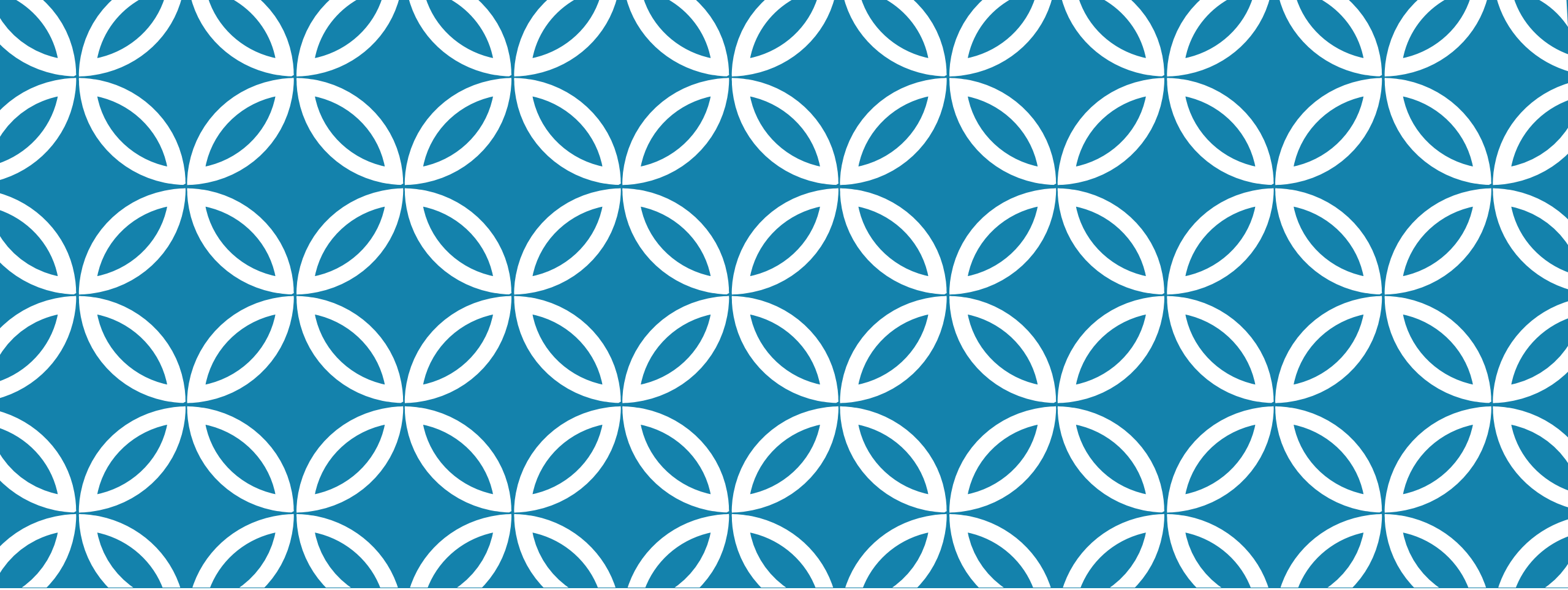


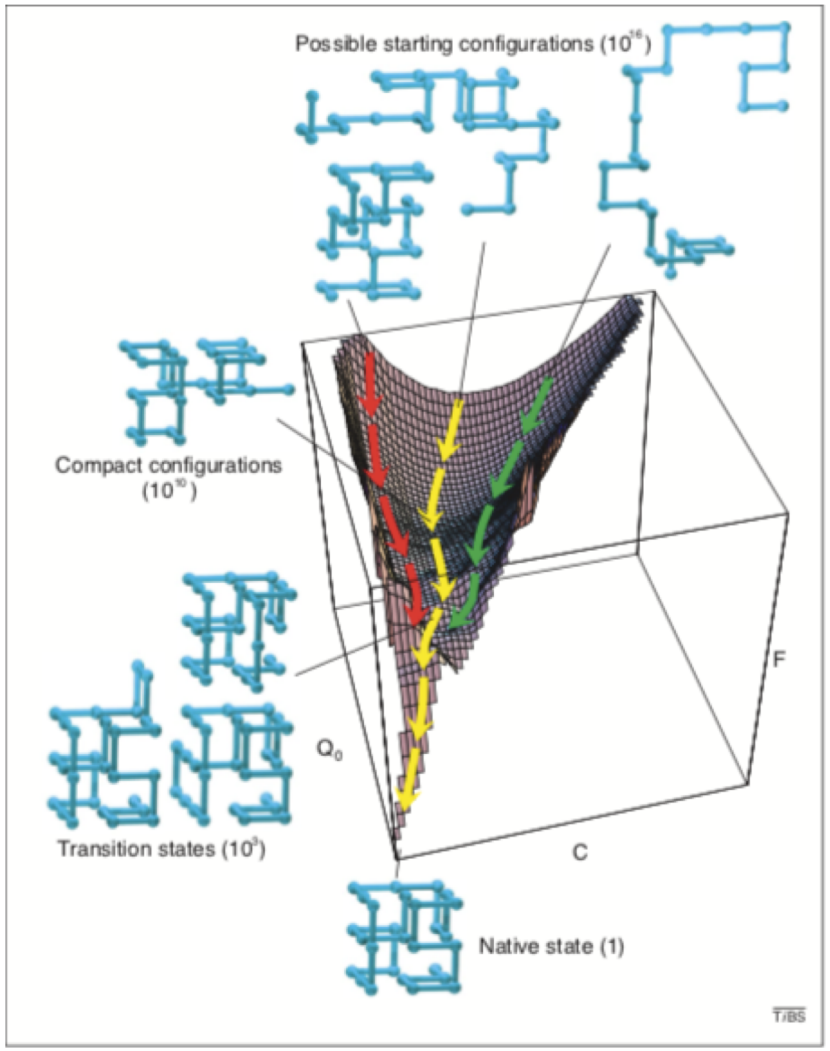
ROSETTA ENERGY FUNCTION

Alican Gulsevin, PhD

VU Rosetta Workshop

04/30/2019

PROTEIN FOLDING PROBLEM



- How to go to the "most stable" conformation?

- Levinthal's paradox: 100 amino acid peptide with 3 rotamers for each amino acid has 3^{198} wrong conformations.

- With random trial and error, folding of this protein would take billions of years… and yet, proteins in the nature can fold at sub-second scales.

Dinner AR, Šali A, Smith LJ, Dobson CM, and Karplus M (2000) Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem Sci* **25**:331–339

THEN, HOW TO FOLD PROTEINS?

- We do not know the rules of cooperativity the Nature has so we need a human-made algorithm.

- How can you fold proteins or antibodies from secondary structure within the timescale of a PhD or a postdoc?

- How can you modify the amino acids of a known protein structure and predict the resulting structure of the modified segment?

Answer: You use Rosetta. Rosetta uses the Monte Carlo method.

MONTE CARLO ALGORITHM

Idea: Collect favorable structural changes until you reach the global minimum.

1) Sampling: Randomly generate conformations to explore the conformational space.

2) Scoring: Somehow measure how favorable the generated structures are.

- Scores are calculated based on a score function.
- Decision to keep or discard a conformation is made based on the Metropolis algorithm.

METROPOLIS MONTE CARLO ALGORITHM

1) Move the structure to create a new conformation.

