# Protein Modeling

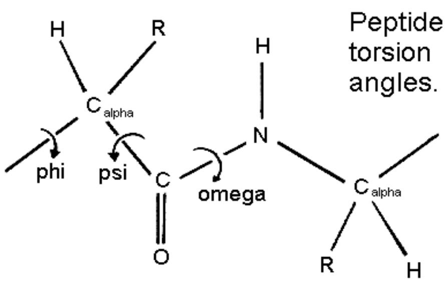
## Complexity of protein structure (Levinthal paradox)

100 residue protein
3 conformations per residue

number of distinct conformations:
$3^{100} \cong 10^{48}$

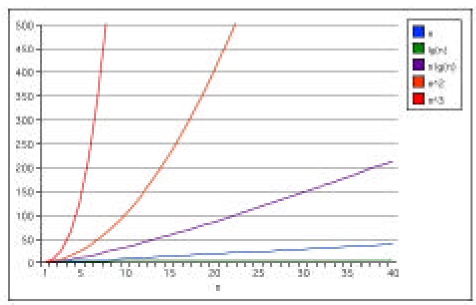
sampling time $\cong 10^{30}$ years

## Levinthal paradox




3 conformations per residue is a very conservative estimate

## Complexity of algorithms




## Complexity

P (Polynomial)

NP (Nondeterministic Polynomial)

NP-Complete

## Protein Folding Problem

Given: sequence
Find: structure

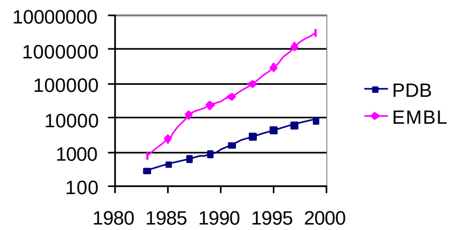
The problem is NP-complete

## Protein Folding Problem

Problem for us, not for proteins.
They just fold...

(Ken Dill)

## Dynamics of Database Growth




## Protein Modeling Methods

- ***Ab initio* methods:**
  solution of a protein folding problem
  search in conformational space
- **Energy-based methods:**
  energy minimization
  molecular simulation
- **Knowledge-based methods:**
  homology modeling
  fold recogniion

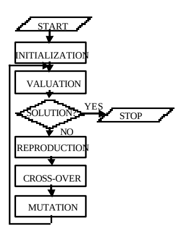
## *Ab initio* Methods

Simplified models
    simplified alphabet (HP)
    simplified representation (lattice)
Build-up techniques
Deterministic methods
    quantum mechanics
    diffusion equations
    DFT
Stochastic searches
    Monte Carlo
    genetic algorithms

## Genetic Algorithms

Search or optimization methods using simulated evolution.
Population of potential solutions is subjected to
natural selection, crossover, and mutation

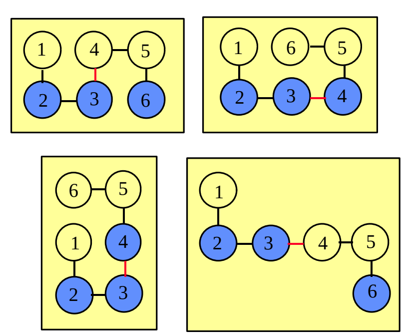
choose initial population
evaluate each individual's fitness
repeat
    select individuals to reproduce
    mate pairs at random
    apply crossover operator
    apply mutation operator
    evaluate each individual's fitness
until terminating condition



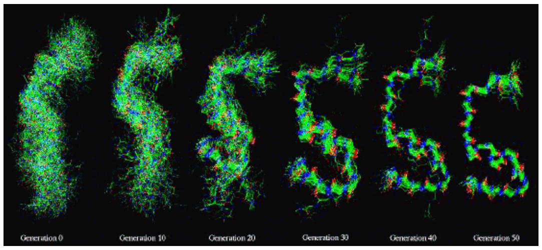

## GA simulation of folding

Parents
10 00 01 00 10
10 00 00 01 11

Children
10 00 10 01 11
10 00 01 00 10



# GA simulation of folding




Membrane binding domain of Blood Coagulation Factor VIII (J.Moult)

