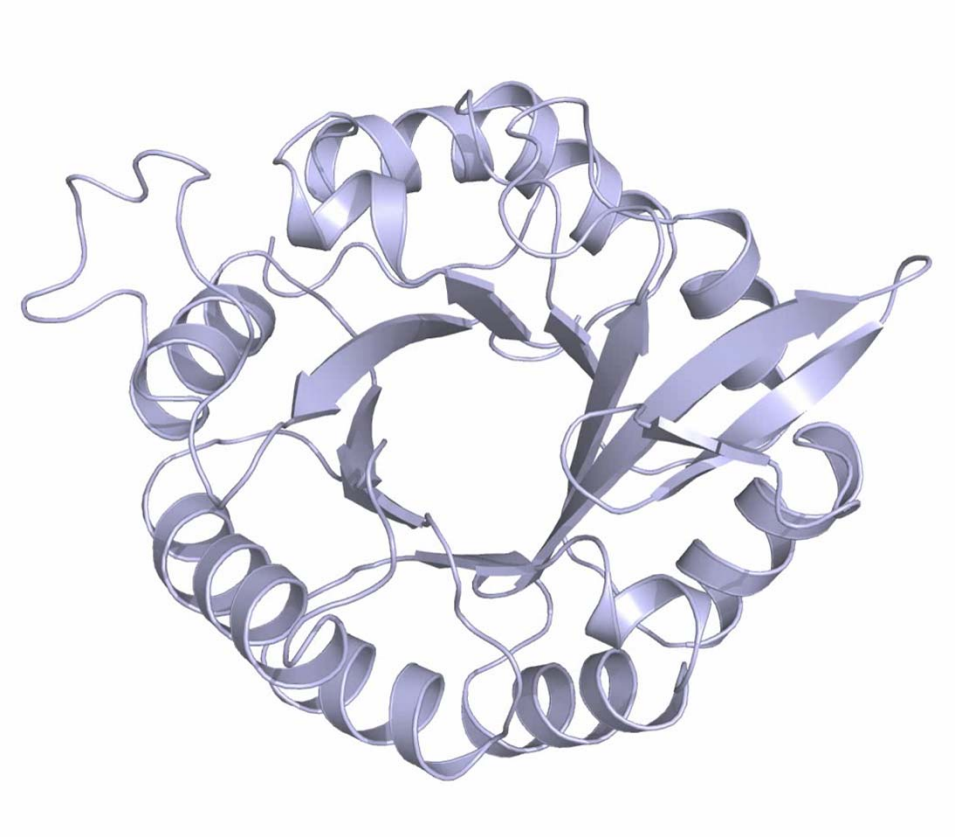
# Rosetta Scoring Function

Steven Combs

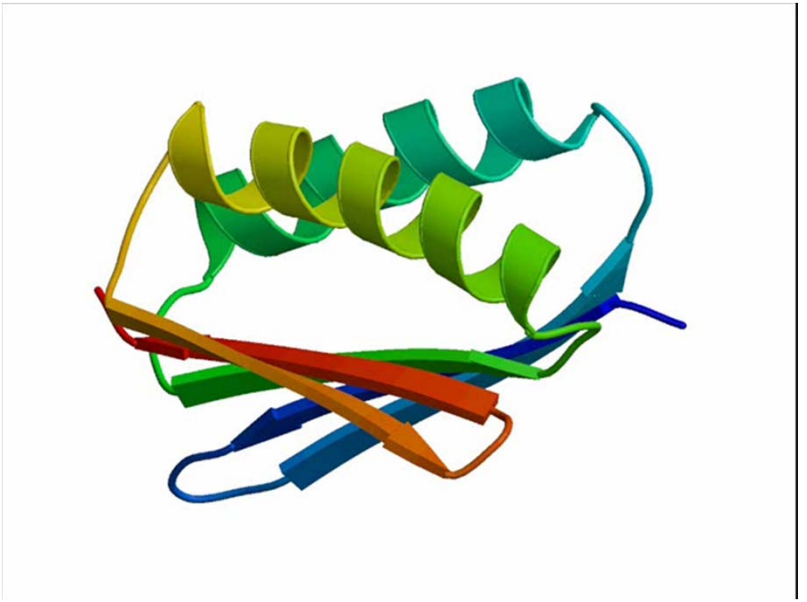
# Rosetta Software Suite is a Premier Software Suite