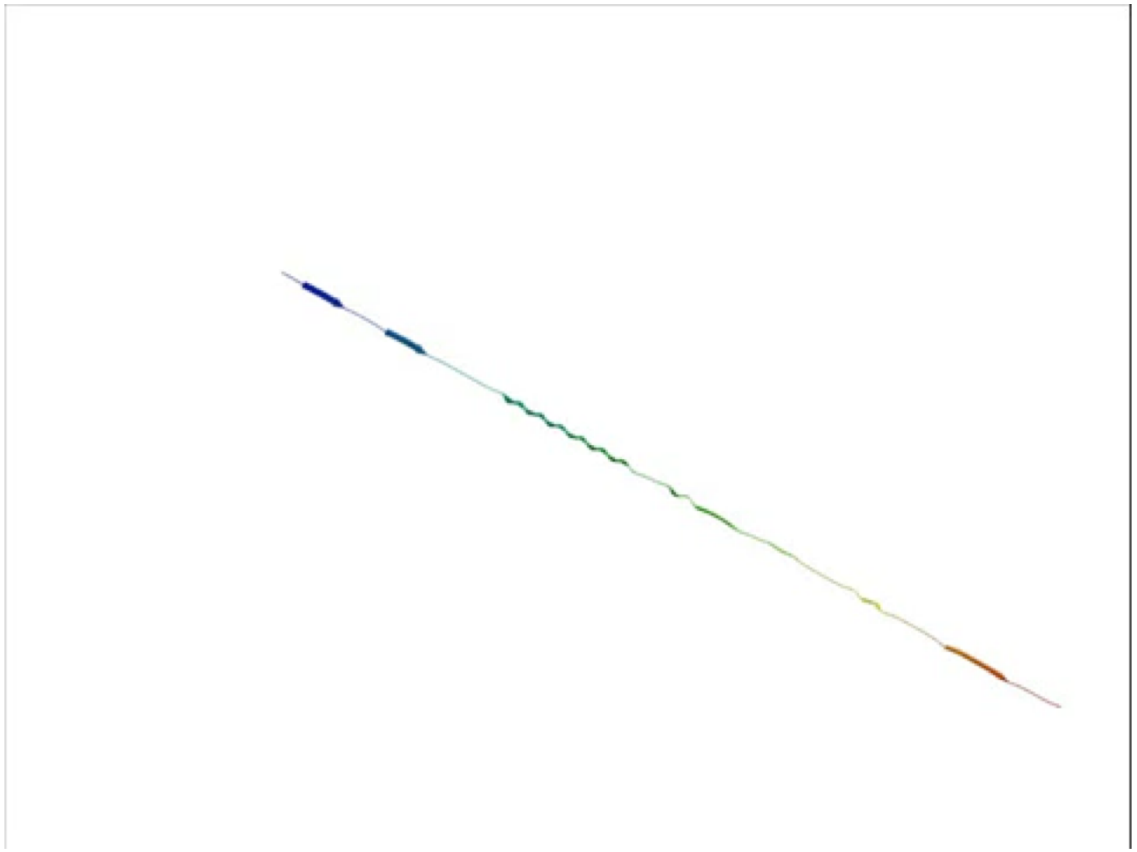
2) Calculate the ΔU as the energy of the $E_{\text{old conformation}} - E_{\text{new conformation}}$

3) If $\Delta U < 0$, then accept

If $\Delta U > 0$, then calculate $W = \exp(-\Delta U/kT)$. If $W >$ a random number R ($0 < R < 1$), then accept. Otherwise, reject.

In lay terms: If a structural change decreases the energy, keep the new conformation. If not, keep only if probability allows.

MONTE CARLO SIMULATION EXAMPLE



Video made by Jens Meiler

WHAT IS A SCORE FUNCTION?

- Score functions in the context of protein modeling consist of energy terms that are believed to represent protein interactions in a physical environment.

- The score terms can represent bonded, non-bonded, environmental, and statistical terms.

- The score terms can be given weights to adjust their contribution to the total energy of the system.

ROSETTA SCORE FUNCTION BASICS

- Rosetta score function has ***physical*** and ***statistical*** terms.

- Has one-body (i.e. within the same residue) and two-body (i.e between different residues) terms.

- The total scores are calculated as **weighted sum** of individual energy terms.

- Score units are Rosetta Energy Units (REU).

- Lower scores indicate more stable structures.

ROSETTA SCORE FUNCTION TYPES

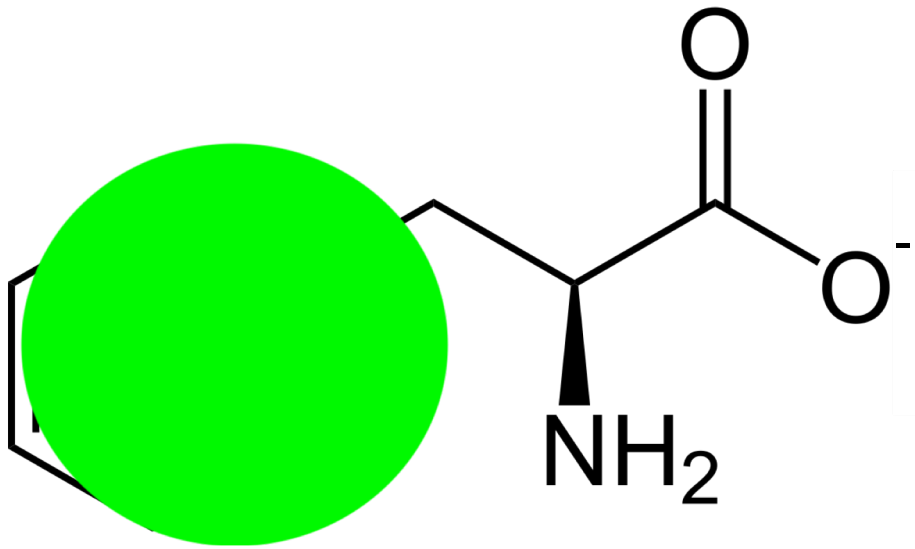
1) **Low-resolution score function:** Uses a simplified residue representation to quickly scan conformational space. Especially useful for *ab initio* folding and loop modeling.

