Simplified Models for Proteins

Employment of Cα Trace

- The starting point for a conformational search is obtained by a build-up method.
- Build-up methods have to quickly locate low-energy, geometrically feasible conformations.
- To this aim, a simplified model and a simplified force field can be used.
- Cα can serve as a model: what about the force field? It can be derived/tuned from observed protein molecules.

Oobatake-Crippen FF (1981)

- Example of 1st gen. FFs; operates on a "Cα trace model"
- $\boldsymbol{x}_i$ is the position of the i-th Cα and the "virtual" bond length r_0 is 3.8 Å
- $\boldsymbol{U} = \boldsymbol{U_b} + \boldsymbol{U_{nb}}$ $\quad U_b = \frac{k_b}{2}\sum_{i=1}^{N-1}\left(r_{i,i+1} - r_0\right)^2$
- $\boldsymbol{U_{nb}} = \boldsymbol{U_1} + \boldsymbol{U_2} + \boldsymbol{U_3}$ one 4/6 LJ pot. + two **gaussian** pot.

$$U_1 = \frac{e_1}{n-m}\left[-m\left(\frac{r_1}{r}\right)^n + n\left(\frac{r_1}{r}\right)^m\right] \quad m=6, n=4$$

$$U_2 = e_2\, e^{-\frac{1}{2}\left(\frac{r-r_2}{d_2}\right)^2}$$

$$U_3 = e_3\, e^{-\frac{1}{2}\left(\frac{r-r_3}{d_3}\right)^2}$$

Constants $r_1, r_2, r_3, e_1, e_2, e_3, d_2$ and d_3 depends on the residue types.

Going Further: Toy Models

- The idea of studying strongly simplified model is supported by some observations:
  1. It is widespread opinion that folding is mainly driven by hydrophobic interaction, while atomic details just determine minor refinements.
  2. The most computationally challenging problem is the calculation of realistic potential.
  3. Systematic comparative studies of (mutated) sequences cannot be afforded by realistic models.
- It is not realistic to generally understand how folding operates by employing realistic models.
- In toy models, the basic building block is the residue, usually represented as a single point element.

Classification of Toy Models

- Toy models can be divided in two wide categories:
  - **Lattice models.**
    Residues are constrained to be placed on a chain of adjacent vertexes of a given lattice. The potential depends on the size of the hydrophobic core.
    Typical example: HP model by Dill (1985)
  - **Off-lattice models.**
    No lattice constraint is present; instead, at least one "bending" term in the FF is taken into account.
    Typical example: AB model by Stillinger (1993)
- In toy models, the distance between conformations can be defined in more convenient ways than RSMD

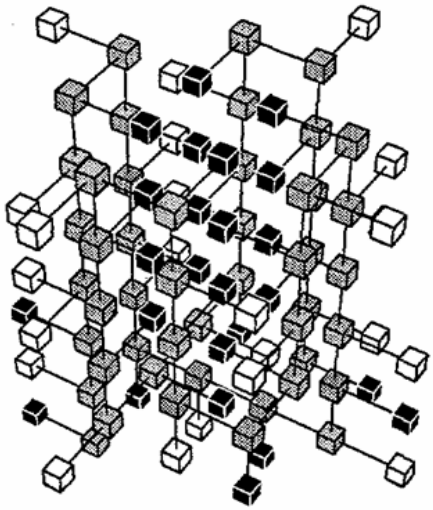
HP Model

- Residues: just two types, H (Hydrophobic) and P (Polar), embedded in a lattice L
- Two hydrophobic residues in a conformation have ***loose contact*** (or simply *contact*) if they are adjacent in the lattice but not connected by a bond
- The potential can be related to the number of contacts: the more the contacts, the lower the potential

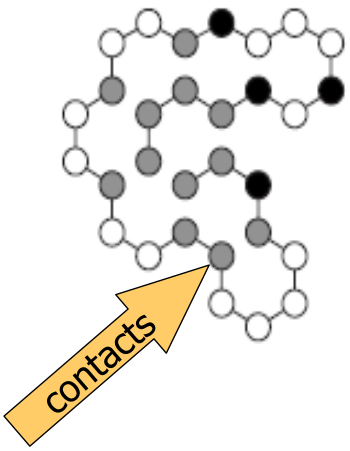
$$U_{HP} = -\frac{1}{2} \left\| \left\{ (a_i, a_j) \,\middle|\, a_i \text{ and } a_j \text{ are of type H and in contact} \right\} \right\|$$

