Structure-based mutation analysis

Practical Bioinformatics 'Protein Structure and Function Analysis'

June 19, 2012

Cedric Staniewski   Julia Rackerseder

Overview

- Introduction

- Protein structure

- Molecular mechanics / force fields

- Methods
  - FOLDX
  - SCWRL4

Protein structure
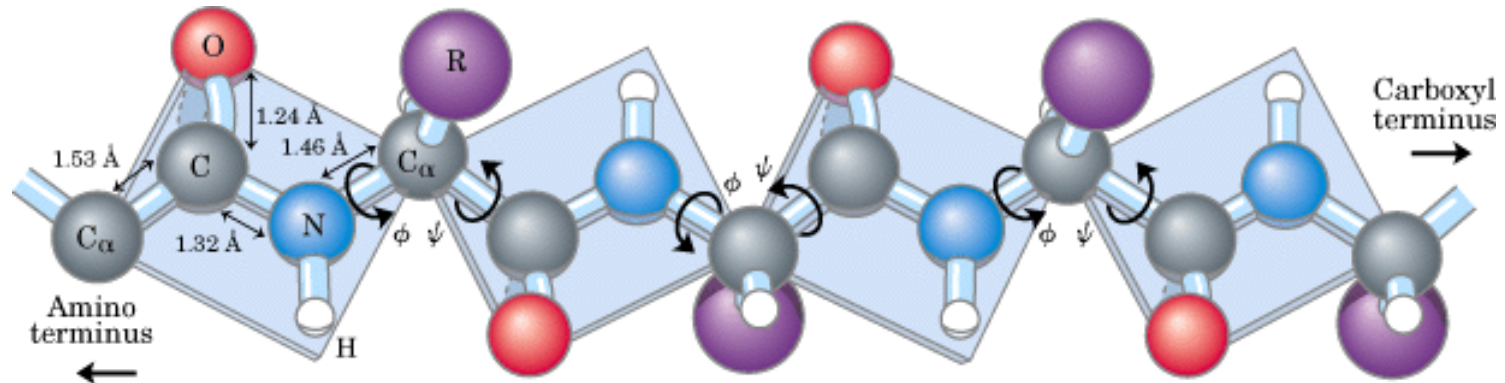
- α-helices
- β-sheets
- Coiled regions

Why does the protein fold that way?
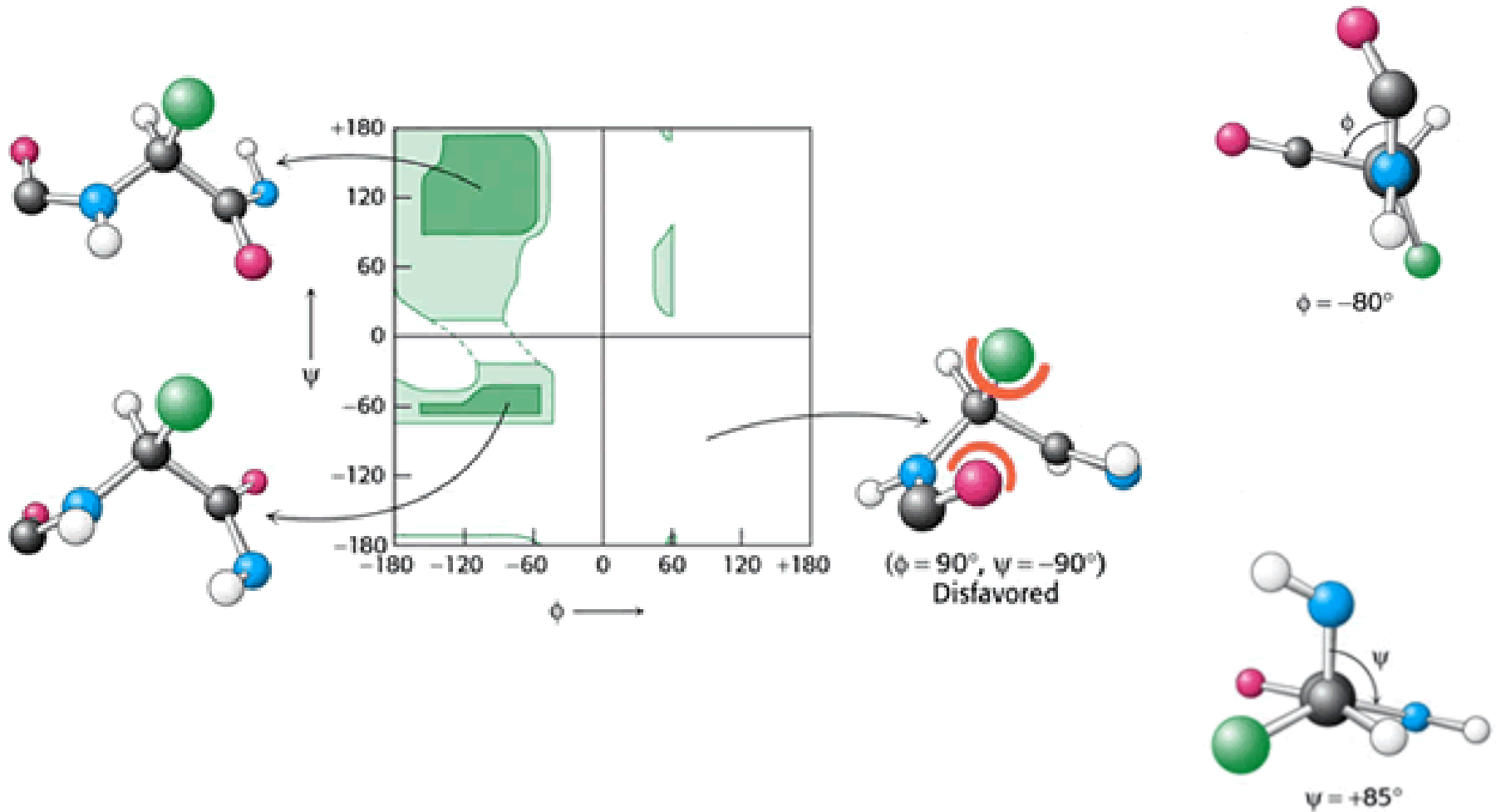Why is there a limited set of folds?
What is the influence of a mutation to the protein structure?