2) **High-resolution score function:** Uses a full-atom residue representation to score conformational terms. More realistic than the low-resolution score function.

LOW-RESOLUTION SCORE FUNCTION

- Side chains are represented as a single “centroid” pseudo-atom.

- The location of the centroid is determined using known structures from PDB as the average location of the side-chain atoms of the same residue type.



Gray et al., J. Mol. Biol. (2003) 331, 281–299

LOW-RESOLUTION SCORE TERMS

Term	Description
env	Residue environment (solvation)
pair	Residue pair interactions (electrostatics, disulfides)
SS	Strand pairing (hydrogen bonding)
sheet	Helix arrangement into sheets
HS	Helix-strand packing
rg	Radius of gyration (vdW attraction, solvation)
cbeta	Cβ density
vdW	Steric repulsion

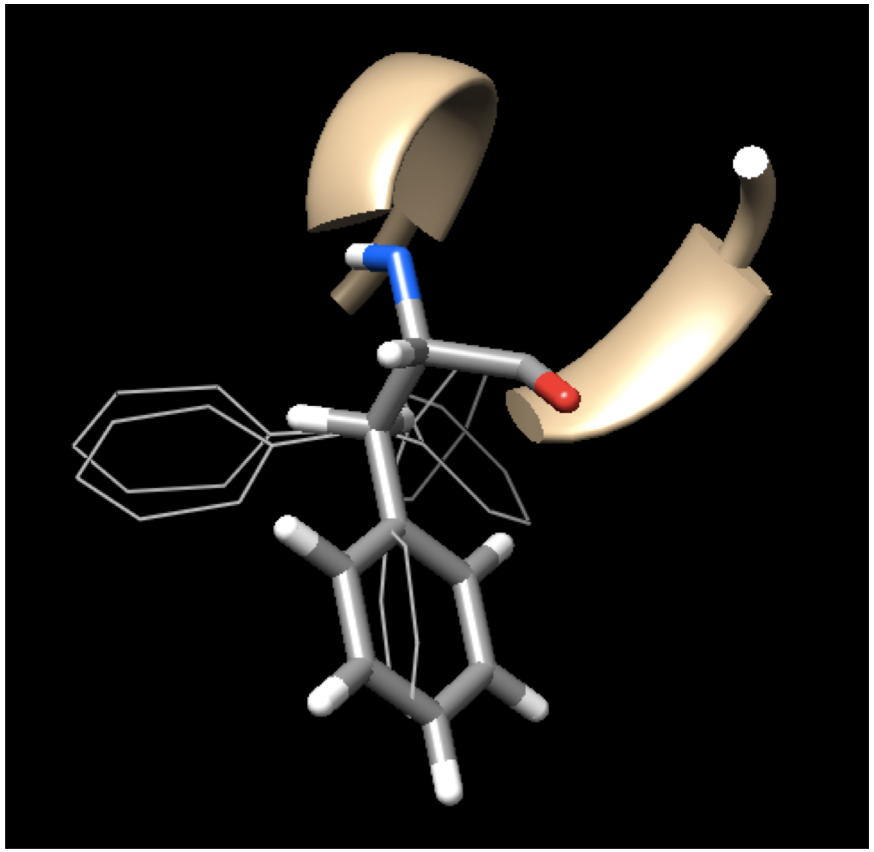
Rohl, C. A., Strauss, C. E. M., Misura, K. M. S., and Baker, D. B. T.-M. in E. (2004)
Protein Structure Prediction Using Rosetta. in *Numerical Computer Methods, Part D*, pp. 66–93, Academic Press, **383**, 66–93

HIGH-RESOLUTION SCORE FUNCTION

- Residues are represented at full atomic level.

- Side-chains are represented explicitly as rotamers and statistical terms are calculated to assess their quality.

- Has several terms to model different types of solvation, electrostatics, hydrogen bonding energies.