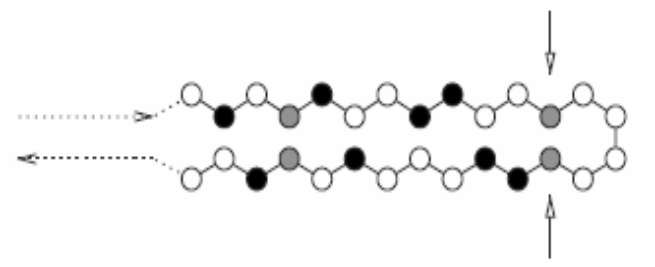
- L can be chosen in 2D, 3D; cubic, triangular, hexagonal

HP in 3D



HP in Hexagonal Lattice



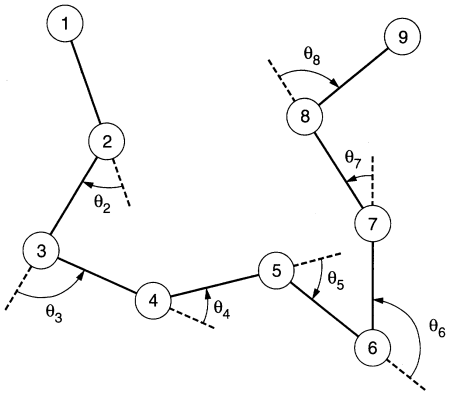


Conformational Search in HP

- The conformational search can be performed using algorithms that try to apply a given set of change rules to determine "adjacent" conformations.
- Suitable algorithms include Metropolis Montecarlo, Branch and Bound, heuristics and genetic algorithms.

AB Model

- Residues: just two types, A (hydrophobic) and B (polar).
- "Covalent" bonds between adjacent residues have unit length
- There have been studied both planar and 3D versions.
- The potential associated to a conformation is composed of one bending and one LJ part.



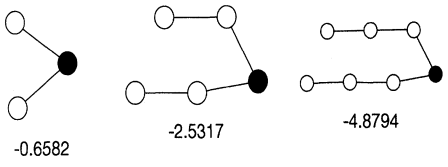
$$U = U_{bending} + U_{LJ}$$

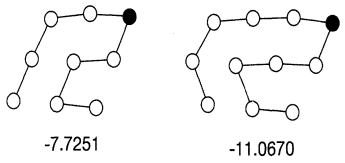
$$U_{bending} = \frac{1}{4}\sum_{k=2}^{N-1}(1-\cos\vartheta_k)$$

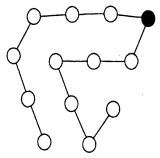
$$U_{LJ} = 4\sum_{j>i+1}\left[\frac{1}{r_{ij}^{12}} - \frac{C(\sigma_i,\sigma_j)}{r_{ij}^{6}}\right]$$

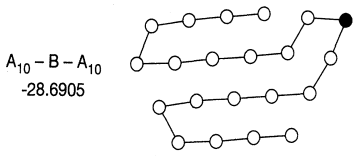
$$C(A,A) = 1$$
$$C(B,B) = 1/2$$
$$C(A,B) = C(B,A) = -1/2$$

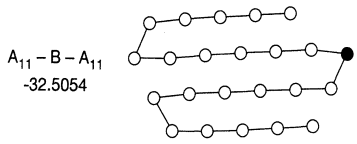
AB Model (I)

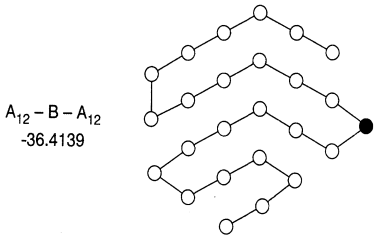












AB Model (II)