

Rosetta Scoring Function

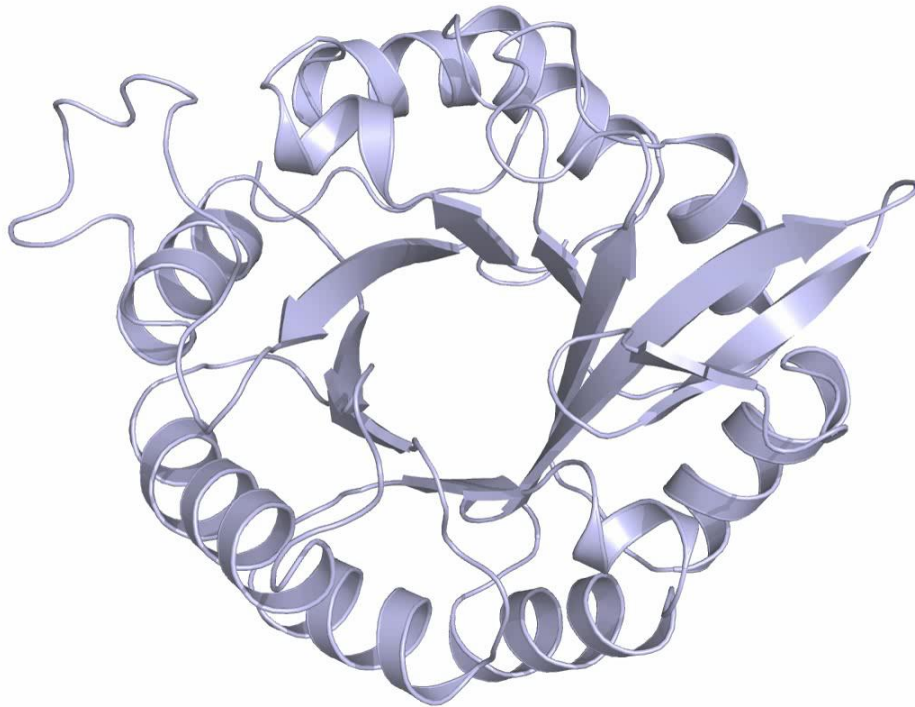
Amanda Duran

amanda.m.duran@vanderbilt.edu

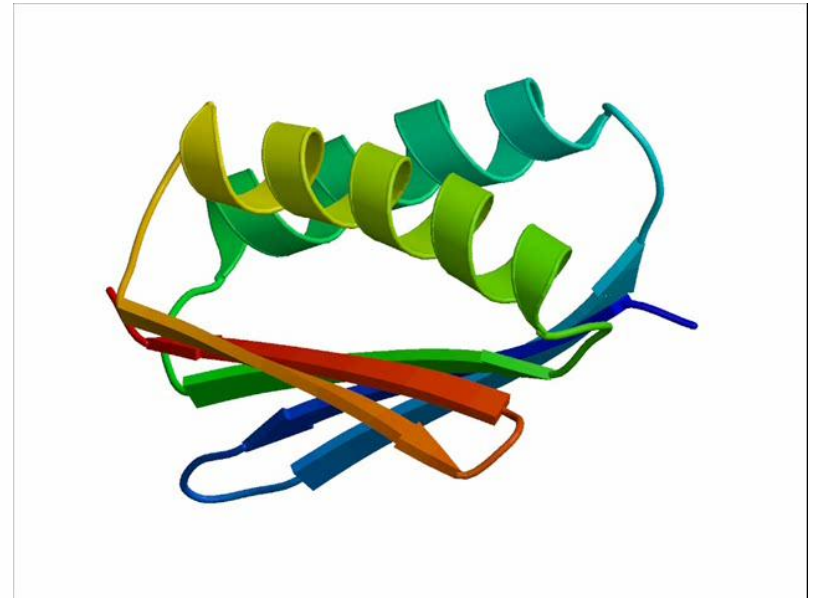
Rosetta Software Suite is a Premier Software Suite

- What it does:
 - *De novo* folding
 - Protein and small molecule docking
 - Protein design
 - Enzyme design
 - DNA/RNA design
- Developed by 12 labs
 - Over 50 developers
- First
 - *De novo* protein
 - *De novo* enzyme
 - *De novo* molecular switch

Rosetta Combines Conformational Sampling and a Robust Scoring Function for Structure Determination



Global Conformational Sampling



Local Side-chain Sampling

Common Computational Scoring Functions

Molecular Mechanics

CHARMM / AMBER

$$\begin{aligned}
 V_{total} = & \sum_{bonds} K_r (r - r_{eq})^2 \\
 & + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 \\
 & + \sum_{dihedrals} K_\phi (1 + \cos(n\phi)) \\
 & + \sum_{i < j} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right] \text{Van der Waals} \\
 & + \sum_{i < j} \left[\frac{q_i q_j}{\epsilon r_{ij}} \right] \text{Electrostatics}
 \end{aligned}$$

