Structure-based mutation analysis

Practical Bioinformatics 'Protein Structure and Function Analysis'

Karolina Worf

02. July 2013

Overview

- Introduction
  - Protein structure level
  - Molecular mechanics
- Force fields
- Methods
  - SCWRL
  - FoldX

Introduction

Level of protein structure:

- **Primary structure:** Amino acid sequence
- **Secondary structure:** areas of folding and coiling of the amino acid sequence, stabilized by hydrogen bonds (α-helices, β-sheets, coiled regions)
- **Tertiary structure:** final 3-D structure of the polypeptide chain of a protein
- **Quarternary structure:** protein consisting of more than one polypeptid chain
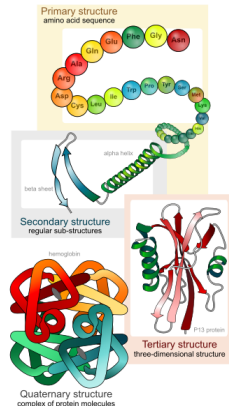
http://upload.wikimedia.org/wikipedia/commons/c/c9/Main_protein_structure_levels_en.svg

Introduction

- What effect has a mutation (SNP) on the protein structure?
  - No effect?
  - Changes the structure? => interactions between molecules (energy of protein)

- How to find out? Required steps?
  - Calculate energy of wild-type and mutated protein
  - Compare the energy results
  - Check whether changes in the structure can be found

Introduction

How to calculate the energy of a protein?

- **Quantum mechanics (QM):**
  - Calculates energy via all electrons in a system
    - + Most accurate method
    - − Time consuming!

- **Molecular mechanics (MM):**
  - Calculates energy of a system as a function of nuclear positions
    - + Fast and easy to calculate
    - − Ignores electronic motions

Force field

The energy function (force field) of a molecular system is defined as a sum of terms, which includes:

- Bonded interactions
  - Bond stretching
  - Angle bending
  - Bond rotation (Torsion)
- Non-bonded interactions
  - Electrostatic interactions
  - Van der Waals interactions
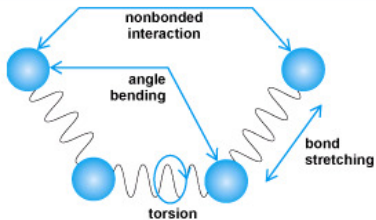
General formula:
$$E = E_{bonds} + E_{angles} + E_{torsions} + E_{vdW} + E_{ES}$$

Force field

Intramolecular energy term which is associated with the deformation of the electronic structure of the molecule.
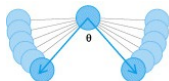
- **Bond stretching**
  - Hooke's law formula: $v(l) = \frac{k}{2}(l - l_0)^2$
- **Angle bending**
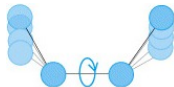 - Also calculated by Hooke's law
- **Bond rotation (Torsion)**
  - Contribution of each bonded quartet:
    $v(\omega) = \sum_{n-0}^{N} \frac{V_n}{2}(1 + cos(n\omega - \gamma))$
    $\omega = torsionangle$

http://www.accessscience.com/loadBinary.
aspx?filename=757328FG0020.gif

Force field

Inter- and intramolecular interactions occuring from electrostatic (through-space) interactions.
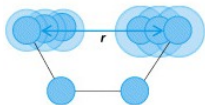
- **Electrostatic interactions**
  - Charge distribution between molecules
  - Calculated by Coulomb law:
    $E = \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \frac{q_i q_j}{4\pi\epsilon_0 r_{-ij}}$
- **Van der Waals interactions**
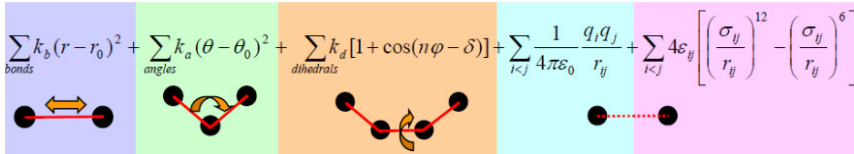 - Described by Lennard-Jones equation:
    $v(r) = 4\epsilon[(\frac{\sigma}{r})^{12} - (\frac{\sigma}{r})^6]$

http://www.accessscience.com/loadBinary.aspx?filename=757328FG0020.gif

Force field

Sum of all terms:

$$\sum_{bonds} k_b (r - r_0)^2 + \sum_{angles} k_a (\theta - \theta_0)^2 + \sum_{dihedrals} k_d [1 + \cos(n\varphi - \delta)] + \sum_{i<j} \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{r_{ij}} + \sum_{i<j} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$

http://www.ch.embnet.org/CoursEMBnet/Basel05/Day1_Intro.pdf

Features:

- Transferability of functional form and parameters
- Empirical - there is no 'correct' formular

Methods

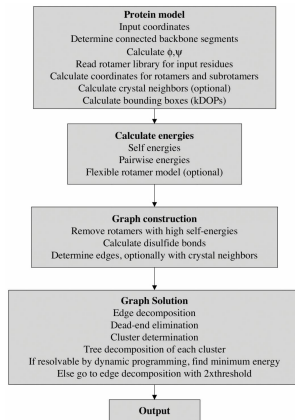# SCWRL4 - Side-chain Confirmation Prediction With Rotamer Library

- predicts side-chain confirmations
- uses backbone-dependent rotamer library
- based on graph theory
- fast and accurate
- ease-of-use

Input:

- PDB-file with backbone coordinates
- Sequence file with mutated sequence
  Example: rhiSpqaKallqdkdviAinqdplgk

SCWRL4 - basic steps

- Inputting data and constructing side chain coordinates with rotamer library
- Calculating energies (self/pairwise)
- Graph computation, with symmetry operators if any
- Combinatorial optimization via edge decomposition, dead-end elimination and tree decomposition of each cluster
- Outputting results

SCWRL

Output:

- PDB-File: coordinates for the structure with predicted side-chains

- StdOut (Examples):
  - Total minimal energy of the graph = 115.41 (without mutations)
  - Total minimal energy of the graph = 972.97 (with mutations)

FoldX

- Fast and accurate estimation of mutational free energy changes (effect of SNPs) on the protein stability
- Approach: **'Multiple mutations using individual list'**

Input (command line version)
- Manual mode:
  - PDB-File
  - Option file
  - Command file
- Runfile mode:
  - Run-File for 1 or more jobs

FoldX - Force fields

$$\Delta G = a \cdot \Delta G_{vdw} + b \cdot \Delta G_{solvH} + c \cdot \Delta G_{solvP} + d \cdot \Delta G_{wb} + e \cdot \Delta G_{hbond} + \\ f \cdot \Delta G_{el} + g \cdot \Delta G_{kon} + h \cdot T \cdot \Delta S_{mc} + k \cdot T \cdot \Delta S_{sc} + l \cdot \Delta G_{clash}$$

a-l:	Relative weights
$\Delta G/\Delta S$:	Gibbs free energy (experimental data/theoretical estimates)
vdw:	Van der Waals
solvH/solvP:	Solvation energy for hydrophobic/polar groups
wb:	Water bridges
hbond:	Hydrogen-bonds
el:	Electrostatic
kon:	Reflects the effect of electrostatic interactions to k_{on}
mc/sc:	Entropy cost fixing backbone in folded state/side chain in particular confirmation
clash:	steric overlaps between atoms in the structure

FoldX

Examples can be found on the FoldX web server:
`http://foldx.crg.es/`

Output:

- *PDB-File*: one for each mutant and corresponding wild-type
- *Dif-File*: difference in energy between the mutation and the corresponding wild-type reference (Positive numbers = less stability)
- *PDBList-File*: List of all created PDB Files
- *Raw-File*: stability for each PDB

Thank you for your attention!
Any questions?

References

- `http://www.vivo.colostate.edu/hbooks/genetics/biotech/` `basics/prostruct.html`
- `www.ch.embnet.org/CoursEMBnet/Basel05/Day1_Intro.pdf`
- Andrew R. Leach (2001): *Molecular Modelling: Principles and Applications (Second Edition)*. Pearson Education Limited 1996, 2001.
- Georgii G. Krivov et al. (2009): *Improved prediction of protein side-chain conformations with SCWRL4*. Proteins Vol. 77(4): 778-795.
- Joost W.H. Schymkowitz et al. (2005): *The FoldX web server: an online force field*. Nucleic Acids Research Vol.33: W382-W388.
- Joost W.H. Schymkowitz et al. (2005): *Prediction of water and metal binding sites and their affinities by using the Fold-X force field*. PNAS Vol.102(29):10147-10152.
- FoldX Description and User Manual Version 3.0