- 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

Movies by Jens Meiler and Steven Combs

# Common Computational Scoring Functions

## Molecular Mechanics

CHARMM / AMBER

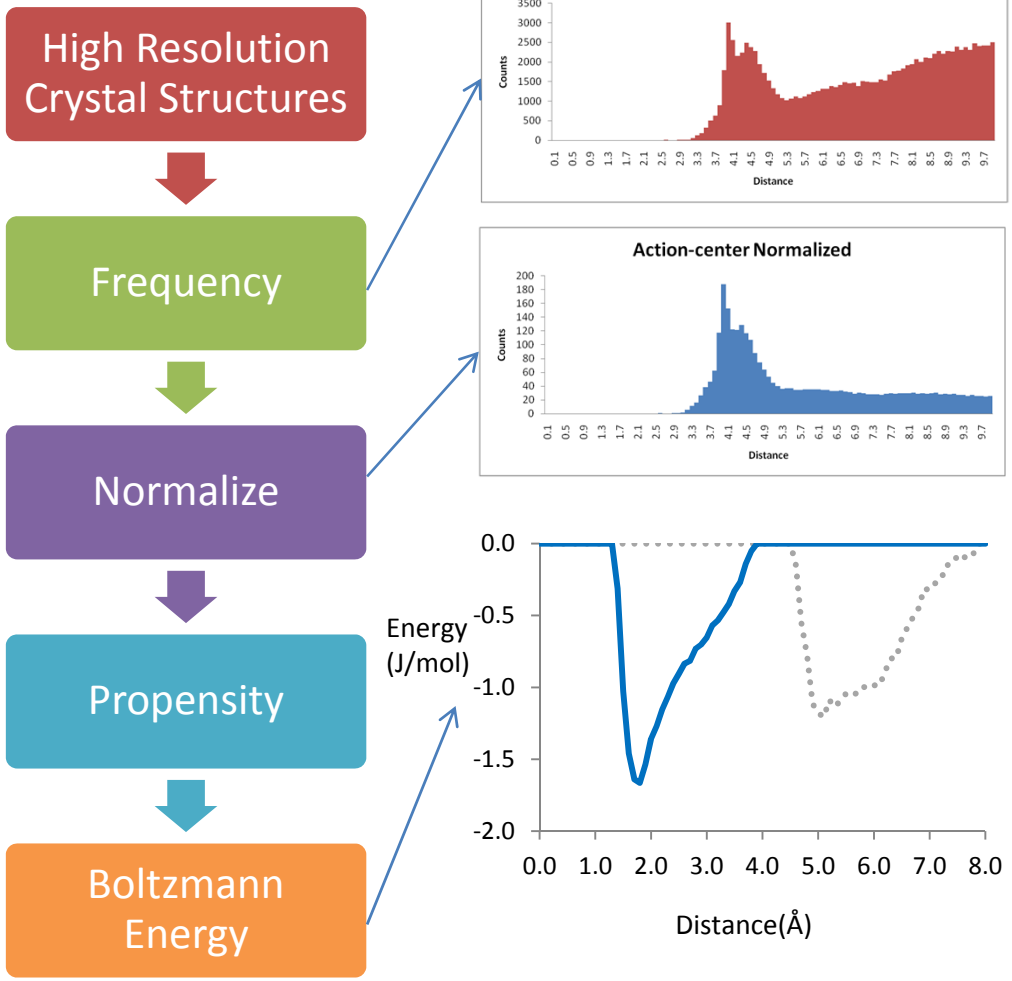
$$V_{total} = \sum_{bonds} K_r \left(r - r_{eq}\right)^2 + \sum_{angles} K_\theta \left(\theta - \theta_{eq}\right)^2 + \sum_{dihedrals} K_\phi \left(1 + \cos\left(n\phi\right)\right) + \sum_{i<j} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right] + \sum_{i<j} \left[\frac{q_i q_j}{\varepsilon r_{ij}}\right]$$

*Van der Waals*

*Electrostatics*

## Knowledge Based

Rosetta




*Energy = $w_1$*term$_1$ + $w_2$*term$_2$ + …*

# Rosetta Scoring Function Classes

**Major Classes:**

**1. Low resolution:**

Reduced atom representation (centroid)

Simple energy function

Aggressively search conformational space

**2. High resolution:**

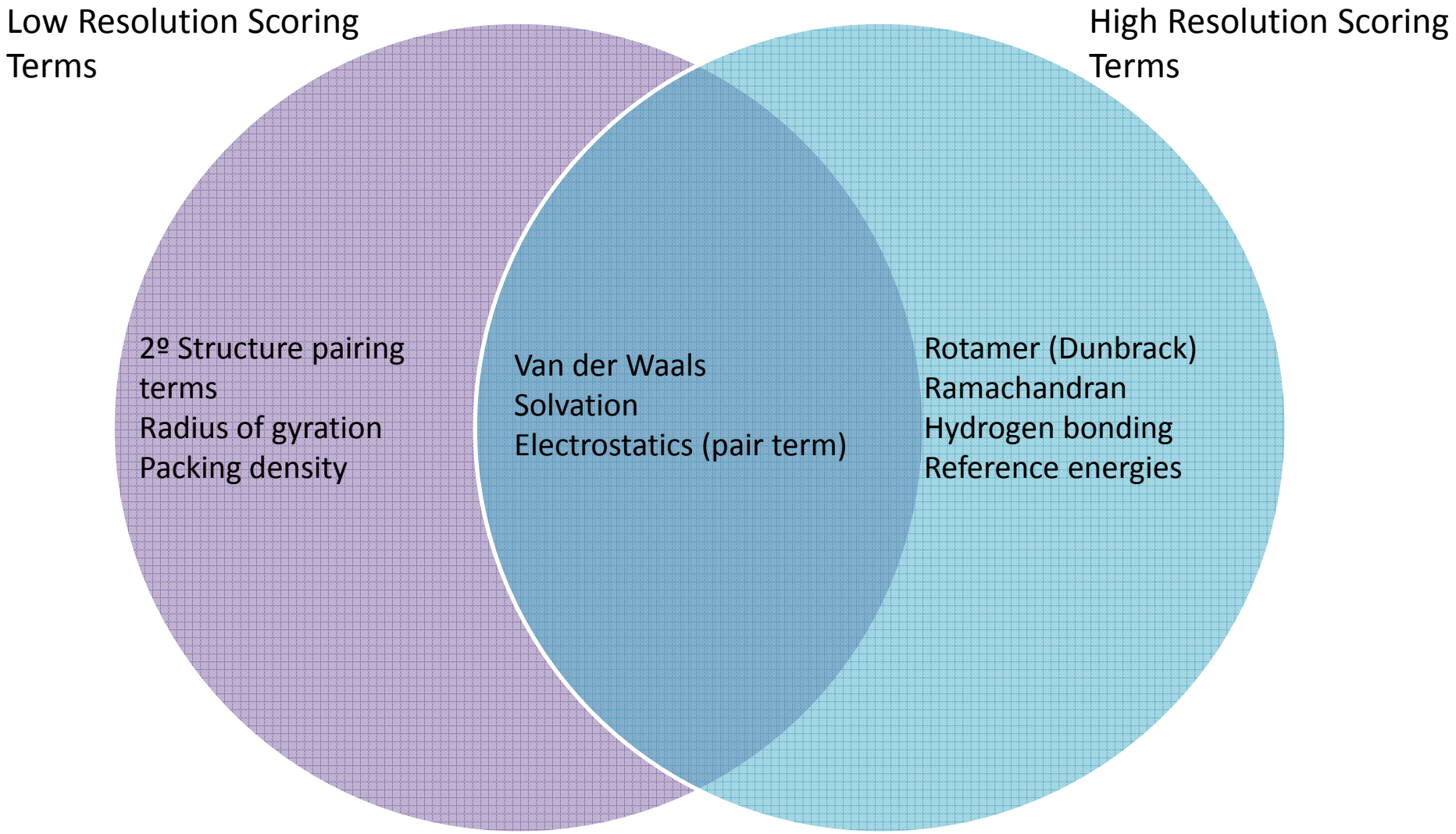
Full atom (FA)