Knowledge Based

Rosetta

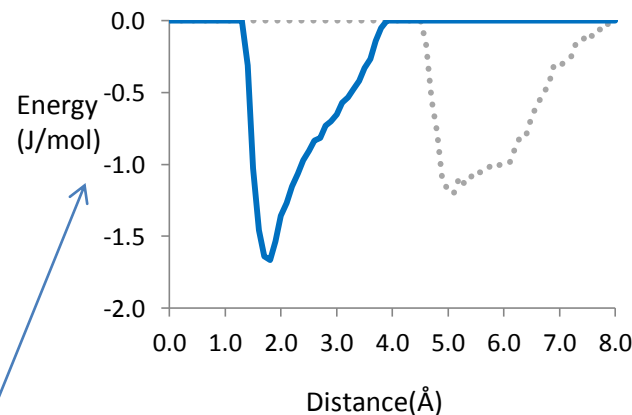
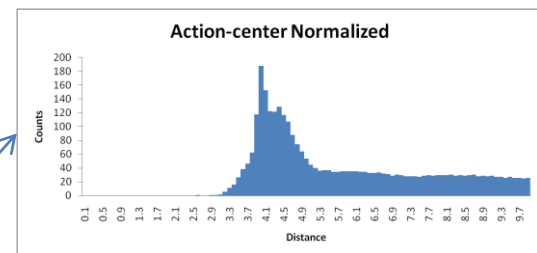
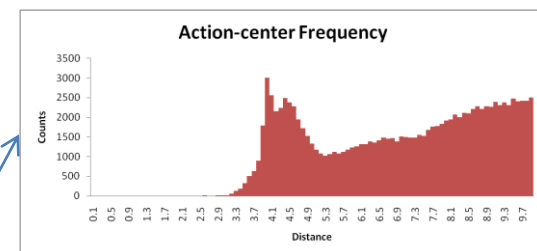
High Resolution
Crystal Structures

Frequency

Normalize

Propensity

Boltzmann
Energy



$$Energy = w_1 * term_1 + w_2 * term_2 + ...$$

Rosetta Scoring Function Classes

1. Low resolution:

Reduced atom representation (centroid)

Simple energy function

Aggressively search conformational space

2. High resolution:

Full atom (FA)