Getting more into detail: peptide bond

- Planar
- Planes are connected at C_α atoms
- They can only rotate around C_α bonds
  - Torsion angles φ and ψ

Ramachandran

$\phi = -80°$

$(\phi = 90°, \psi = -90°)$
Disfavored

$\psi = +85°$

Levinthal's Paradox

"Let us ask ourselves <u>how proteins fold to give such a unique structure. By going to a state of lowest free energy?</u> Most people would say yes and indeed, this is a very logical assumption. [...]

Even if we knew these angles to better than a tenth of a radian, there would be 10^{300} <u>possible configurations</u> in our theoretical protein. <u>In nature, proteins</u> apparently do not sample all of these possible configurations since they <u>fold in a few seconds</u>, and even postulating a minimum time for going from one conformation to another, the proteins would have time to try on the order of 10^8 different conformations at most before reaching their final state. We feel that <u>protein folding is speeded and guided by the rapid formation of local interactions which then determine the further folding of the peptide</u>. This suggests local amino acid sequences which form stable interactions and serve as nucleation points in the folding process."

Cyrus Levinthal. How to fold graciously. In J. T. P. Debrunnder and E. Munck, editors, Mossbauer Spectroscopy in Biological Systems: Proceedings of a meeting held at Allerton House, Monticello, Illinois, pages 22–24. University of Illinois Press, 1969.
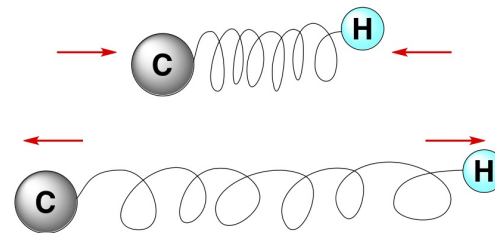
Molecular mechanics: motivation

- Most accurate simulation method available: Quantum mechanics
- But: very time-consuming, because every electron is modeled
    - → Approximation desired

- Molecular mechanics: approximation of molecular systems using the classical Newtonian mechanics

Molecular mechanics: All-atomic model

- Each atom is considered as one unit
    - No separate electrons

- Has several properties, usually
    - (van der Waals) radius
    - Polarizability
    - Constant net charge

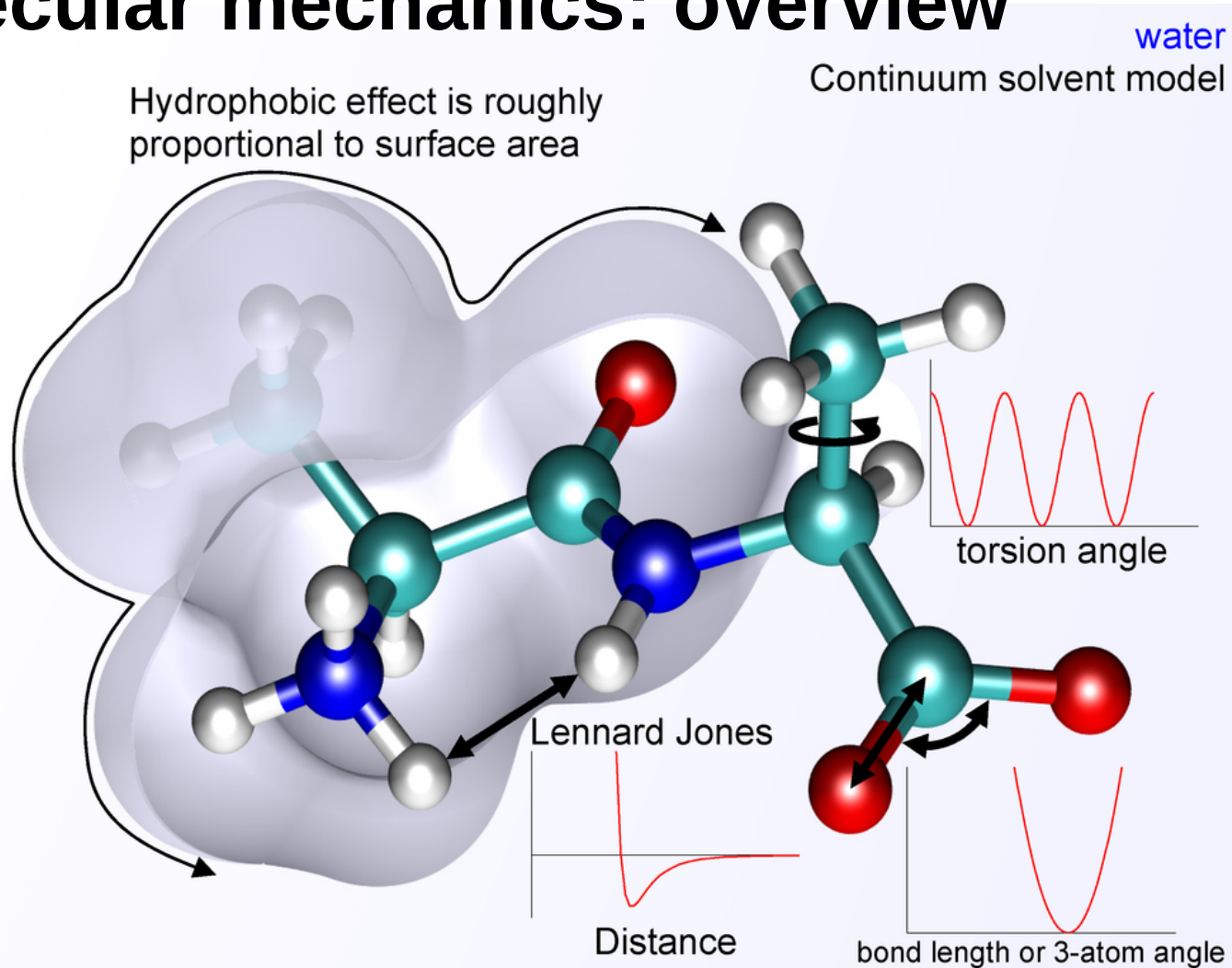- Covalent bonds are simulated as springs
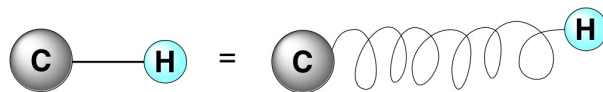
Any bond can be thought of as a spring...   **...constantly stretching and contracting.**

Molecular mechanics: overview

Force fields

- A set of functions that are used in molecular mechanics to describe the potential energy of a system