HIGH-RESOLUTION SCORE TERMS

Term	Description
fa_atr	Lennard-Jones attractive between atoms in different residues
fa_rep	Lennard-Jones repulsive between atoms in different residues
fa_sol	Lazaridis-Karplus solvation energy
fa_intra_sol_xover4	Intra-residue Lazaridis-Karplus solvation energy
lk_ball_wtd	Asymmetric solvation energy
fa_intra_rep	Lennard-Jones repulsive between atoms in the same residue
fa_elec	Coulombic electrostatic potential with a distance-dependent dielectric
pro_close	Proline ring closure energy and energy of psi angle of preceding residue
hbond_sr_bb	Backbone-backbone hbonds close in primary sequence
hbond_lr_bb	Backbone-backbone hbonds distant in primary sequence
hbond_bb_sc	Sidechain-backbone hydrogen bond energy
hbond_sc	Sidechain-sidechain hydrogen bond energy
dslf_fa13	Disulfide geometry potential
rama_prepro	Ramachandran preferences (with separate lookup tables for pre-proline positions and other positions)
omega	Omega dihedral in the backbone. A Harmonic constraint on planarity with standard deviation of ~6 degrees
p_aa_pp	Probability of amino acid, given torsion values for phi and psi
fa_dun	Internal energy of sidechain rotamers as derived from Dunbrack's statistics
yhh_planarity	A special torsional potential to keep the tyrosine hydroxyl in the plane of the aromatic ring
ref	Reference energy for each amino acid. Balances internal energy of amino acid terms. Plays role in design
METHOD_WEIGHTS	Not an energy term itself, but the parameters for each amino acid used by the ref energy term.

Source: https://rosettacommons.org/demos/latest/tutorials/scoring/scoring#scoring-in-rosetta Taken on: 04/17/2019

SCORE FUNCTION WEIGHTS

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

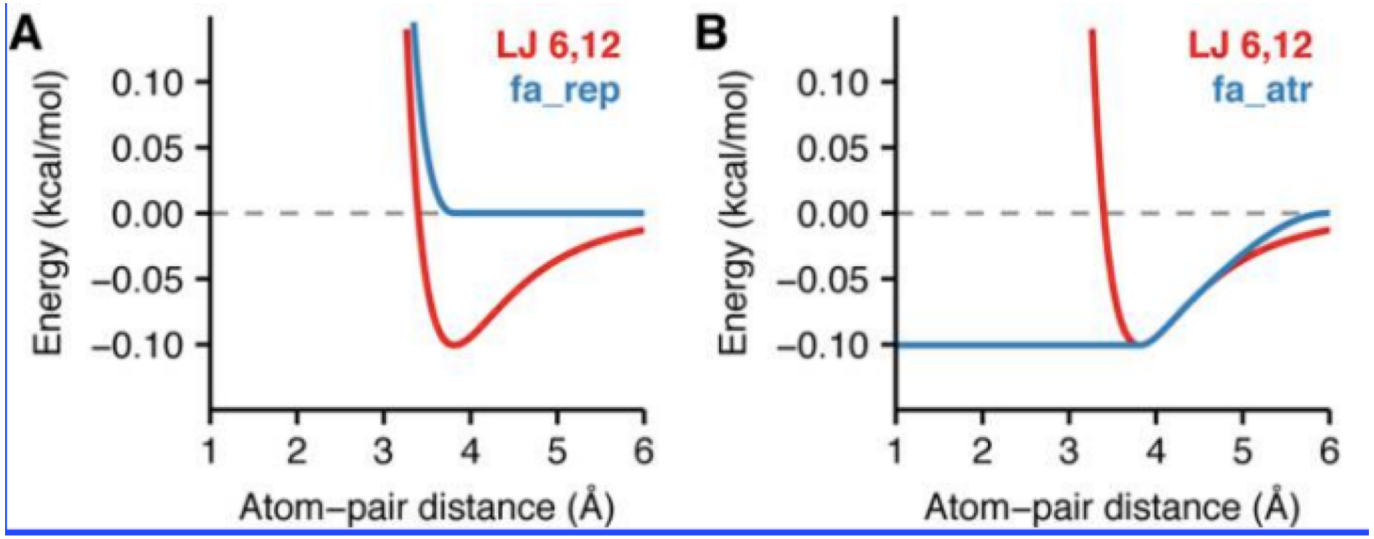
fa_atr 1 fa_rep 0.55 fa_sol 0.9375 fa_intra_rep 0.005 fa_elec 0.875 pro_close 1.25 hbond_sr_bb 1.17 hbond_lr_bb 1.17 hbond_bb_sc 1.17 hbond_sc 1.1 dslf_fa13 1.25 rama 0.25 omega 0.625 fa_dun 0.7 p_aa_pp 0.4 yhh_planarity 0.625 ref 1