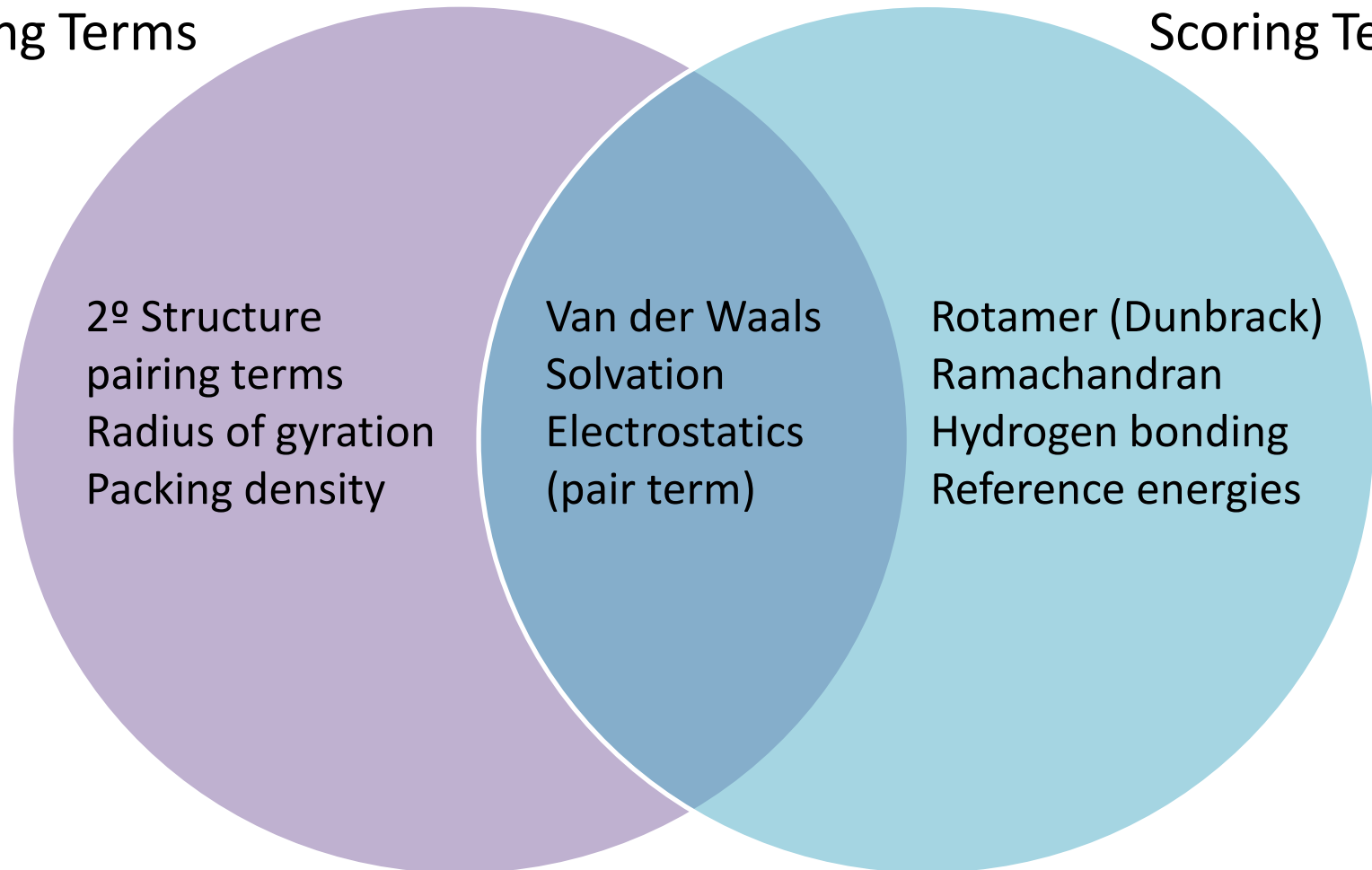
More sophisticated energy function

“Local” search of conformational (and sequence) space

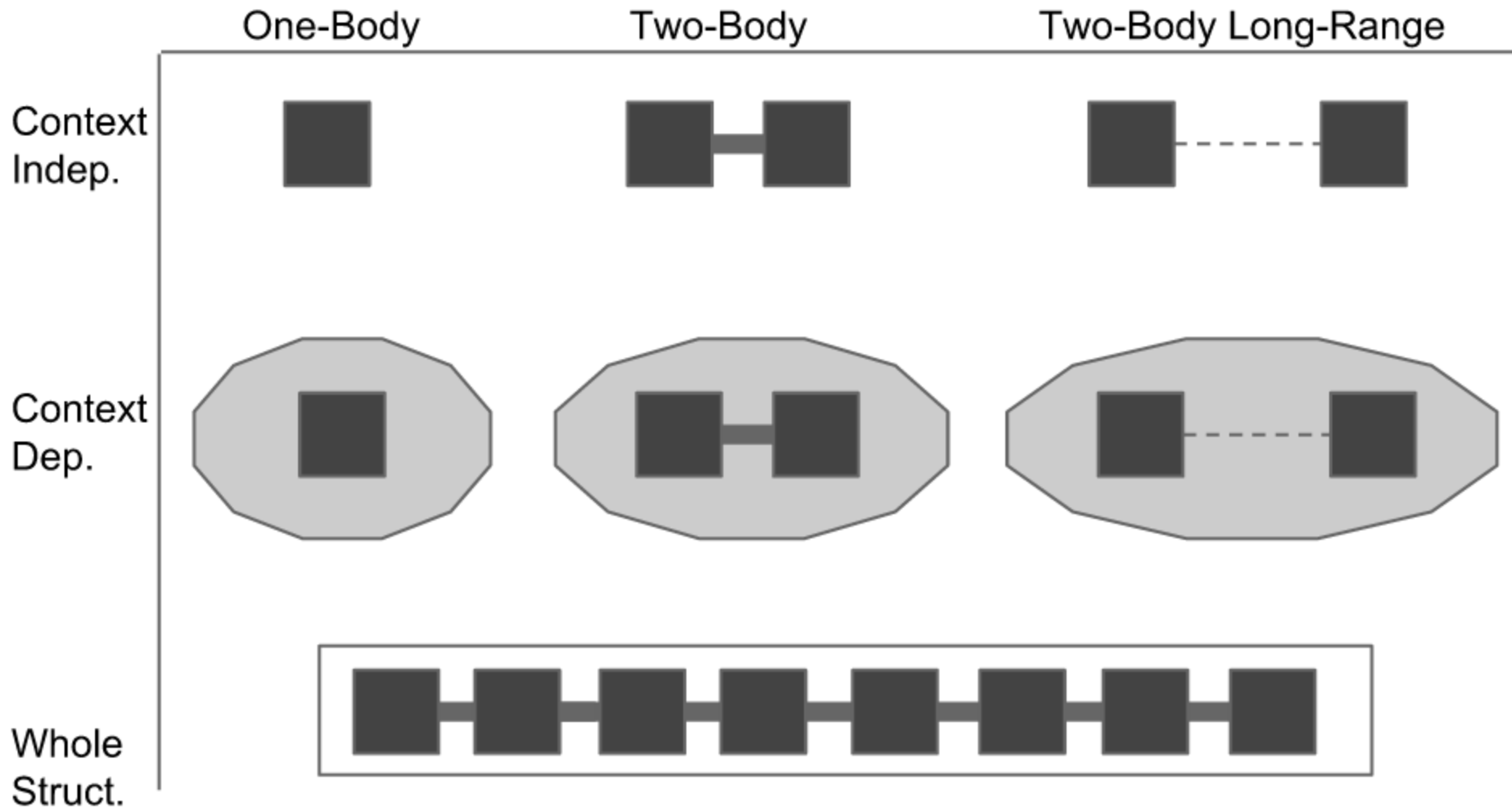
Common Score Types

Low Resolution
Scoring Terms

High Resolution
Scoring Terms



EnergyMethod Classes



Example Score Types

	One-Body	Two-Body	Two-Body Long-Range
Context Indep.	rama fa_dun	fa_atr fa_rep fa_sol	PB_elec elec_dens_fast
Context Dep.	env neigh_vect occ_sol_exact	hbond_* fa_pair	gb_elec lk_ball
Whole Struct.	rg pack_stat co rdc		

Where to Find Score Table

All scores below are weighted scores, not raw scores.

#BEGIN_POSE_ENERGIES_TABLE design_3UKM.pdb Score terms

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	0.02487	1.03873	0	-0.2	-1.74471	-0.15705					

Where to Find Score Table

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_mbsolv 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

Weights

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 0.02487 1.03873 0 -0.2 -1.74471 -0.15705

Where to Find Score Table

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

Total energies for score terms

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
------	----------	---------	---------	---------	---------	---------	----------	----------	----------	----------	----------	----------	---------	---------	---------	---------	---------	---------	----------	--------	----------	----------

← Total energy for protein

LEU_p:NtermProteinFull_1	-4.60762	0.57643	0.01812	4.16856	0.56856	0	0	0	0	0	0	0.02487	1.03873	0	-0.2	-1.74471	-0.15705
--------------------------	----------	---------	---------	---------	---------	---	---	---	---	---	---	---------	---------	---	------	----------	----------