General functional form:

$$E = E_{stretch} + E_{bend} + E_{tors} + E_{vdW} + E_{ES}$$

Physical background

- Protein structure is largely determined by intramolecular forces
- Covalent bonds
  - Also, disulfid bonds
- Noncovalent bonds/interactions
  - Van der Waals interactions
  - Electrostatic interactions
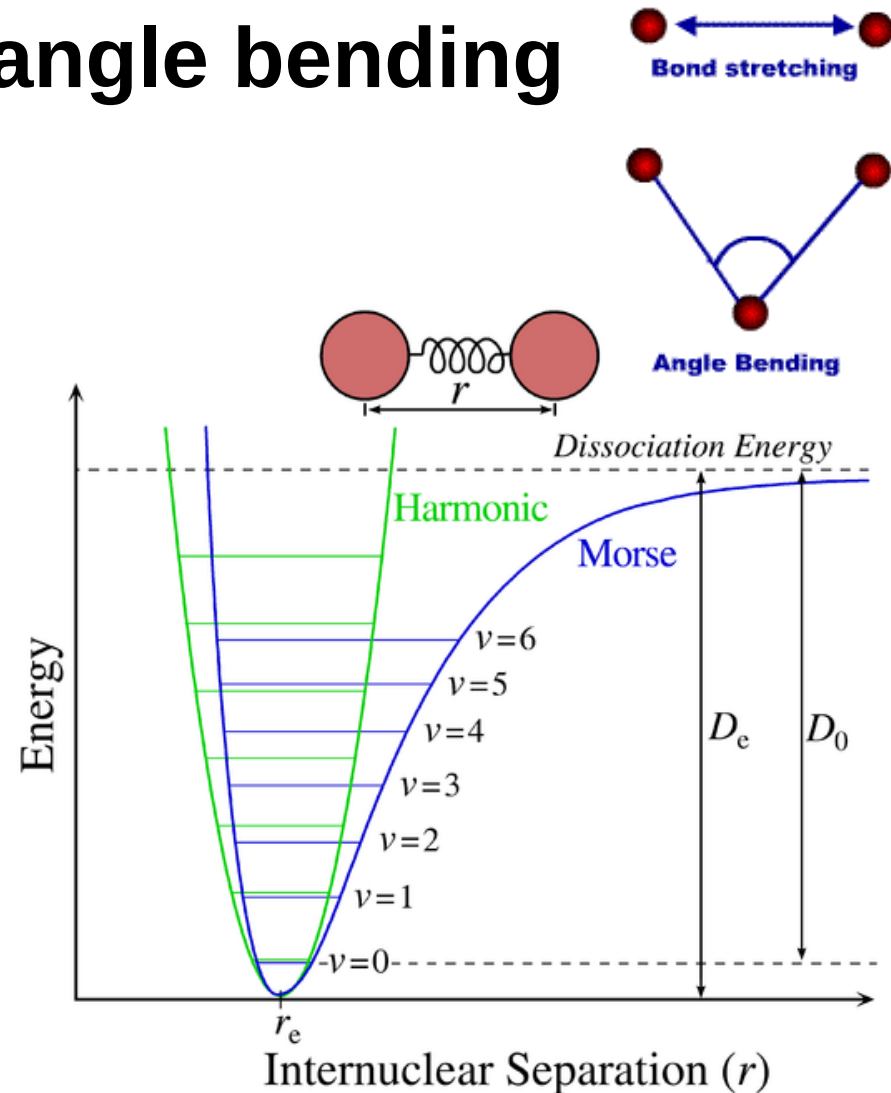   - Ionic bonds
    - Hydrogen bonds