More sophisticated energy function

“Local” search of conformational (and sequence) space

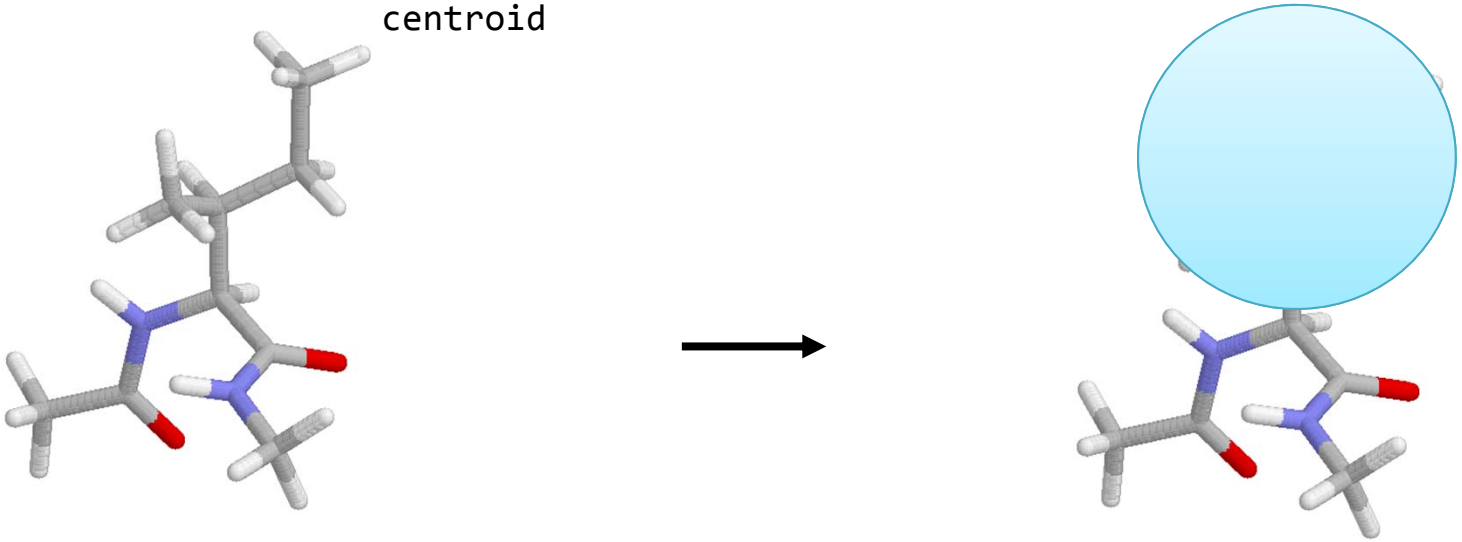
Slide by: Glenn Butterfoss
–Rosetta Design Group