VAN DER WAALS ATTRACTION/REPULSION

$$V_{\mathrm{LJ}} = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] = \varepsilon\left[\left(\frac{r_{\mathrm{m}}}{r}\right)^{12} - 2\left(\frac{r_{\mathrm{m}}}{r}\right)^{6}\right],$$

repulsive attractive

ε = energy well depths
r_m = sum of atom radii
r = atom-pair distance

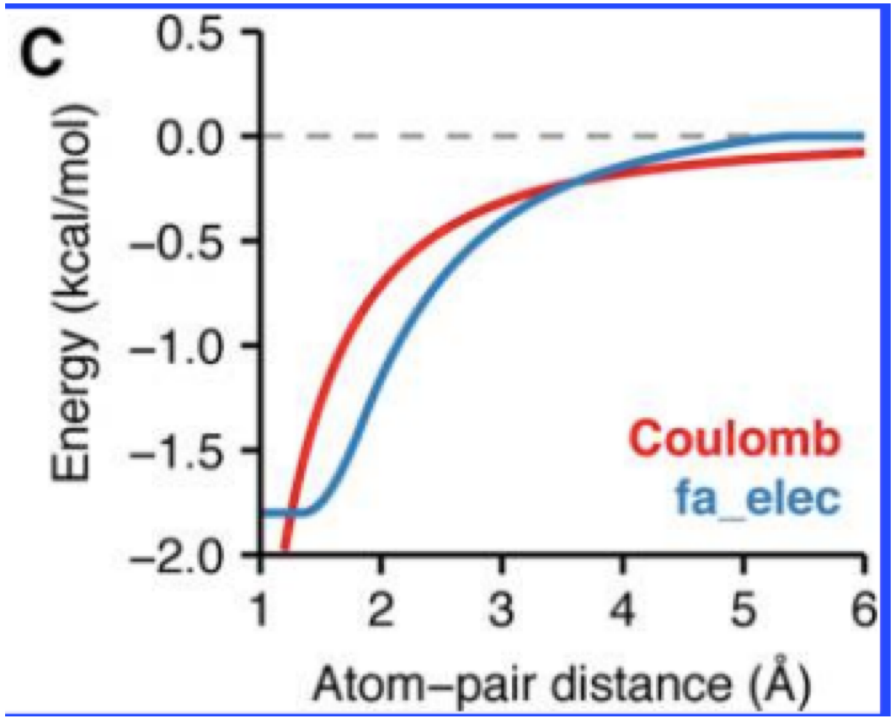


Alford, R. F., Leaver-Fay, A., Jeliazkov, J. R., O'Meara, M. J., DiMaio, F. P., Park, H., Shapovalov, M. V., Renfrew, P. D., Mulligan, V. K., Kappel, K., Labonte, J. W., Pacella, M. S., Bonneau, R., Bradley, P., Dunbrack, R. L., Das, R., Baker, D., Kuhlman, B., Kortemme, T., and Gray, J. J. (2017) The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **13**, 3031–3048