Empirical Potential Energy Function

Bonds

Angles

Improper Dihedrals

Torsions

Electrostatics

van der Waals

11
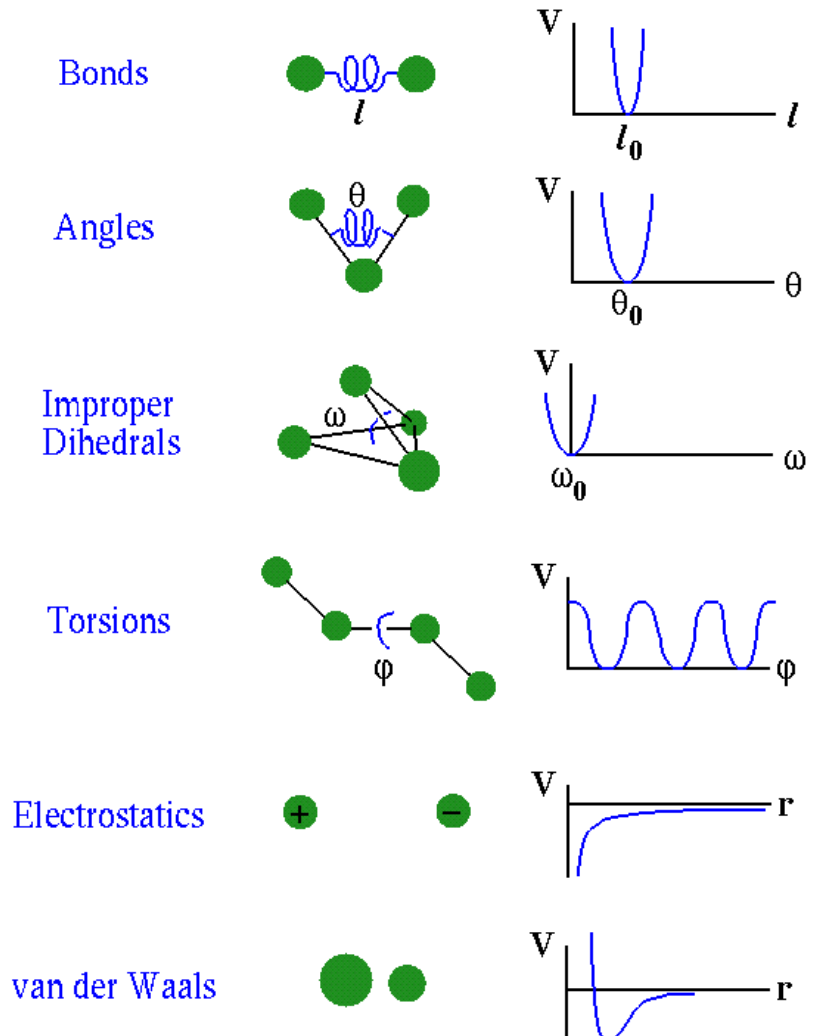
Bond stretching and angle bending

- Approximated by harmonic potential

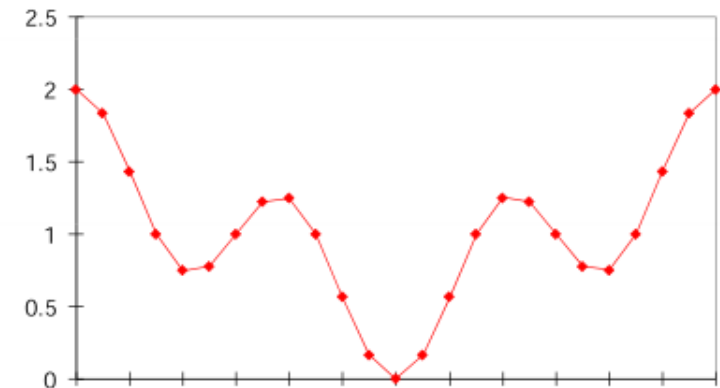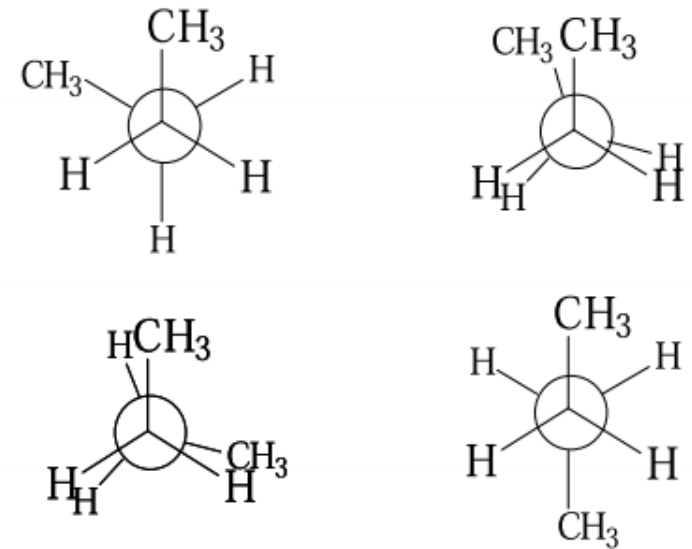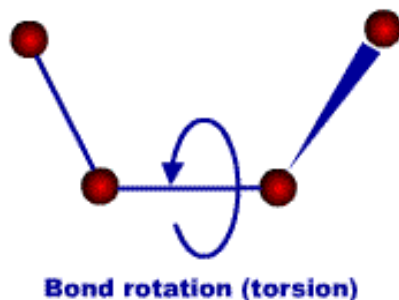- More accurate, but more time-consuming: Morse potential

$$V(r) = \frac{k}{2}(r - r_e)^2$$

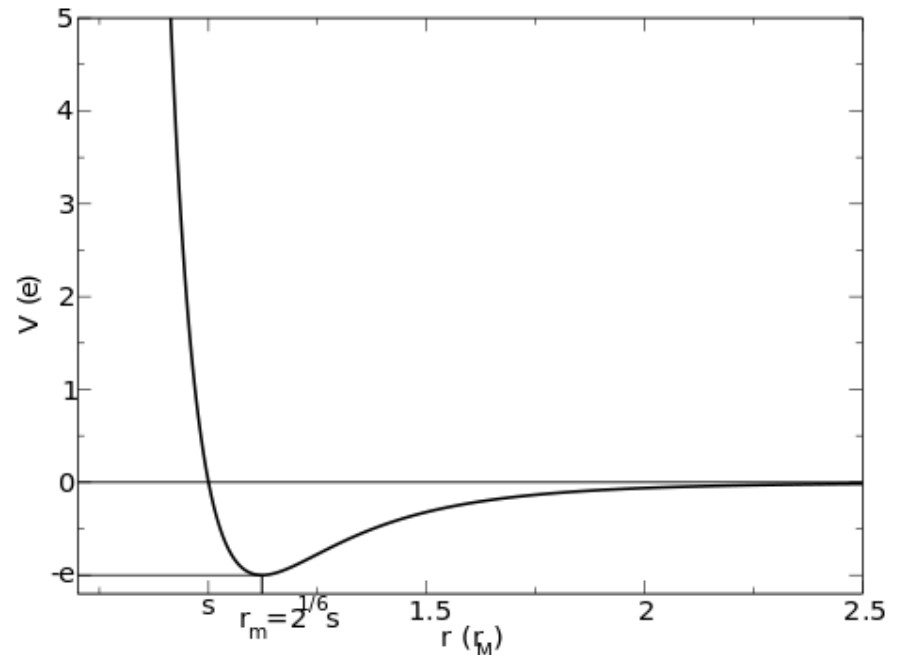# Torsions

- More than one minimum

- Described by Multi-minima harmonic potential

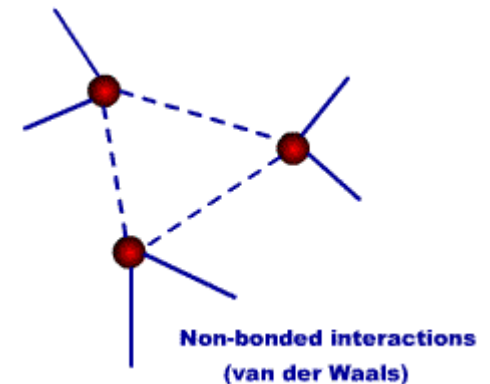$$V(x) = \sum_{torsions\ (ijkl)} \frac{k^{(ijkl)}}{2}\left(1 + \cos(n^{(ijkl)}\tau - \tau_0^{(ijkl)})\right)^2$$

Bond rotation (torsion)

Van der Waals interactions

Non-bonded interactions (van der Waals)