# Common Score Types Between Low Resolution and High Resolution



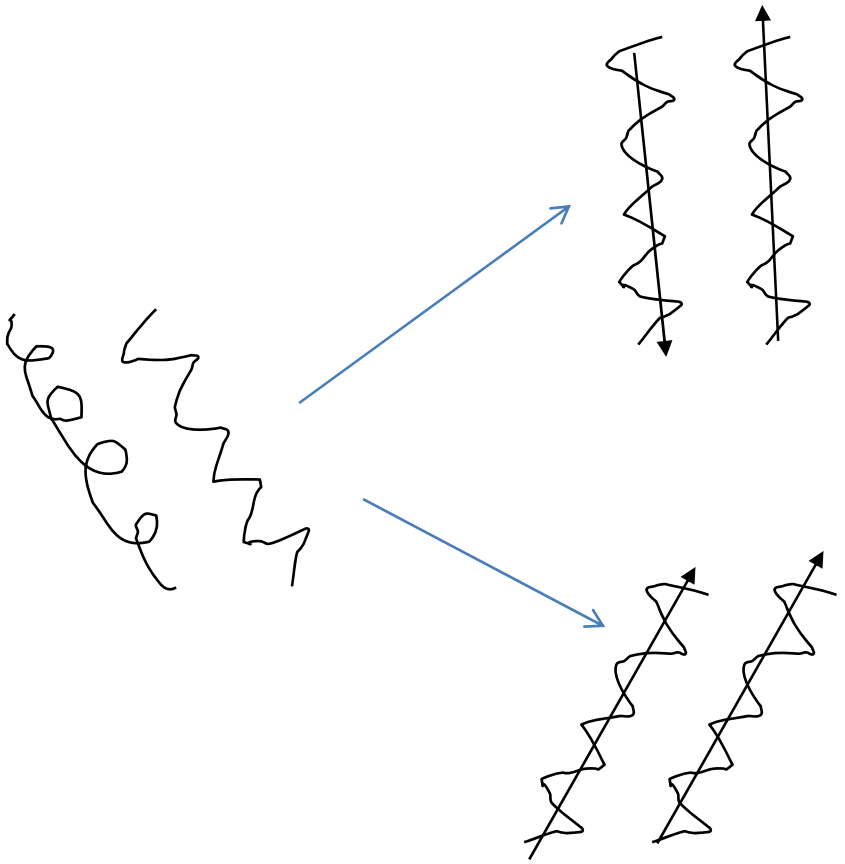
# Low Resolution – Atom Representation

`database/chemical/residue_type_sets/centroid`



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

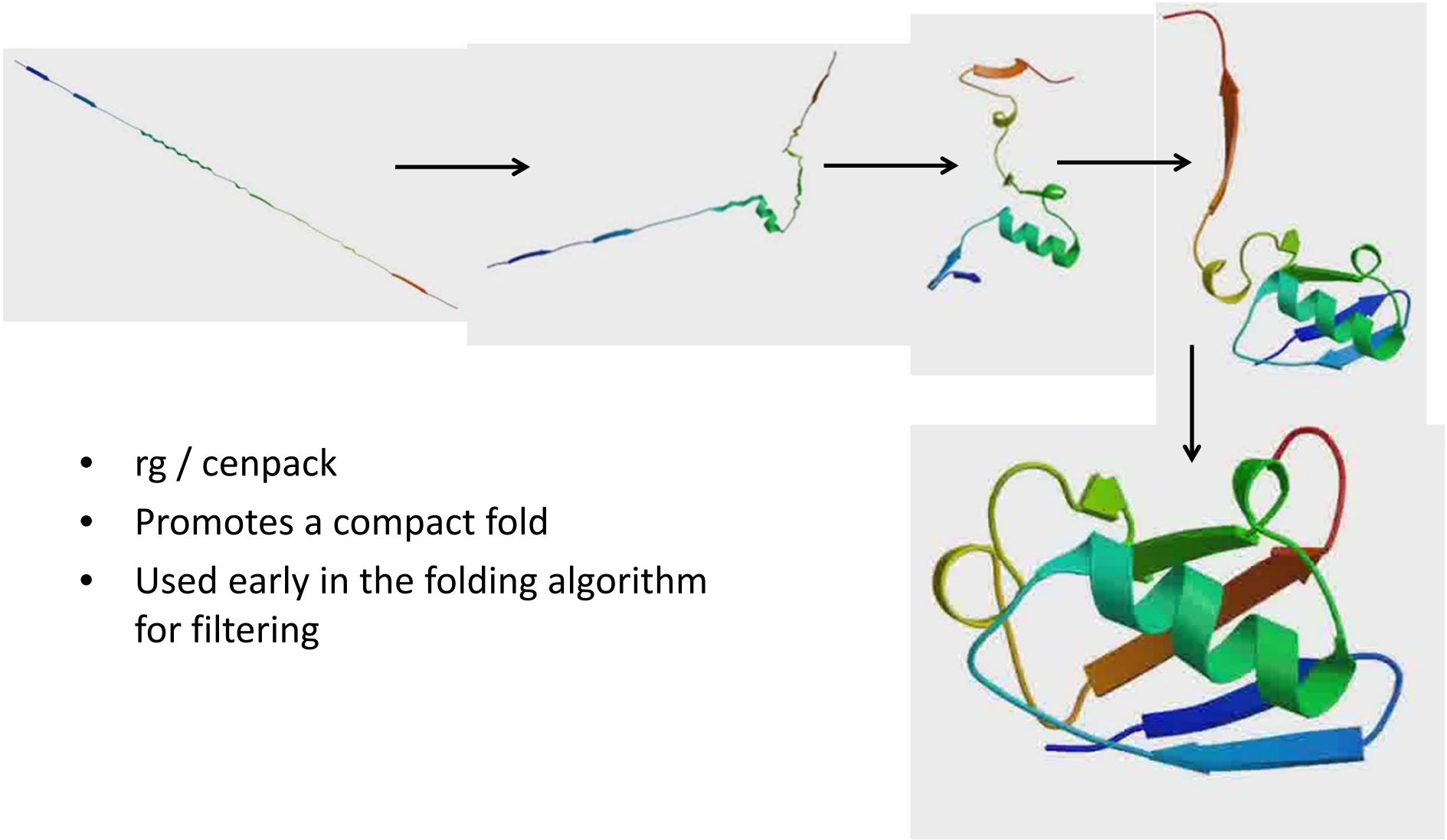
# Low Resolution - 2º Structure pairing terms



