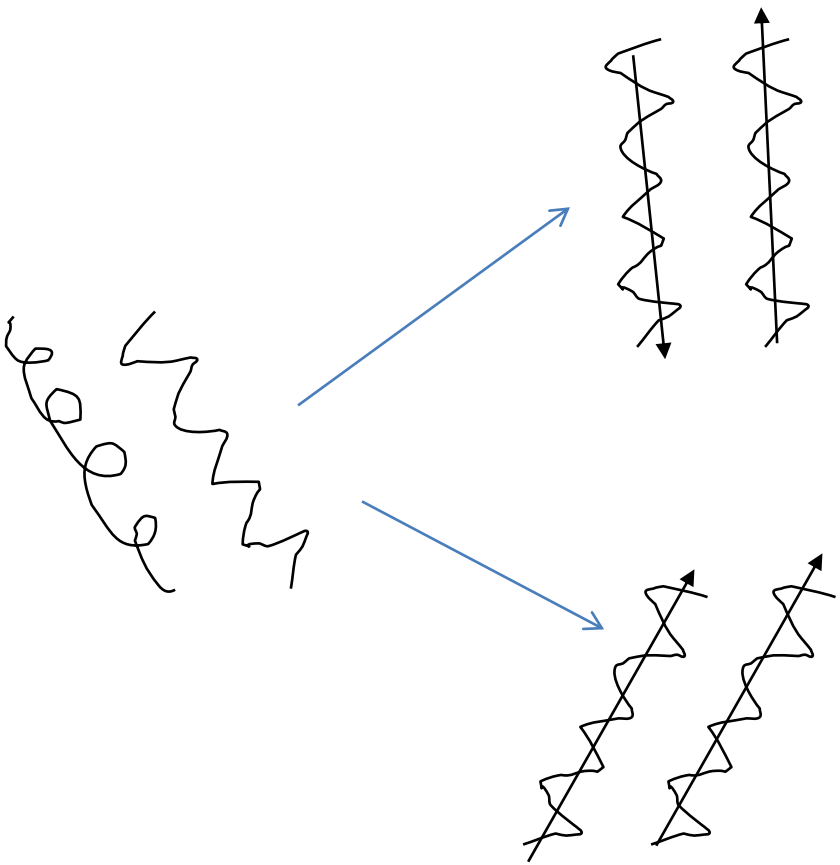
REF2015 References (coming soon):

http://biorxiv.org/content/early/2017/02/07/106054 (preprint link)

Useful Links

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

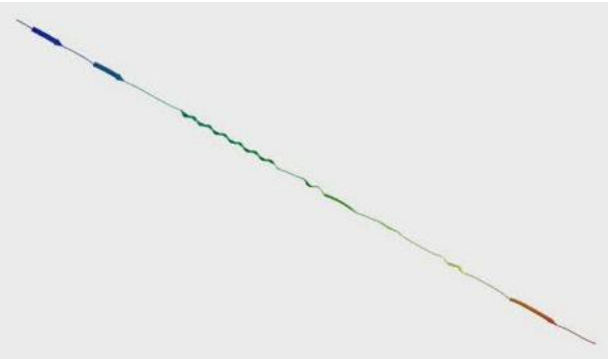
2º Structure pairing terms

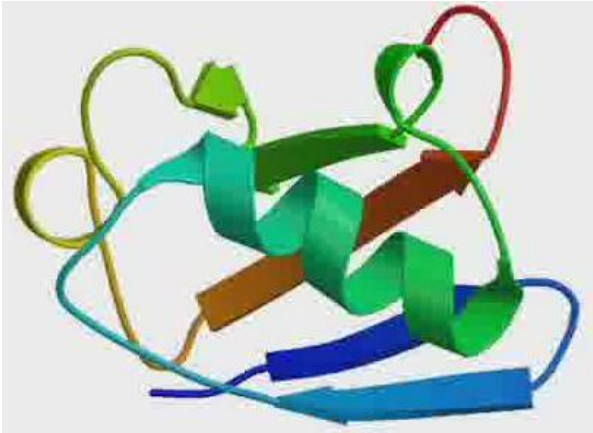


- `hs_pair / ss_pair / sheet`
- Aligns secondary structures to form helix/strand secondary structures
- Represent protein as vectors of 2 residue “strands”
- Scores selected to discriminate “near native structures for “non native”