- Described by Lennard-Jones potential

- Increases with distance

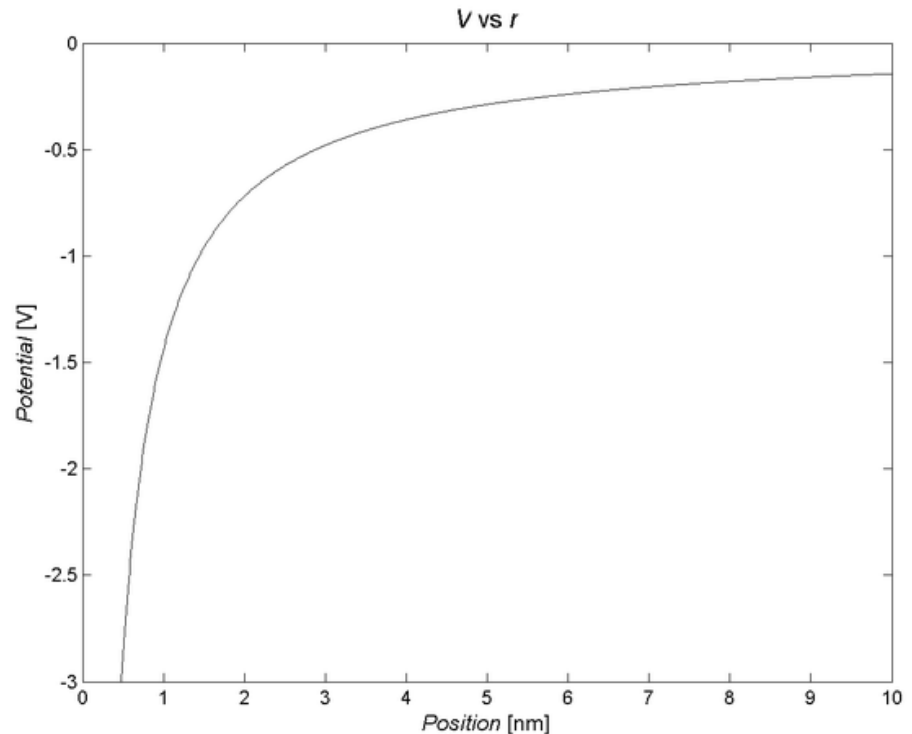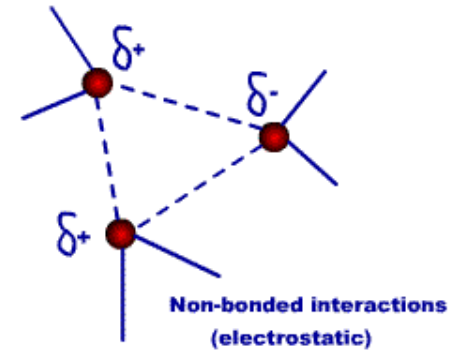- Usually used with a cut-off radius

$$V(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

14

Electrostatic interactions

- Approximated by
  Coulomb potential

- Increases with distance

- "Long-range interaction"
  → difficult to calculate
    correctly

$$V(r) = \frac{1}{4\pi\varepsilon_o} \frac{q_1 q_2}{r}$$

Putting all together: Basic force field

$$E = E_{stretch} + E_{bend} + E_{tors} + E_{vdW} + E_{ES} =$$

$$\sum_{bonds(ij)} \frac{k^{(ij)}}{2}(r_{ij} - r_0^{ij})^2 + \sum_{angles(ijk)} \frac{k^{(ijk)}}{2}(\phi_{ijk} - \phi_0^{ijk})^2 +$$

$$+ \sum_{torsions(ijkl)} \frac{k^{(ijkl)}}{2}(1 + cos(n^{ijkl}\tau - \tau_0^{ijkl})^2 +$$

$$+ \sum_{pairs(ij)} 4\epsilon \left(\frac{\sigma_{(ij)}}{r_{ij}^{12}} - \frac{\sigma_{(ij)}}{r_{ij}^6} \right) + \frac{1}{4\pi\epsilon_0} \sum_{pairs(ij)} \frac{q_i q_j}{r_{ij}}$$
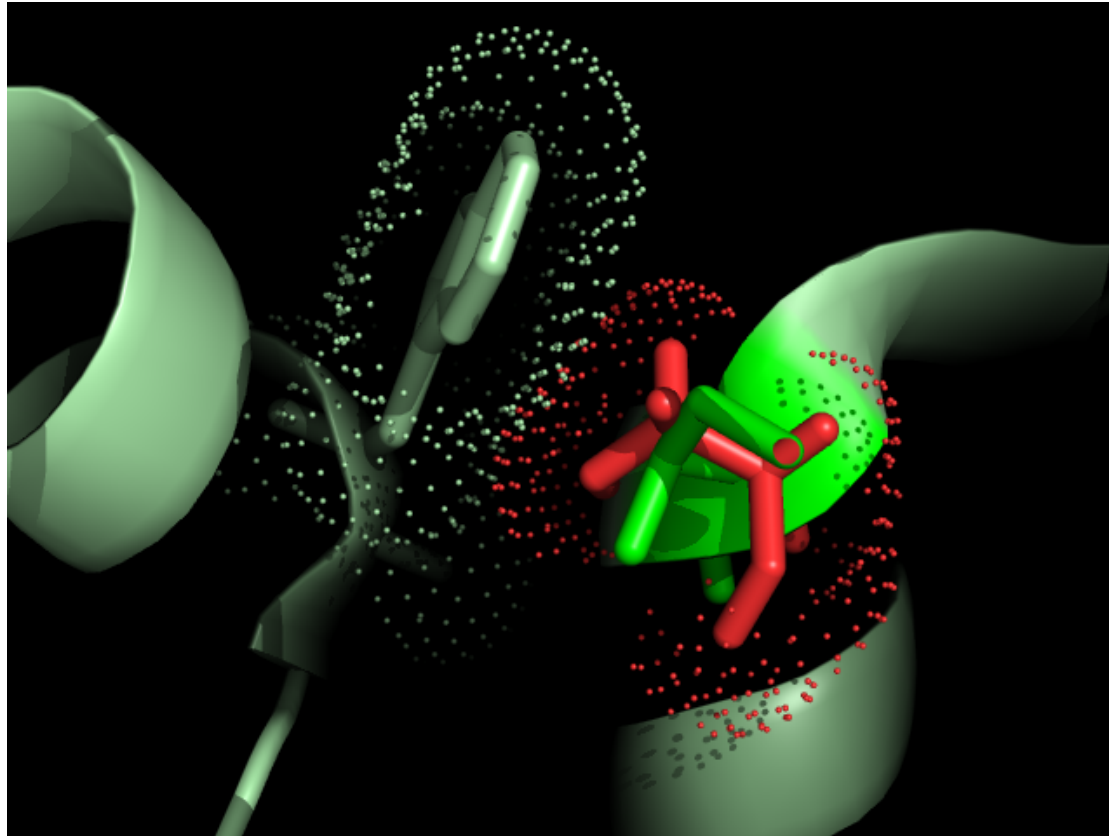
Motivation