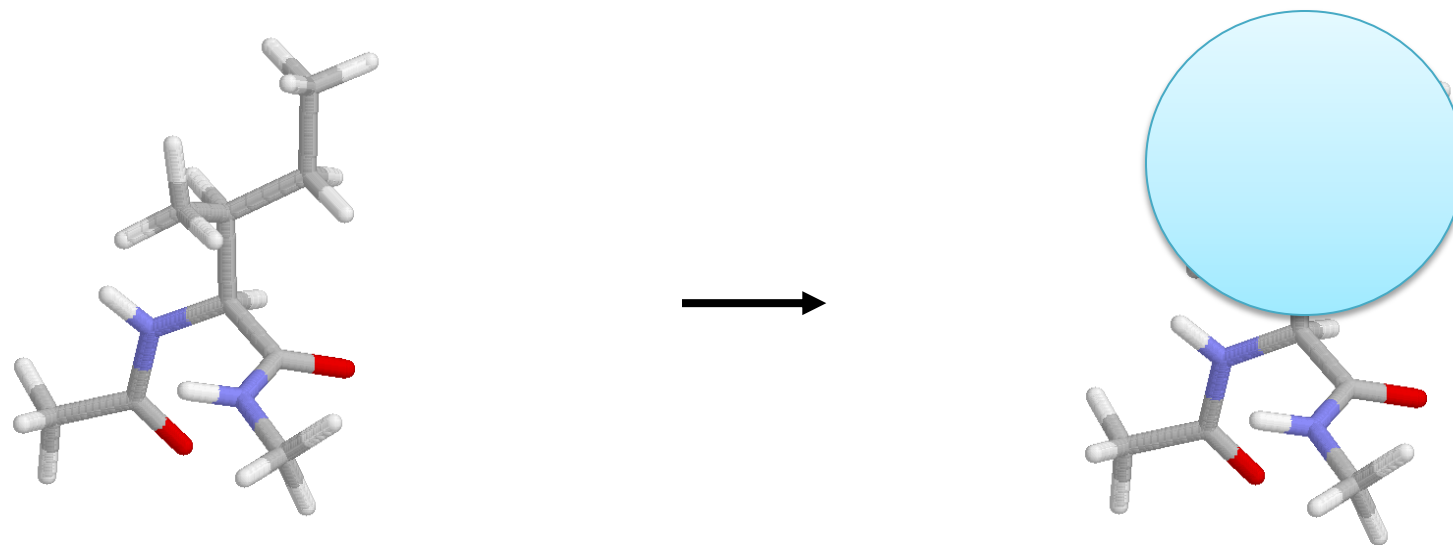
Where to Find Score Table

```
# 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
```

Individual residue energies for score terms

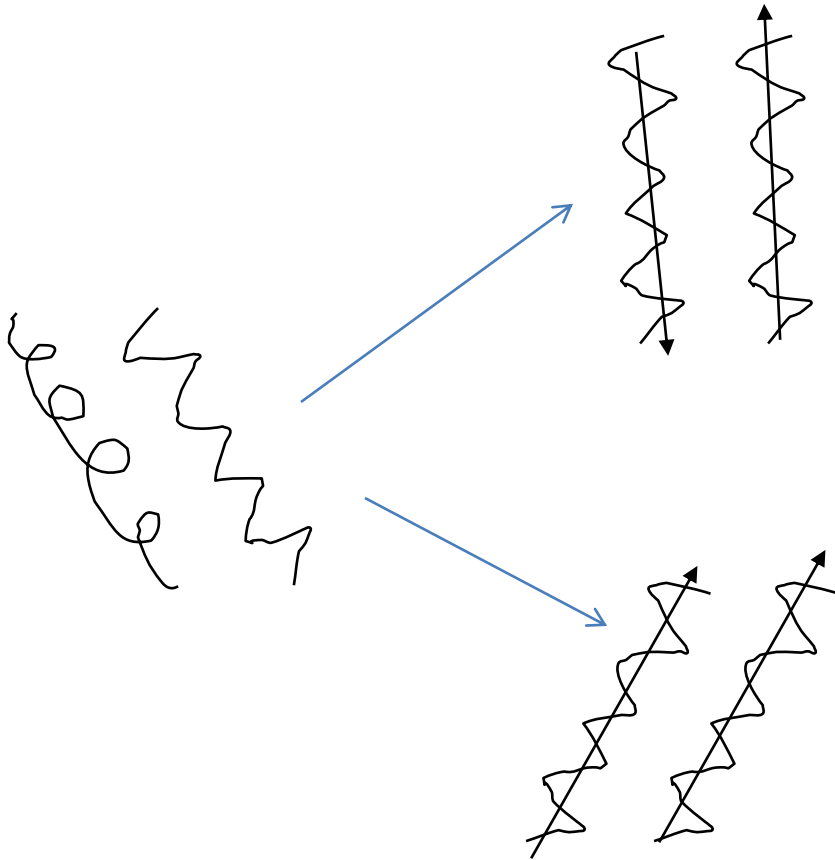
Low Resolution – Atom Representation: Centroid