Image by: Glenn Butterfoss –Rosetta Design Group

Radius of Gyration / Packing Density





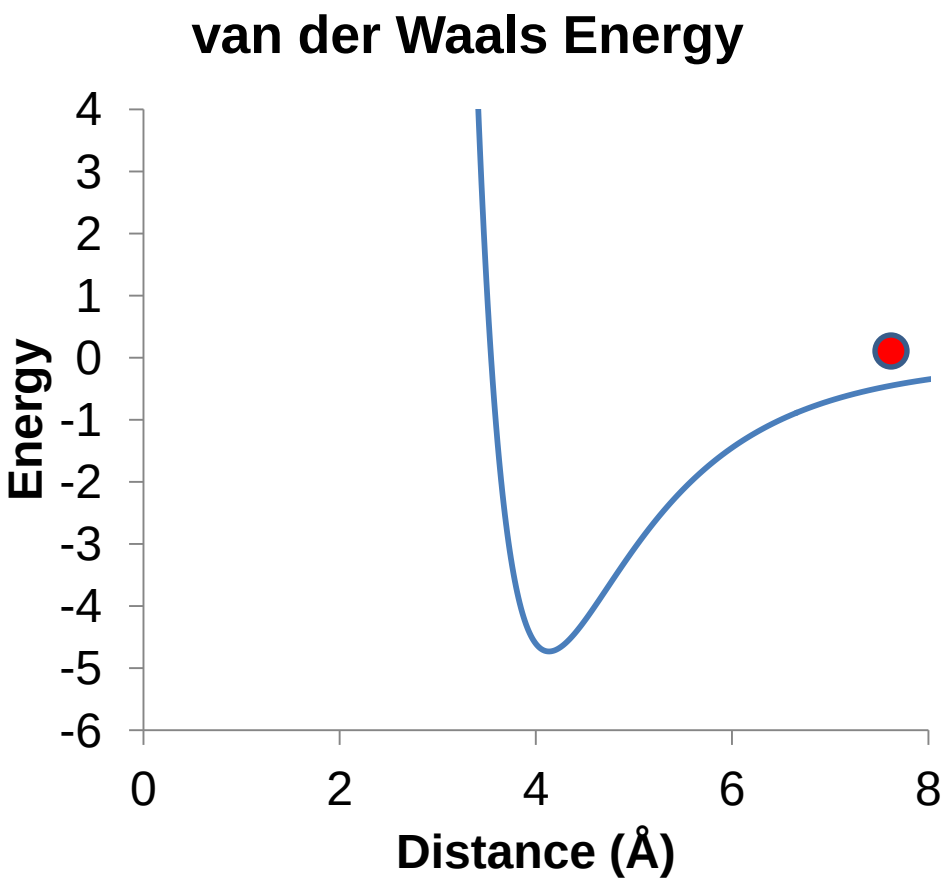
- rg / cenpack
- Promotes a compact fold
- Used early in the folding algorithm for filtering

Reference Energy

- `ref`
- Unique “cost” for *designing* in each residue type
- Free energy of given aa in unfolded state
- New meaning
  - Optimized for aa composition recovery

Van der Waals

- `fa_atr / fa_rep / fa_intra_rep`
- 12-6 Lennard-Jones potential



Solvation

- `fa_sol`
- Based on Lazaridis Karplus term
- Implicit consideration of water
- Both penalty and bonus

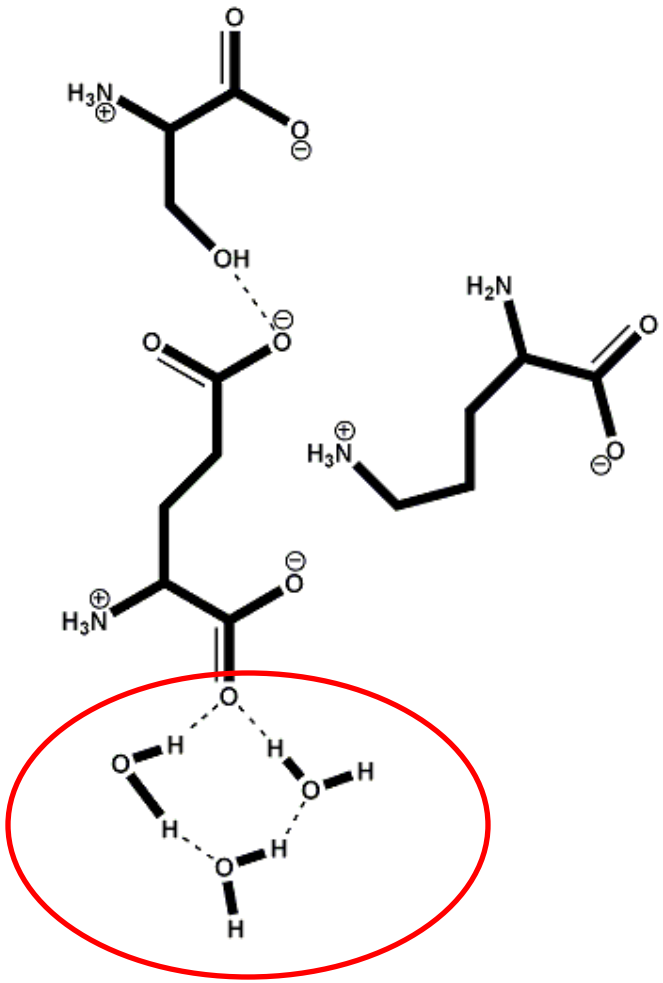
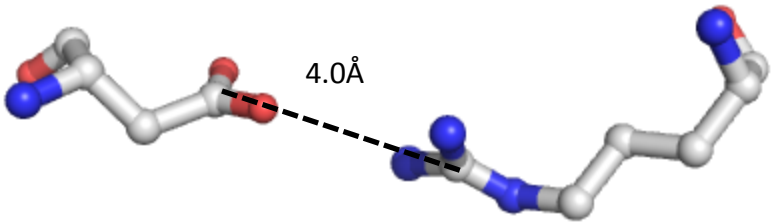