Motivation
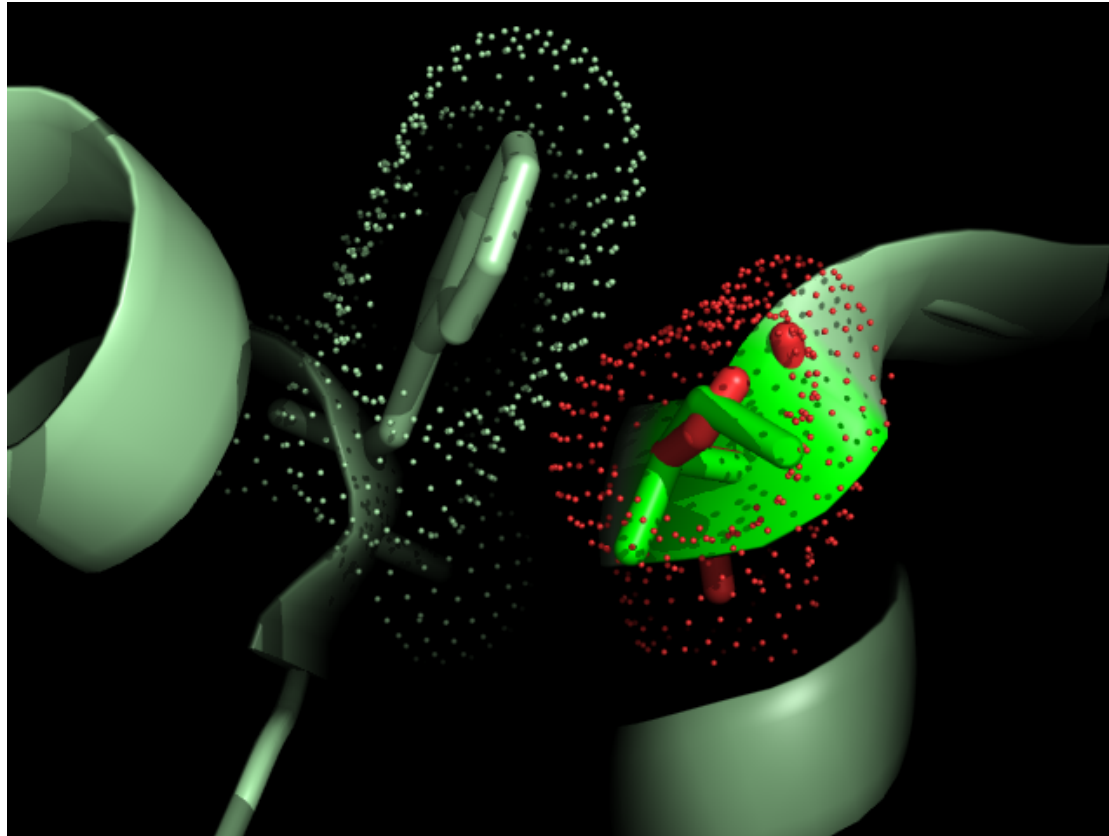
SCWRL

- SCWRL: <u>S</u>ide-Chain <u>C</u>onformation Prediction <u>w</u>ith

    <u>R</u>otamer <u>L</u>ibrary

- prediction of protein side-chain conformations

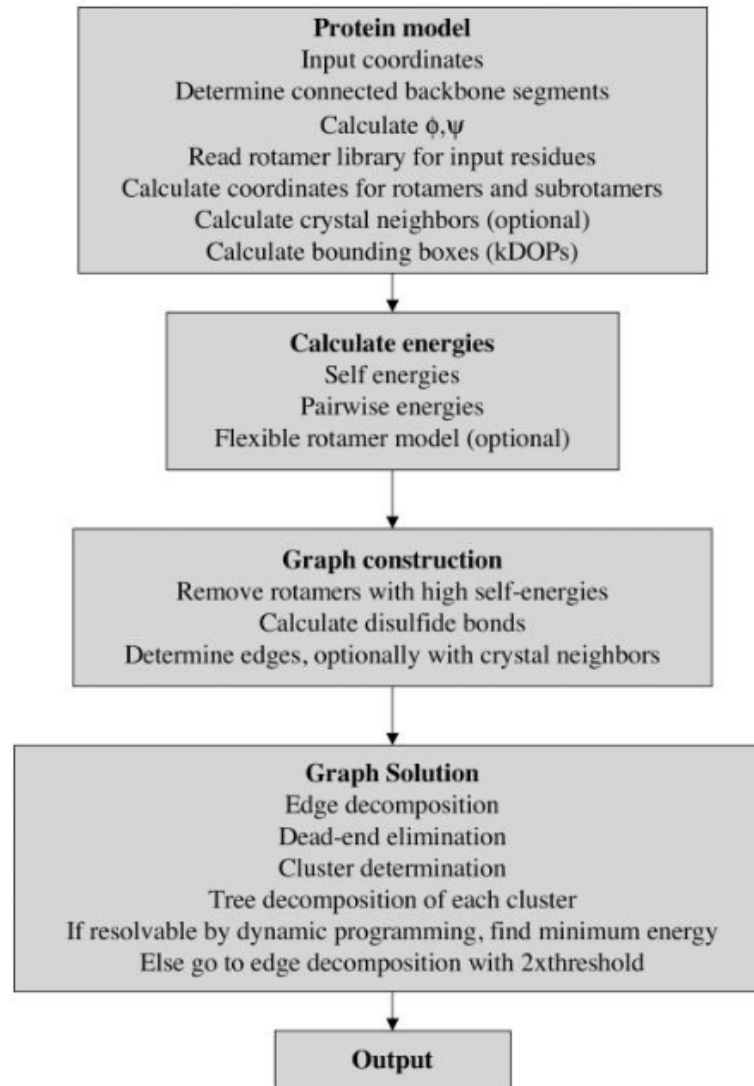- algorithm based on graph theory

- solves combinatorial problems in side-chain prediction

- Dunbrack Lab, FCCC, Philadelphia (1997)

- SCWRL 4.0 (2009)

SCWRL

Protein model
Input coordinates
Determine connected backbone segments
Calculate ϕ, ψ
Read rotamer library for input residues
Calculate coordinates for rotamers and subrotamers
Calculate crystal neighbors (optional)
Calculate bounding boxes (kDOPs)

Calculate energies
Self energies
Pairwise energies
Flexible rotamer model (optional)

Graph construction
Remove rotamers with high self-energies
Calculate disulfide bonds
Determine edges, optionally with crystal neighbors

Graph Solution
Edge decomposition
Dead-end elimination
Cluster determination
Tree decomposition of each cluster
If resolvable by dynamic programming, find minimum energy
Else go to edge decomposition with 2xthreshold

Output

SCWRL 4.0

New features:

1) backbone-dependent rotamer library

2) averaging over samples of conformations

3) anisotropic hydrogen bonding function

4) van der Waals atom-atom interaction potential

5) collision detection using k-discrete oriented polytopes

6) tree decomposition algorithm

7) optimization of all parameters by determining the
   interaction graph

SCWRL 4.0

1) backbone-dependent rotamer library
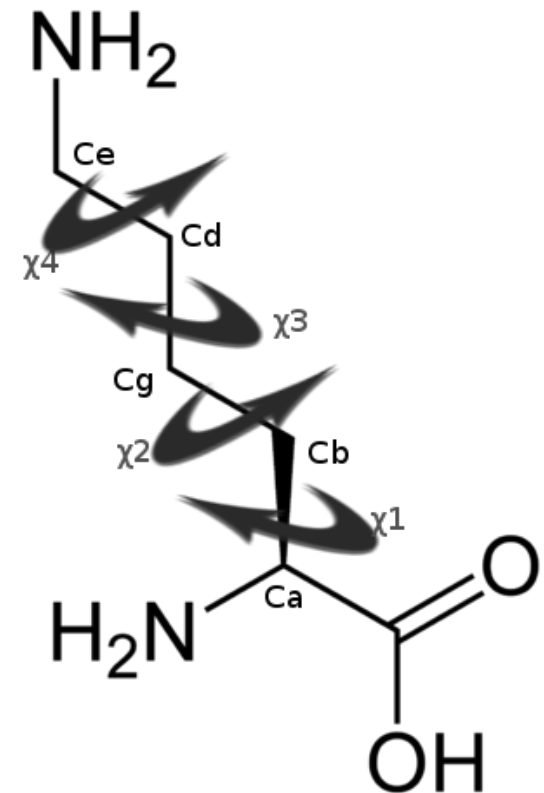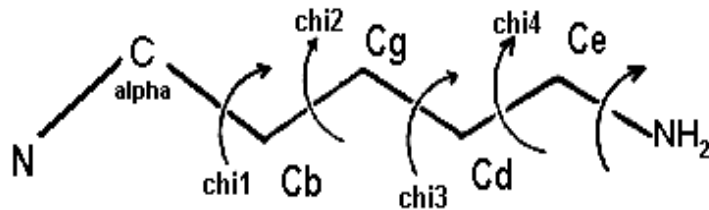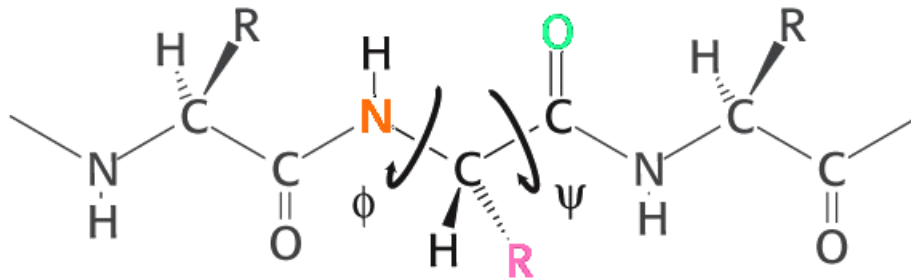
SCWRL 4.0

2) averaging over samples of conformations

- free-energy-like scoring function

- sampled χ angles around the library values

- average the energy of interaction between rotamers of different side chains

- to account for variation of dihedral angles given in the rotamer library

SCWRL 4.0

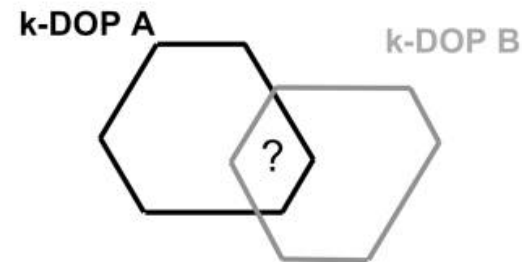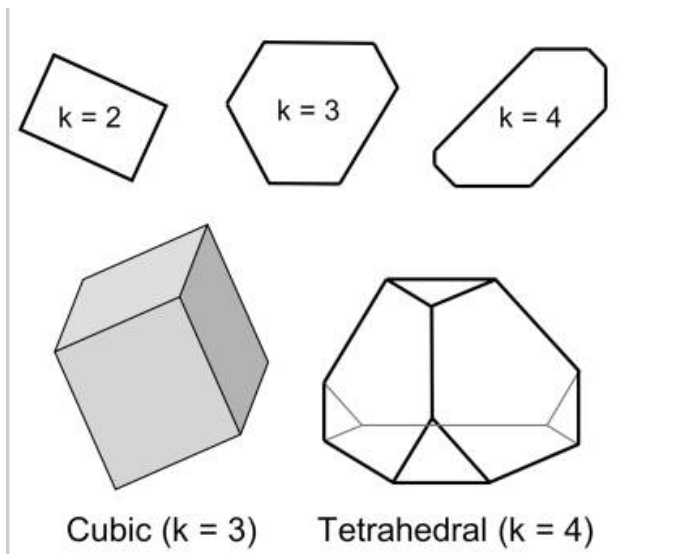3) anisotropic hydrogen bonding function

4) van der Waals atom-atom interaction potential

- directionally dependent

- no linear repulsive-only function
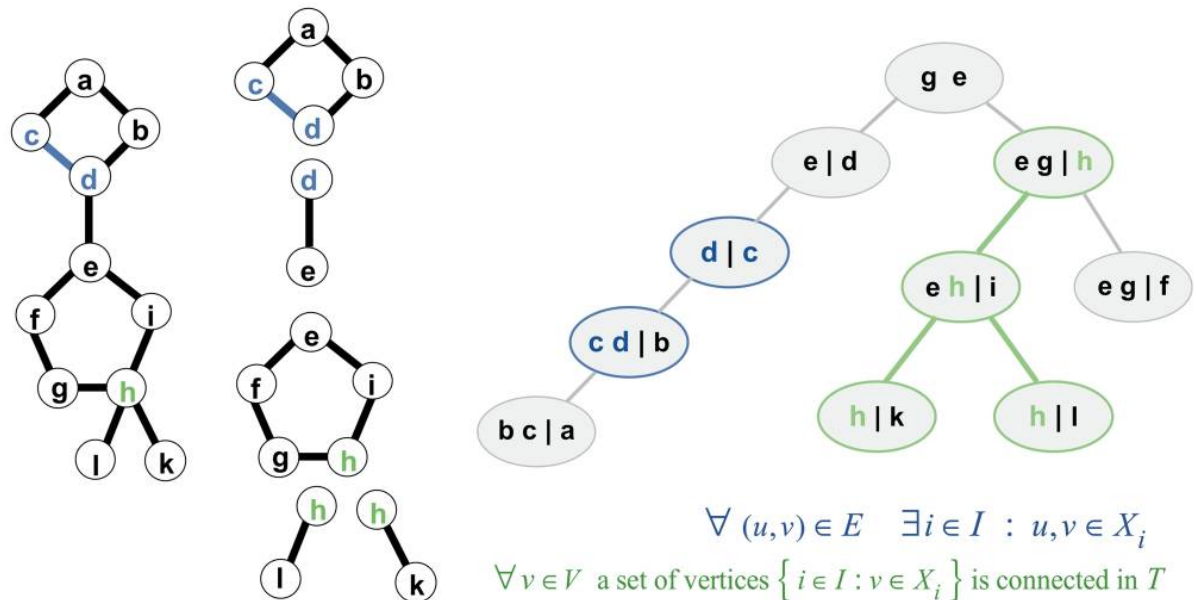
SCWRL 4.0