- `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

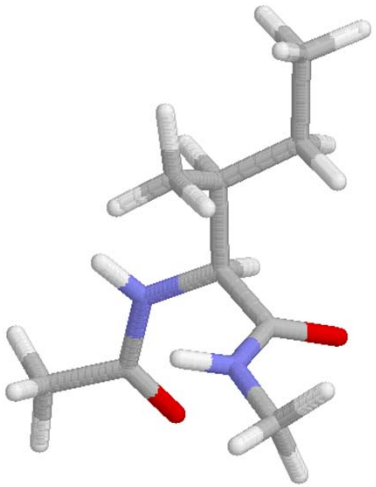
# Low Resolution – Radius of Gyration / Packing Density

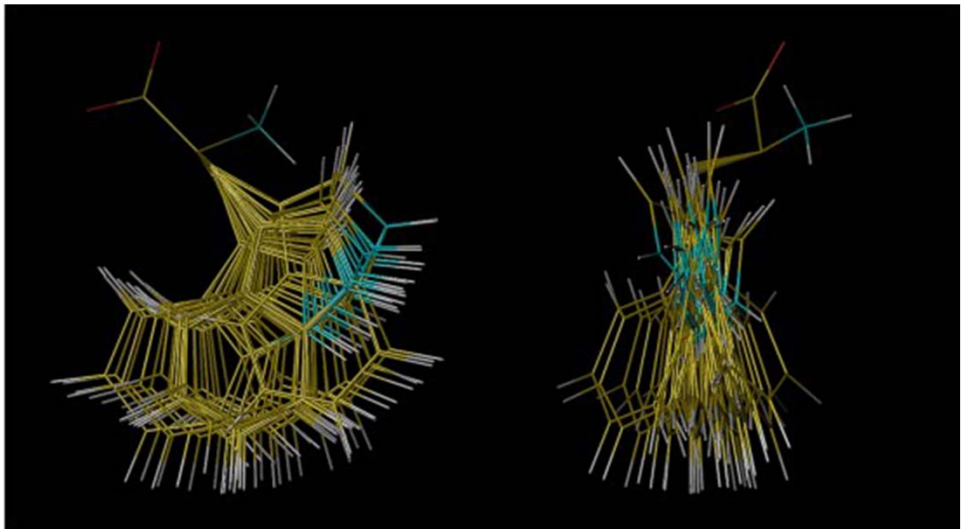


- rg / cenpack
- Promotes a compact fold
- Used early in the folding algorithm for filtering

# High Resolution – Atom Representation

```
database/chemical/residue_type_sets/fa_st
andard
```

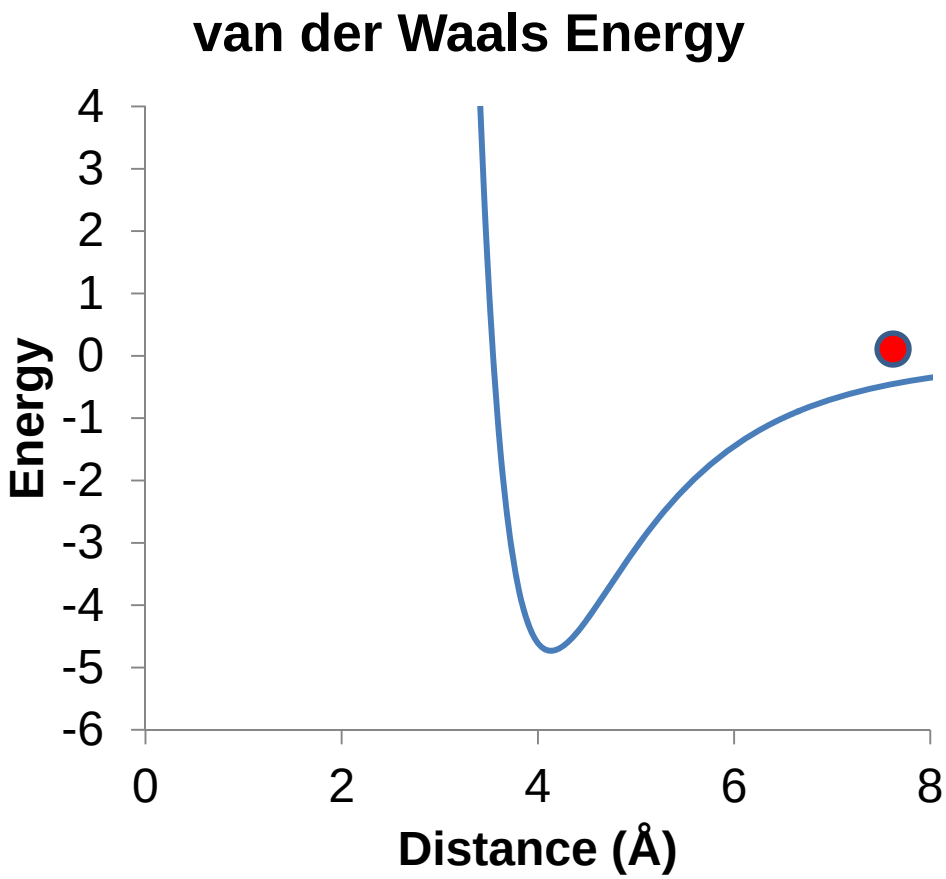




- All atoms represented
- Side-chain represented as rotamers

# 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`




Animation by:
Kristian Kaufmann

# High Resolution – Solvation

- `fa_sol`
- Based on Lazaridis Karplus term
- Implicit term
- Divided into penalty and bonus energies
- Lives in `src/core/scoring/etable/Etable.cc`