database/chemical/residue_type_sets/centroid



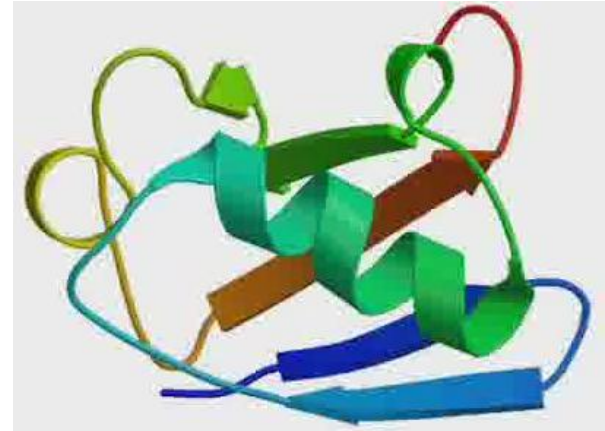
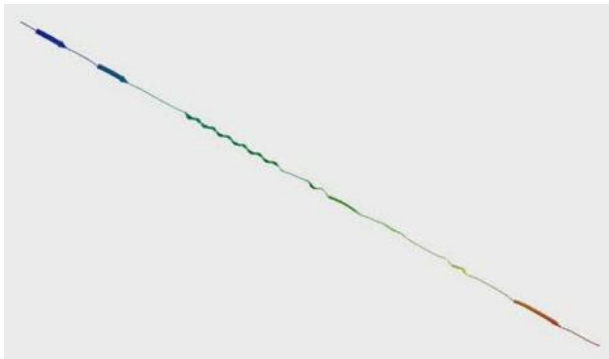
- Backbone full atom
- Side-chain represented as a “super atom”
 - “Centroid” in Rosetta jargon
- Centered at CB

Low Resolution - 2^o 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”

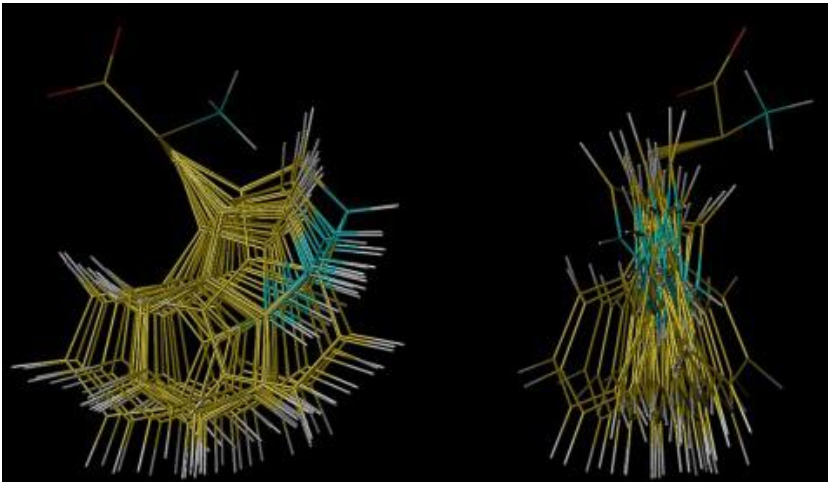
Low Resolution – Radius of Gyration / Packing Density



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

High Resolution – Atom Representation: Full-atom

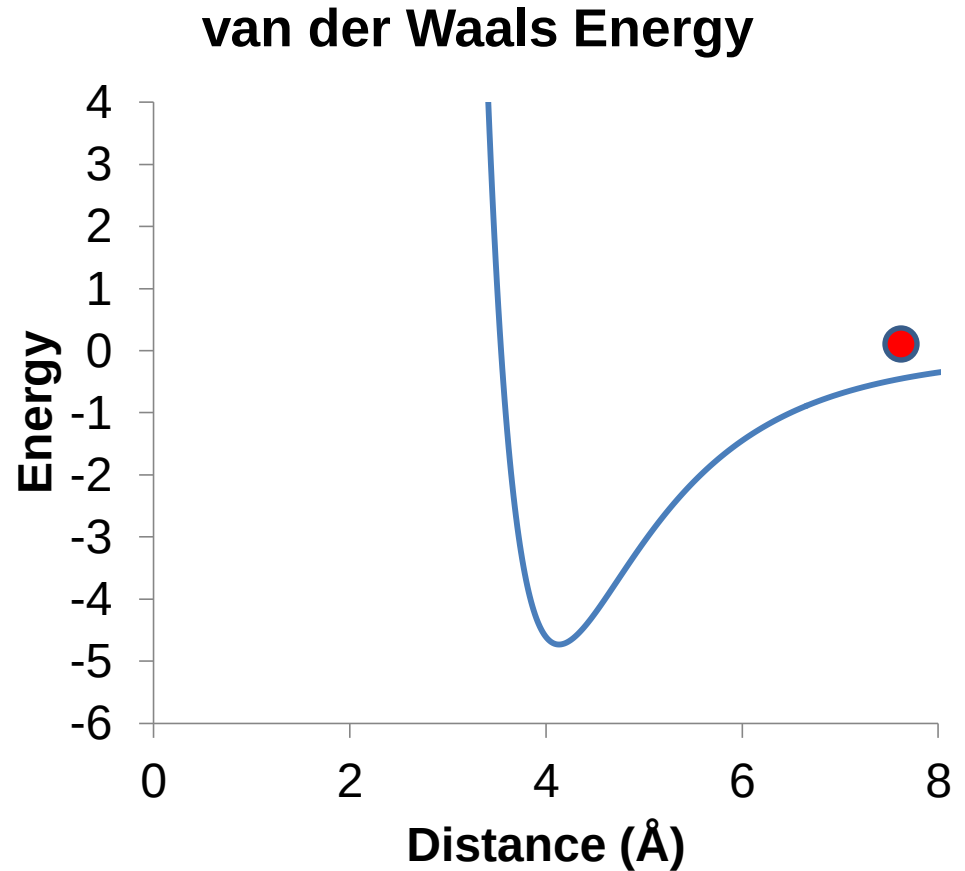
database/chemical/residue_type_sets/fa_standard