ELECTROSTATIC INTERACTIONS

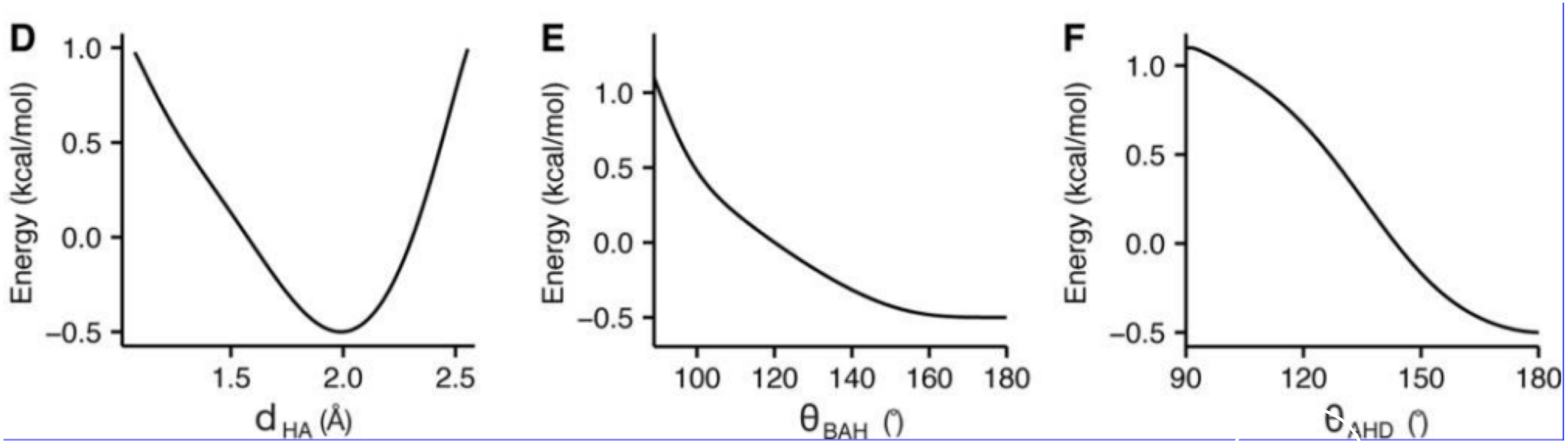
$$E_{\text{Coulomb}}(i, j) = \frac{C_0 q_i q_j}{\epsilon} \frac{1}{d_{i,j}}$$

q_i, q_j = charges of atoms i and j
ε = dielectric constant
d_{ij} = distance between atoms i and j



Alford, R. F., Leaver-Fay, A., Jeliazkov, J. R., O'Meara, M. J., DiMaio, F. P., Park, H., Shapovalov, M. V., Renfrew, P. D., Mulligan, V. K., Kappel, K., Labonte, J. W., Pacella, M. S., Bonneau, R., Bradley, P., Dunbrack, R. L., Das, R., Baker, D., Kuhlman, B., Kortemme, T., and Gray, J. J. (2017) The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **13**, 3031–3048

HYDROGEN BONDING



Acceptor – donor distance

Base-acceptor-hydrogen angle

Acceptor-hydrogen-donor angle

Alford, R. F., Leaver-Fay, A., Jeliazkov, J. R., O'Meara, M. J., DiMaio, F. P., Park, H., Shapovalov, M. V., Renfrew, P. D., Mulligan, V. K., Kappel, K., Labonte, J. W., Pacella, M. S., Bonneau, R., Bradley, P., Dunbrack, R. L., Das, R., Baker, D., Kuhlman, B., Kortemme, T., and Gray, J. J. (2017) The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **13**, 3031–3048

REFERENCE ENERGIES

- Unfolded state energies of individual amino acids.

- Residues with a large gap between folded and unfolded states are less likely to be found in proteins.

- Helps adjusting frequencies of amino acids during protein design.

$$E_{\text{ref}} = \sum_i \Delta G_i^{\text{ref}}(\text{aa}_i)$$

Alford, R. F., Leaver-Fay, A., Jeliazkov, J. R., O'Meara, M. J., DiMaio, F. P., Park, H., Shapovalov, M. V., Renfrew, P. D., Mulligan, V. K., Kappel, K., Labonte, J. W., Pacella, M. S., Bonneau, R., Bradley, P., Dunbrack, R. L., Das, R., Baker, D., Kuhlman, B., Kortemme, T., and Gray, J. J. (2017) The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **13**, 3031–3048

STATISTICAL TERMS

Statistical terms are calculated based on the probability of finding certain rotamers/dihedrals in experimental structures deposited to PDB.

- Backbone dihedral angles
- Side-chain dihedral angles
- Side-chain rotamer probability

ADDITIONAL SCORE TERMS

Rosetta also has additional score terms for both the low- and high-resolution score functions for more specialized applications. The additional terms can be used for several purposes:

1) To model non-protein biomolecules such as DNA or RNA.

2) To add constraints based on experimental data such as SAXS, NMR, and cryo-EM experiments to guide refinement procedures.

3) To add specific interactions that are not part of the default score functions.

SCORE FUNCTIONS CAN CHANGE

- Score functions are re-evaluated over time to add or adjust energy terms based on the emergence of new structural data.

- The current high-resolution score function in Rosetta is ref2015.

- Examples to previous score functions are talaris2013 and score12. These score functions are still used for some applications because they perform better than ref2015.