5) collision detection using k-discrete oriented polytopes

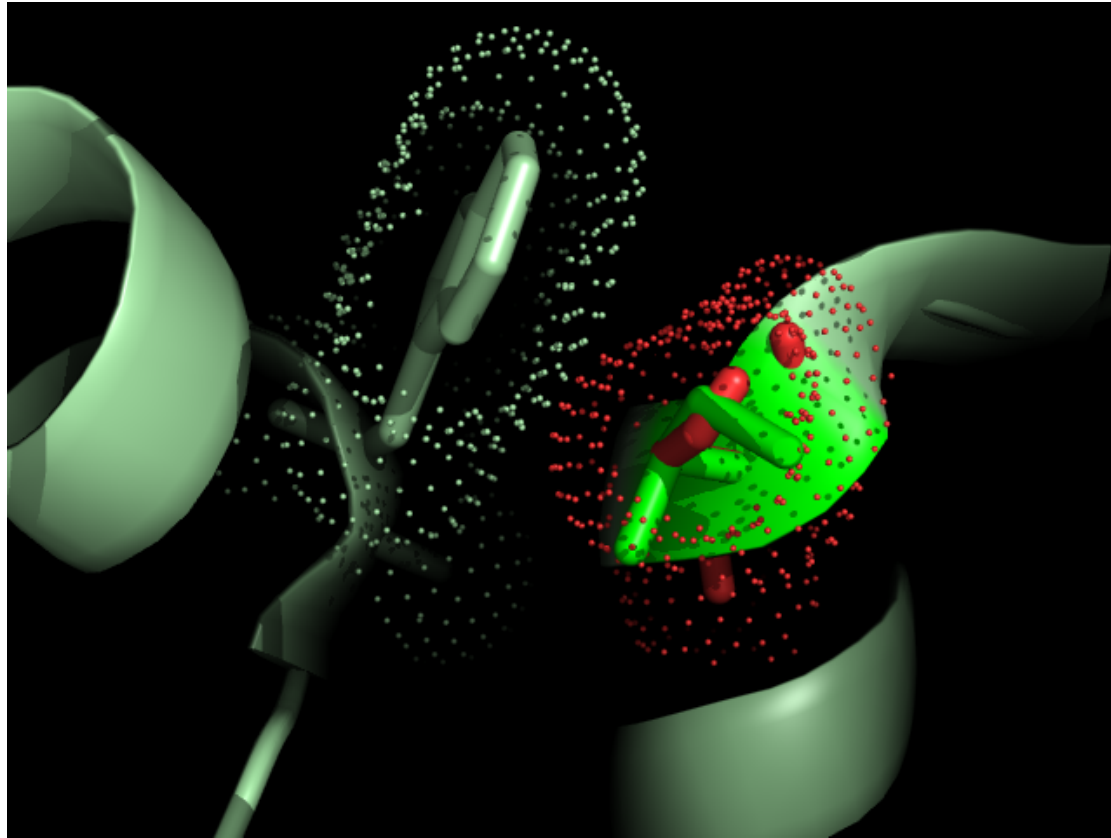# **SCWRL 4.0**

6) tree decomposition algorithm

7) optimization of all parameters by determining the
interaction graph

$$\forall\,(u,v)\in E \quad \exists\,i\in I \;:\; u,v\in X_i$$

$\forall\,v\in V$ a set of vertices $\left\{\,i\in I : v\in X_i\,\right\}$ is connected in T

Motivation

Stable?

FoldX

- Predict stability of proteins and protein complexes

- estimation of the importance of the interactions contributing to the stability

- empirical force field (EEEF - using empirical data obtained form experiments ran on proteins)

- FOLD-X energy function (FOLDEF) calculating the free energy of unfolding ΔG

- Fast

- Accurate

Energy terms in FOLDEF

$$\Delta G = W_{vdw} * \Delta G_{vdw} + W_{solvH} * \Delta G_{solvH} + W_{solvP} * \Delta G_{solvP} + \Delta G_{wb} + \Delta G_{hbond} +$$
$$\Delta G_{el} + \Delta G_{Kon} + W_{mc} * T * \Delta S_{mc} + W_{sc} * T * \Delta S_{sc}$$

ΔG :   Gibbs free energy

W :   Weight

vdw: Van der Waals

solvH/solvP: differences in solvation energy, when changing from
                unfolded to folded state

wb: extra stabilising free energy provided by water bridges

hbond: energy difference between formation of intra-molecular
                compared to inter-molecular hydrogen-bond formation

el: electrostatic contribution of charged groups

mc/sc: entropy cost of fixing the backbone/side chain in the
                folded state/particular conformation

References

- Andrew R. Leach (2001) Molecular Modelling: Principles and Applications (2nd Edition)

- Iris Antes, Lecture slides "Modellierung und Simulation biologischer Simulation biologischer Makromoleküle"

References (continued)

- G. G. Krivov, M. V. Shapovalov, and R. L. Dunbrack, Jr „*Improved prediction of protein side-chain conformations with SCWRL4*." Proteins (2009).

- Guerois, R., J. E. Nielsen, et al. (2002). "Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations." J Mol Biol 320(2): 369-87.

- Schymkowitz, J. W. et al. Prediction of water and metal binding sites and their affinities by using the Fold-X force field. Proc Natl Acad Sci U S A 102, 10147-52 (2005).