Image by:
Kristian Kaufmann

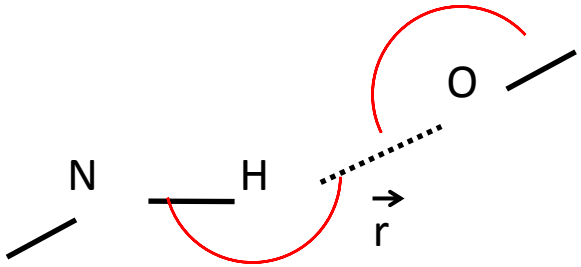
Electrostatics



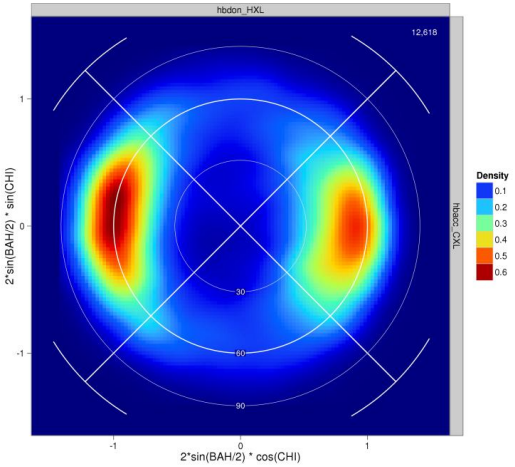
- `fa_elec`
- coulombic electrostatic potential with a distance-dependant dielectric
- Improved decoy discrimination
- Previously fa_pair – statistics based pair term

Hydrogen Bond Terms

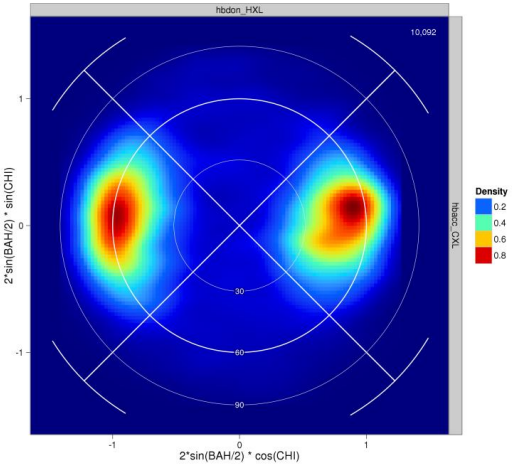
- hbond_lr_bb: long-range backbone
- hbond_sr_bb: short-range backbone
- hbond_bb_sc: backbone-sidechain
- hbond_sc: sidechain-sidechain
- Geometry dependent