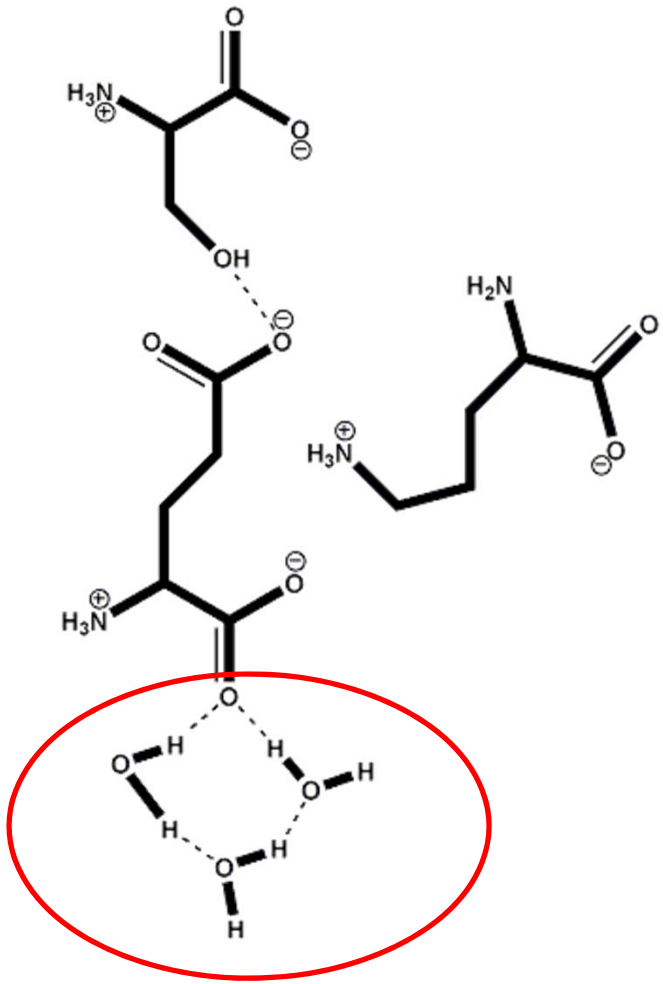


Image by:
Kristian Kaufmann

# High Resolution – Electrostatics

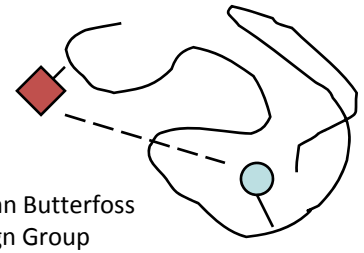
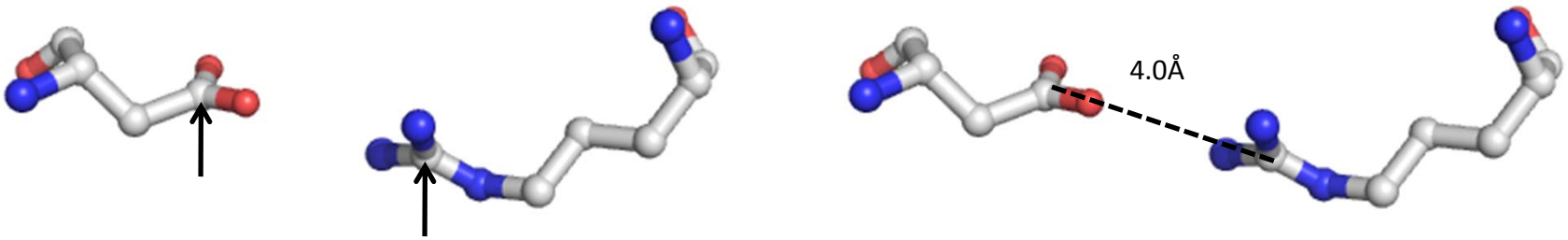


Image by: Glenn Butterfoss
–Rosetta Design Group

- Different applications have additional terms
  - Ligand -> Coulomb (hack_elec)
  - DNA -> General Born (gen_born)

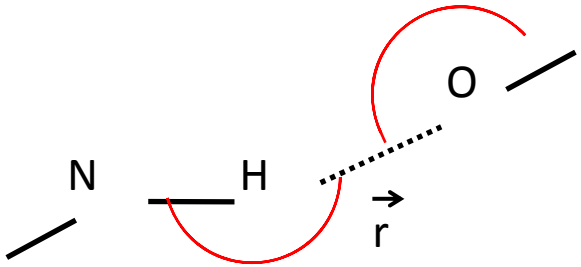
- `fa_pair`
- Probability of finding polar aa given a certain distance
  - Distance measured between action centers
- Lives in: `src/core/scoring/PairEPotential.cc`