MANUALLY CHANGING THE SCORE FUNCTION

1) **Custom weights file:** A weights file with a desired set of weights for different terms can be inputted by the user. Example: your system environment is somehow significantly different than the assumptions made to generate the existing score functions.

2) **Patch file to modify existing weights:** A weights file can be used to "patch" the terms of an existing score function. Example: you want to re-scale the electrostatic terms without changing the rest of the term weights.

3) **Set weights to specific terms from the command line:** Changes to individual terms can be done using the command line instead of providing an external weights file. Example: you want to apply changes *ad hoc* to individual terms on the go.

SPECIALIZED SCORE FUNCTIONS

Membrane score function(s): Rosetta has low- and high-resolution membrane scoring functions to model proteins in a membrane environment. These functions have additional terms to include residue-membrane interactions and changes to non-bonded term weights to better represent the membrane environment.

Orbital score function: Uses an orbital representation to calculate partial covalent interactions such as hydrogen bonds, π-cation, π–π interactions, and salt bridges more accurately.

Barth P, Schonbrun J, Baker D. (2007 Oct 2) Toward high-resolution prediction and design of transmembrane helical protein structures. Proc Natl Acad Sci U S A. 104(40):15682-7

Yarov-Yarovoy V, Schonbrun J, Baker D. (2006 Mar 1) Multipass membrane protein structure prediction using Rosetta. Proteins. 62(4):1010-25.

Combs, S. A., Mueller, B. K., and Meiler, J. (2018) Holistic Approach to Partial Covalent Interactions in Protein Structure Prediction and Design with Rosetta. *J. Chem. Inf. Model.* **58**, 1021–1036

SCORING WITH NON-PROTEIN RESIDUES

- If parameters for a molecule do not exist, you need to create parameters files to define the new atom types and conformers of the new molecule.

- The new atom types and properties are used to calculate the energy terms with the existing weights of the score function.

- Parameters for some D-amino acids, non-canonical amino acids, peptoids, sugars, DNA, and RNA are present in the Rosetta database under database/chemical/residue_type_sets.

Drew, K., Renfrew, P. D., Craven, T. W., Butterfoss, G. L., Chou, F. C., Lyskov, S., Bullock, B. N., Watkins, A., Labonte, J. W., Pacella, M., Kilambi, K. P., Leaver-Fay, A., Kuhlman, B., Gray, J. J., Bradley, P., Kirshenbaum, K., Arora, P. S., Das, R., and Bonneau, R. (2013) Adding Diverse Noncanonical Backbones to Rosetta: Enabling Peptidomimetic Design. *PLoS One*.

SUMMARY

- Rosetta score functions consist of a number of physical and statistical weighted energy terms.

- The score of a pose is a relative measure of its stability rather than a global metric such as free energy.

- Low-resolution (centroid) score function is useful when exploring conformational space with large backbone motions whereas high-resolution (full atom) score function is useful for applications involving side-chain motions.

- The term weights of the score functions can be modified (though not suggested) or additional terms can be included to account for specific interactions.

- Scoring non-protein species or unnatural amino acids is possible, but requires additional work and care.

ADDITIONAL RESOURCES

Rosetta score function overview

Alford, R. F., Leaver-Fay, A., Jeliazkov, J. R., O'Meara, M. J., DiMaio, F. P., Park, H., Shapovalov, M. V., Renfrew, P. D., Mulligan, V. K., Kappel, K., Labonte, J. W., Pacella, M. S., Bonneau, R., Bradley, P., Dunbrack, R. L., Das, R., Baker, D., Kuhlman, B., Kortemme, T., and Gray, J. J. (2017) The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **13**, 3031–3048

Rosetta scoring tutorial:

https://rosettacommons.org/demos/latest/tutorials/scoring/scoring#scoring-in-rosetta

Additional terms

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

Membrane scoring

https://www.rosettacommons.org/docs/latest/application_documentation/membrane_proteins/RosettaMP-GettingStarted-Overview

Orbital scoring

https://www.rosettacommons.org/docs/latest/rosetta_basics/scoring/NC-scorefunction-info#Partial-Covalent-Interactions-Energy-Function-(Orbitals)

Non-protein residue modeling

https://www.rosettacommons.org/docs/latest/rosetta_basics/non_protein_residues/non-protein-residues

Pyrosetta scoring examples

https://graylab.jhu.edu/pyrosetta/downloads/documentation/Workshop3_PyRosetta_Scoring.pdf