- All atoms represented
- Side-chain represented as rotamers

High Resolution – Van der Waals

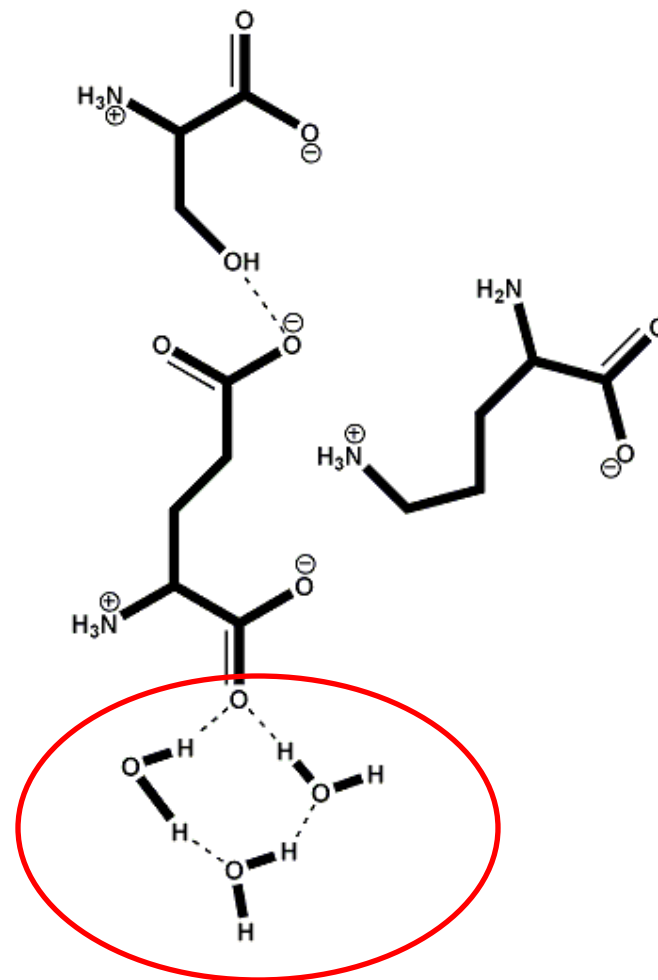
- `fa_atr / fa_rep / fa_intra_rep`
- 12-6 Lennard-Jones potential
- Lives in `src/core/scoring/etable/Etable.cc`



High Resolution – Solvation

- fa_sol
- Based on Lazaridis Karplus term
- Implicit term
- Divided into penalty and bonus energies

`src/core/scoring/etable/Etable.cc`



High Resolution – Electrostatics

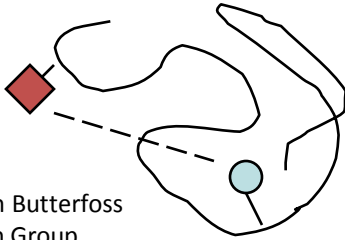
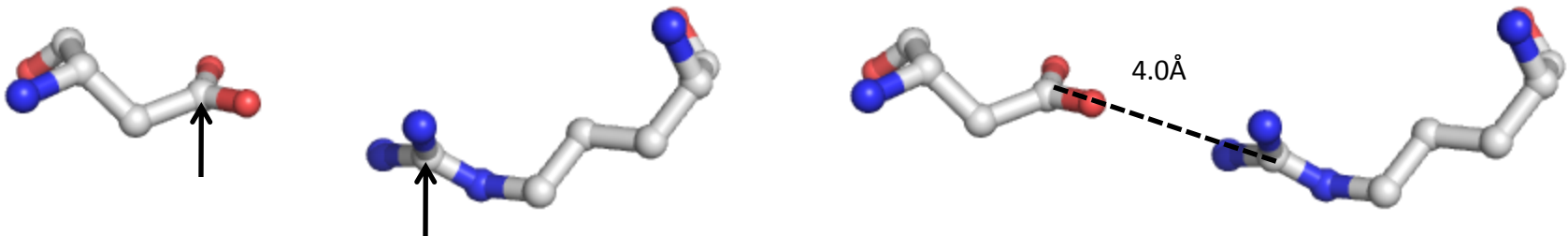