Crystal Structure



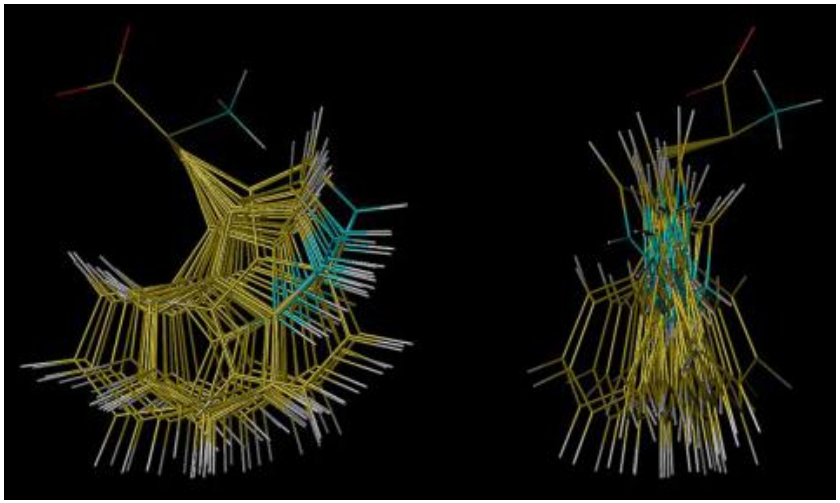
talaris2013



yhh_planarity: fixes geometry of hydrogen to lay within plane of Tyrosine ring

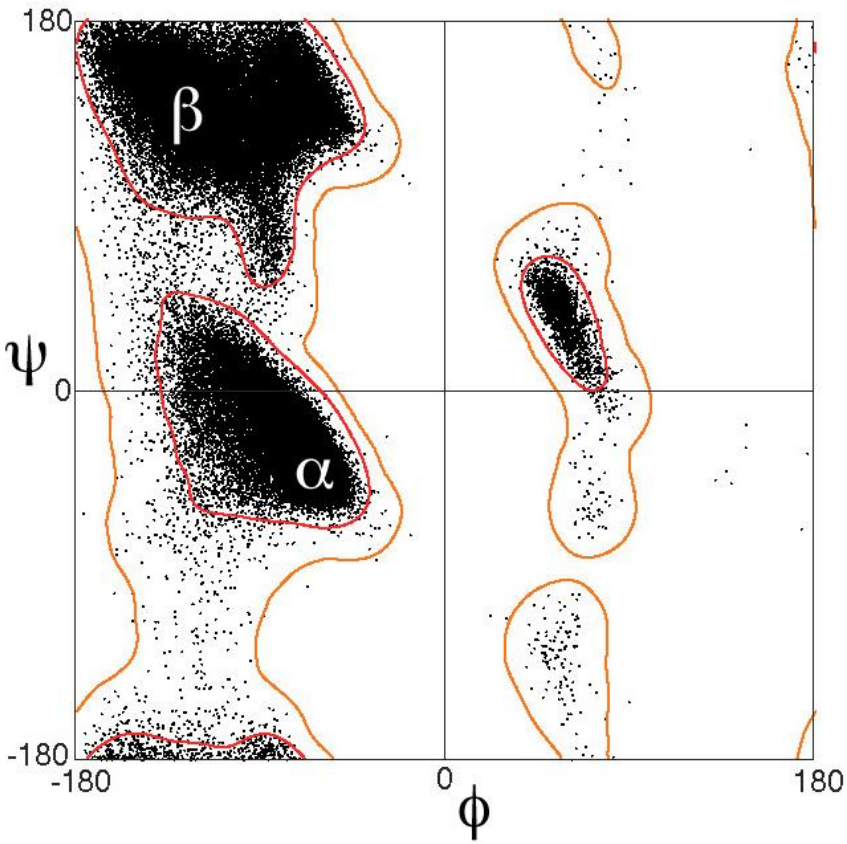
Dunbrack Rotamer Energy

- fa_dun
- Internal energy of a given rotamer
- Based on statistics of conformations of sidechains in the PDB
  - Compiled by Roland Dunbrack for the widely-used Dunbrack rotamer library.



Ramachandran and Phi Psi angles

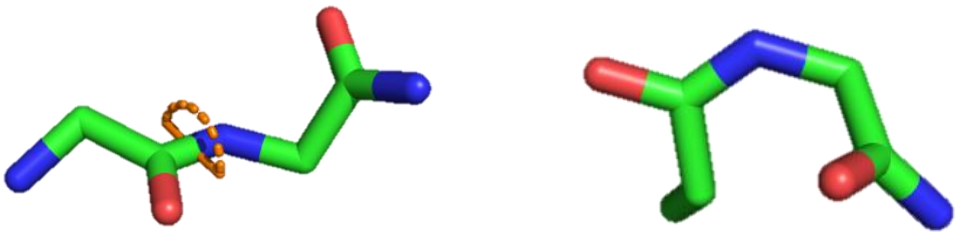
- `rama-` ramachandran preference
- `p_aa_pp-` probability of amino acid for phi, psi

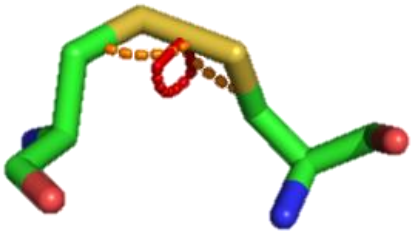


Lovell SC et al. (2003) *Proteins* **50** (3): 437–450.

Bonding penalties

- omega – omega angle penalty



- pro_close – proline ring closure

- dslf_fa13 – disulfide closure
  - Statistics based. (distance, angles, dihedrals)