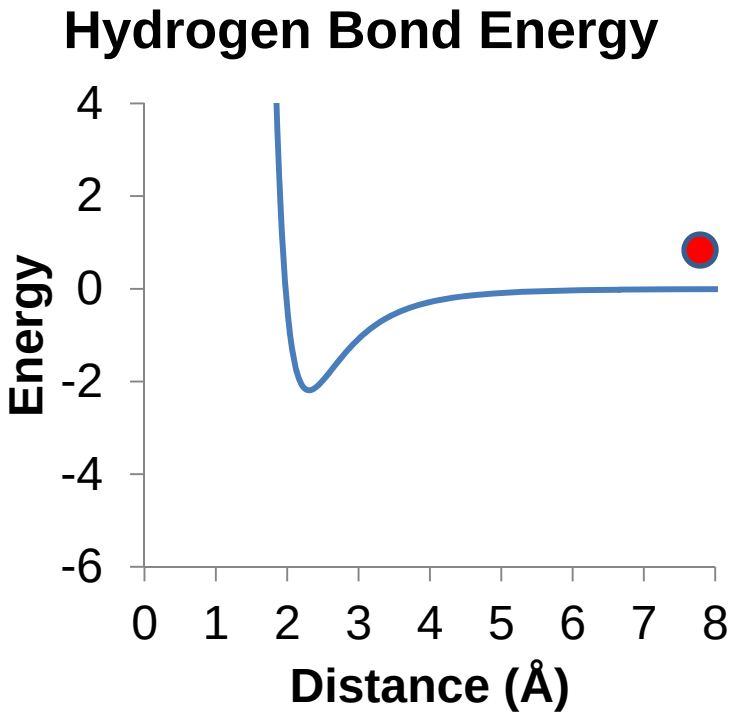
Action-Center

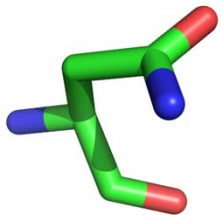


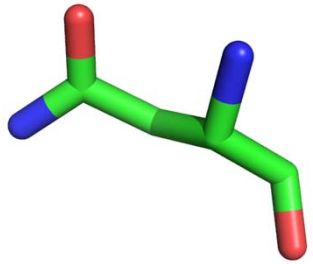
# High Resolution – Hydrogen Bond

- `Hbond_lr_bb / hbond_sr_bb / hbond_bb_sc / hbond_sc`
- Geometry dependent
  - 2 angles, 1 distance
- Lives in: `src/core/scoring/hbonds/HbondEnergy.cc`



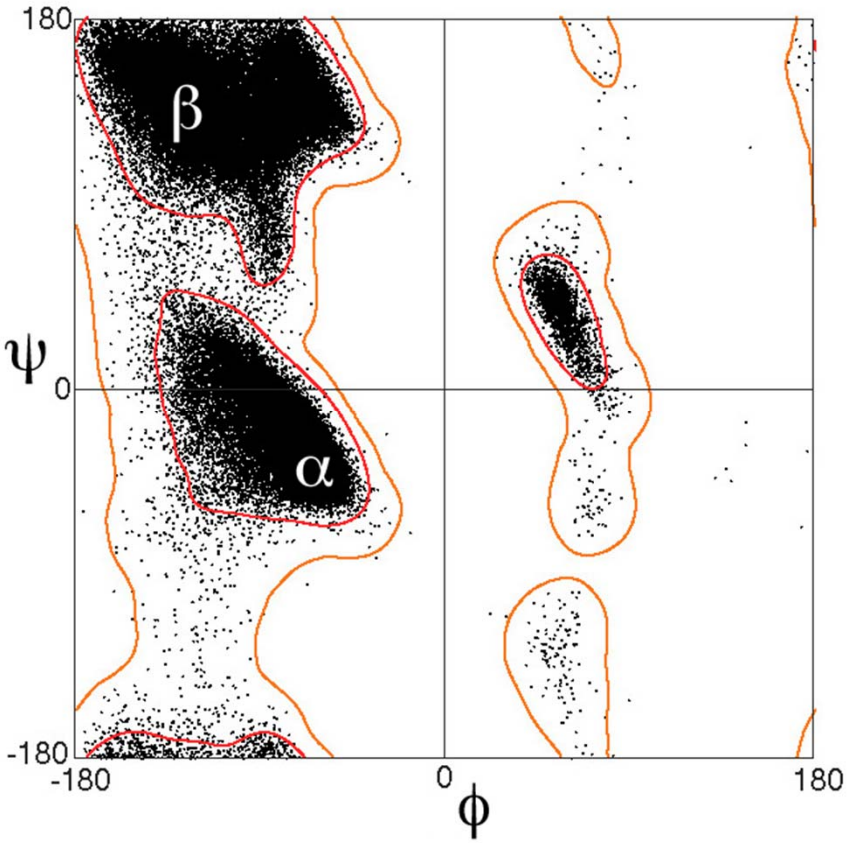





Animation by:
Kristian Kaufmann

# High Resolution – Ramachandran and Phi Psi angles

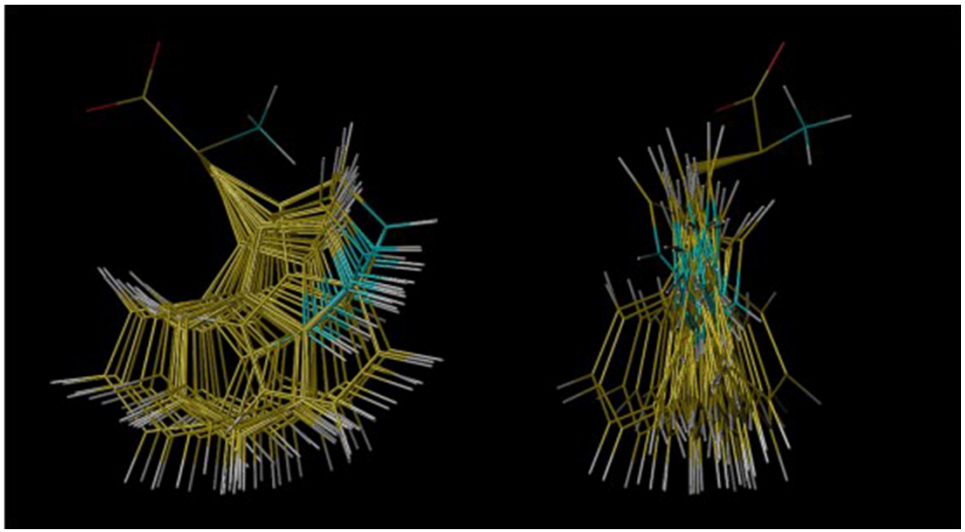
- `rama / p_aa_pp`
- Lives in:
  `src/core/scoring/Ramachandran.cc`
  `src/core/scoring/P_AA.cc`