Image by: Glenn Butterfoss
–Rosetta Design Group

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

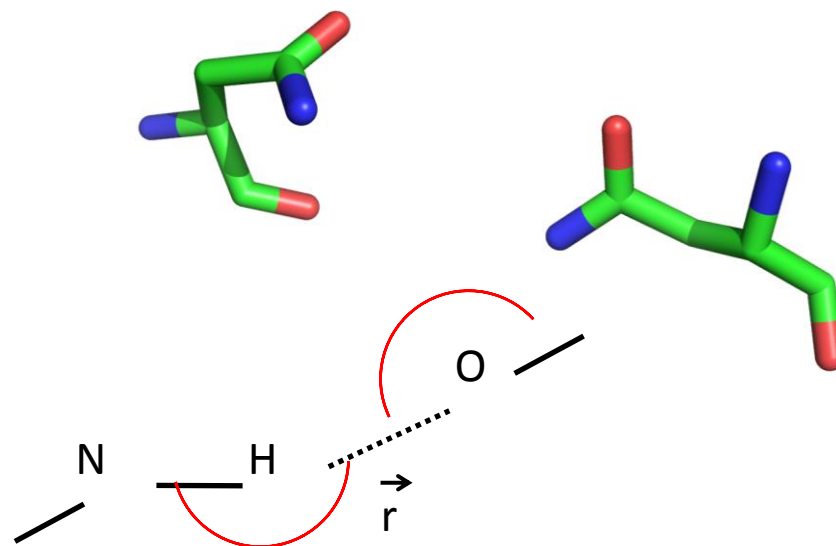
Action-Center



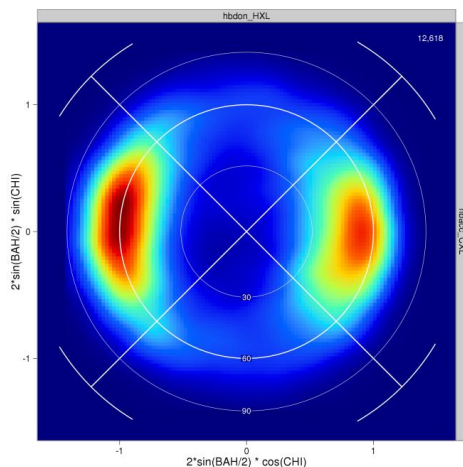
High Resolution – Hydrogen Bond

- Hbond_lr_bb / hbond_sr_bb / hbond_bb_sc / hbond_sc
- Geometry dependent
- Recently, inclusion of extra chi angle improved geometries