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

# High Resolution – Dunbrack Rotamer Energy

- fa_dun
- Probability of a given rotamer found in PDB



# 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

# Where to Find Score Table

Score a protein

1 score_jd2.release –database <database> -s 1thfD.pdb -output

2 open 1thfD_0001.pdb in a txt editor

```
# All scores below are weighted scores, not raw scores.
#BEGIN_POSE_ENERGIES_TABLE 1thfD_0001
label fa_atr fa_rep fa_sol fa_intra_rep 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 total
weights 0.8 0.44 0.65 0.004 1 0.49 0.585 1.17 1.17 1.1 0.5 2 5 5 0.2
0.5 0.56 0.32 1 NA
pose -969.902 170.65 470.483 2.21791 1.29029 -20.756 -63.4763 -83.4728
-21.9036 -15.8397 0 0 0 0 -3.64882 15.9837 265.068 -24.2314 -49.34 -
326.878
MET_p:NtermProteinFull_1 -1.08126 0.0581536 0.871402 0.00966834 0 0 0 0
0 0 0 0 0 0 0 0 0.0943278 2.50593 0 -0.34 2.11822
LEU_2 -1.92514 0.187852 1.29337 0.00686622 0 0 0 0 0 0 0 0 0 0 -
0.267489 0.00337648 0.906386 -0.00995851 -0.1 0.0952614
```

Score terms

# Where to Find Score Table

```
# All scores below are weighted scores, not raw scores.
#BEGIN_POSE_ENERGIES_TABLE 1thfD_0001
label fa_atr fa_rep fa_sol fa_intra_rep 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 total
weights 0.8 0.44 0.65 0.004 1 0.49 0.585 1.17 1.17 1.1 0.5 2 5 5 0.2
0.5 0.56 0.32 1 NA
pose -969.902 170.65 470.483 2.21791 1.29029 -20.756 -63.4763 -83.4728
-21.9036 -15.8397 0 0 0 0 -3.64882 15.9837 265.068 -24.2314 -49.34 -
326.878
MET_p:NtermProteinFull_1 -1.08126 0.0581536 0.871402 0.00966834 0 0 0 0
0 0 0 0 0 0 0 0 0 0.0943278 2.50593 0 -0.34 2.11822
LEU_2 -1.92514 0.187852 1.29337 0.00686622 0 0 0 0 0 0 0 0 0 0 -
0.267489 0.00337648 0.906386 -0.00995851 -0.1 0.0952614
```

Weights

# Where to Find Score Table

```
# All scores below are weighted scores, not raw scores.
#BEGIN_POSE_ENERGIES_TABLE 1thfD_0001
label fa_atr fa_rep fa_sol fa_intra_rep 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 total
weights 0.8 0.44 0.65 0.004 1 0.49 0.585 1.17 1.17 1.1 0.5 2 5 5 0.2
0.5 0.56 0.32 1 NA
pose -969.902 170.65 470.483 2.21791 1.29029 -20.756 -63.4763 -83.4728
-21.9036 -15.8397 0 0 0 0 -3.64882 15.9837 265.068 -24.2314 -49.34 -
326.878
MET_p:NtermProteinFull_1 -1.08126 0.0581536 0.871402 0.00966834 0 0 0 0
0 0 0 0 0 0 0 0 0.0943278 2.50593 0 -0.34 2.11822
LEU_2 -1.92514 0.187852 1.29337 0.00686622 0 0 0 0 0 0 0 0 0 0 -
0.267489 0.00337648 0.906386 -0.00995851 -0.1 0.0952614
```

Total energies for score terms

Total energy for protein

# Where to Find Score Table

```
# All scores below are weighted scores, not raw scores.
#BEGIN_POSE_ENERGIES_TABLE 1thfD_0001
label fa_atr fa_rep fa_sol fa_intra_rep 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 total
weights 0.8 0.44 0.65 0.004 1 0.49 0.585 1.17 1.17 1.1 0.5 2 5 5 0.2
0.5 0.56 0.32 1 NA
pose -969.902 170.65 470.483 2.21791 1.29029 -20.756 -63.4763 -83.4728
-21.9036 -15.8397 0 0 0 0 -3.64882 15.9837 265.068 -24.2314 -49.34 -
326.878
```

Individual residue energies for score terms

```
MET_p:NtermProteinFull_1 -1.08126 0.0581536 0.871402 0.00966834 0 0 0 0
0 0 0 0 0 0 0 0 0.0943278 2.50593 0 -0.34 2.11822
LEU_2 -1.92514 0.187852 1.29337 0.00686622 0 0 0 0 0 0 0 0 0 0 0 -
0.267489 0.00337648 0.906386 -0.00995851 -0.1 0.0952614
```