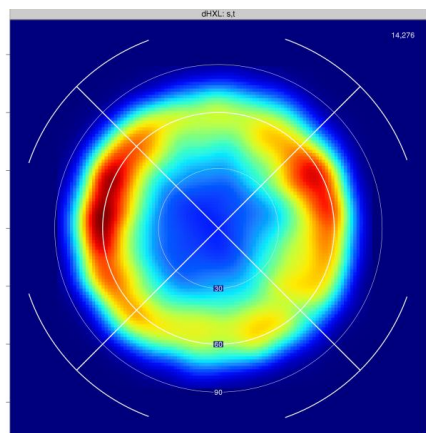
src/core/scoring/hbonds/HbondEnergy.cc



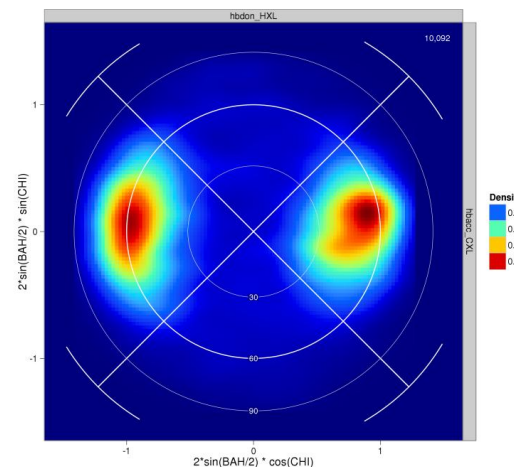
Crystal Structure



Score12



talaris2013

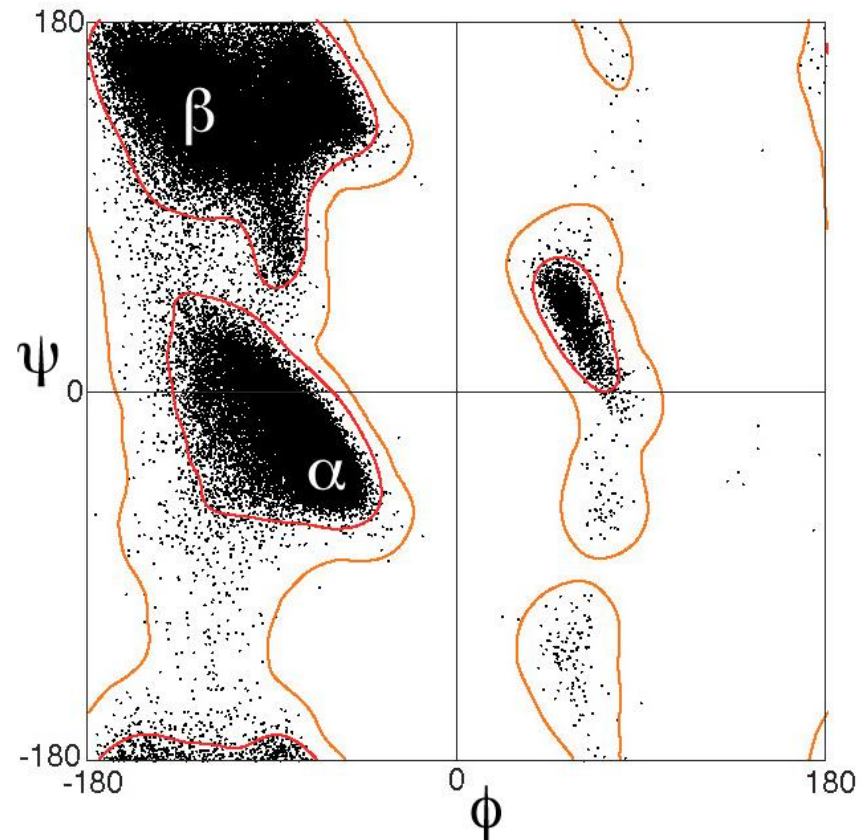


High Resolution – Ramachandran and Phi Psi angles

- rama- ramachandran preference
- p_aa_pp- probability at phi, psi

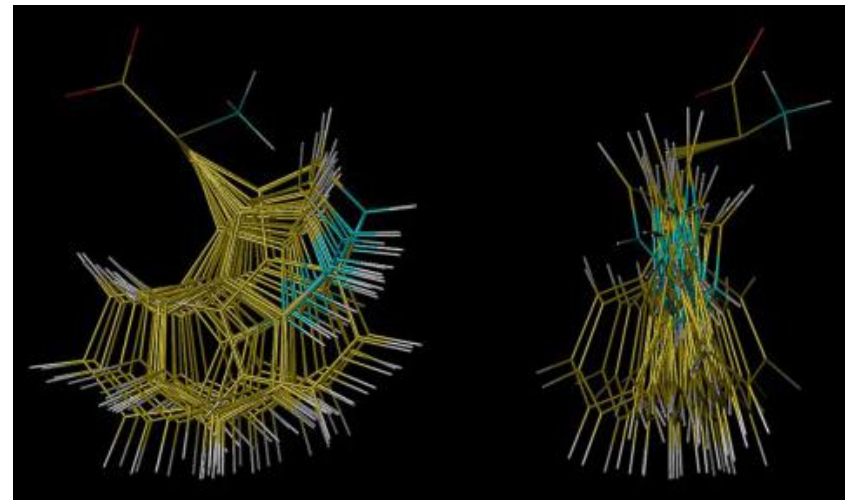
src/core/scoring/Ramachandran.cc

src/core/scoring/P_AA.cc



High Resolution – Dunbrack Rotamer Energy

- fa_dun
- Internal energy of a given rotamer found in PDB
- In Talaris2013, library has been updated
 - Chi distributions improve
 - Rotamer Recovery improves
 - Decoy discrimination unharmed

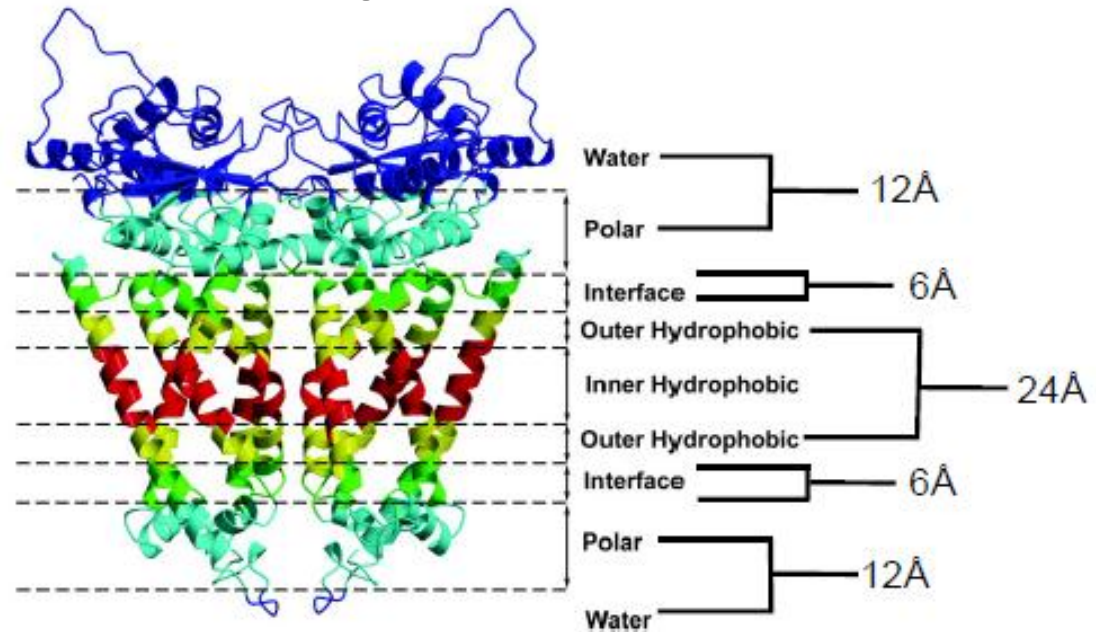


High Resolution – Reference Energy

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

Rosetta Membrane Specific Terms

Membrane center
and normal defined
by input TM span
information



Yarov-Yarovoy, Schonbrun, and Baker 2006.

Fa_mbsolv - replaces the solvation term with change in solvation energy of an atom upon burial. (IMM1) transfer of free energy from water to cyclohexane

Fa_mbenv – change in solvation free energy of isolated atoms when transferred. Deals with burial binning and layer placement of the residue

Rosetta Membrane Specific Terms

RosettaMembrane penalties

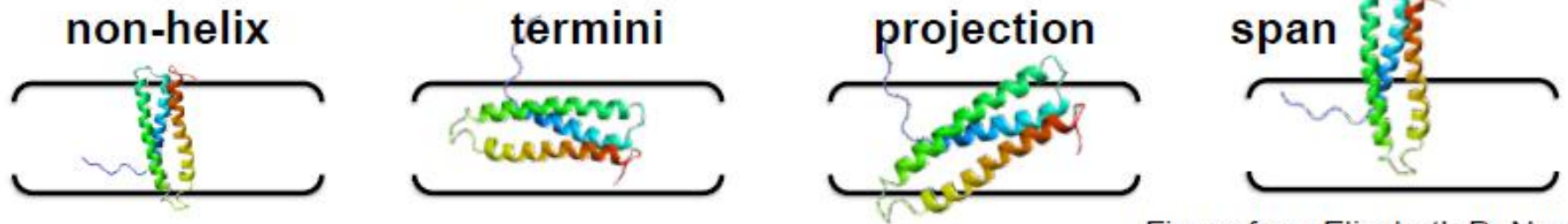


Figure from Elizabeth D. Nguyen

Menv_penalties:

- no non-helical secondary structure fragments in predicted TM helical regions in the hydrophobic layer of the membrane.
- no N- and C- termini residues of predicted TM helical regions in the hydrophobic layer of the membrane.
- no TM helices with orientation >45 degrees relative to the membrane normal vector.