# Protein Modeling Methods

- ***Ab initio* methods:**
  solution of a protein folding problem
  search in conformational space
- **Energy-based methods:**
  energy minimization
  molecular simulation
- **Knowledge-based methods:**
  homology modeling
  fold recogniion

# Molecular structure representation

Elementary particles

Atoms

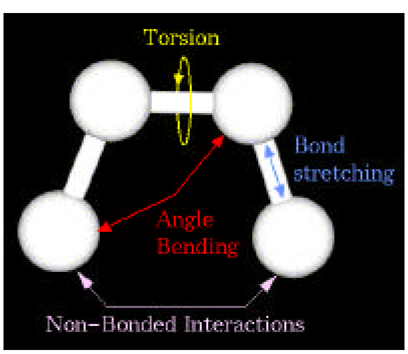
Groups of atoms

# Potential Energy Function

$$PEF(R) = \sum_{bonds} K_b\{b(R)-b_{eq}\}^2 + \sum_{angles} K_\theta\{\theta(R)-\theta_{eq}\}^2 + \sum_{dihedrals} \frac{K_\phi}{2}\{1+\cos[n\phi(R)-\gamma]\} + \sum_{\substack{\text{non-bonded}\\ \text{atom pairs } i,j}} \left[ \frac{A_{ij}}{r_{ij}(R)^{12}} - \frac{B_{ij}}{r_{ij}(R)^{6}} + \frac{q_i q_j}{\varepsilon_r \varepsilon_0 r_{ij}(R)} \right] \quad (1)$$

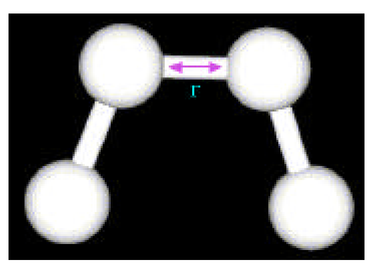
Forcefields: AMBER, CHARMM, CVF, ECEPP, GROMOS

# Non-Bonded Interactions

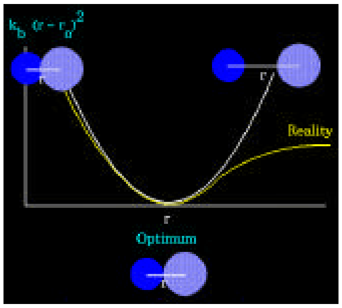



# Bond length

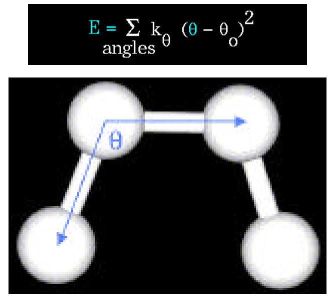
$$E = \sum_{bonds} k_b (r - r_o)^2$$



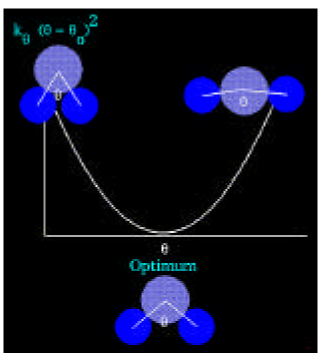
Bond length



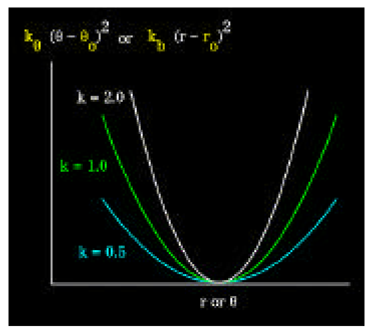

Bond angle



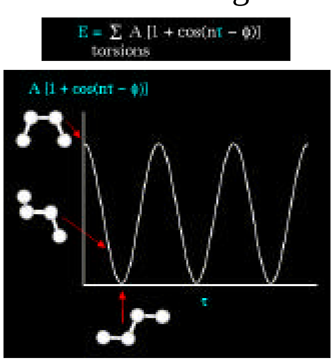

Bond angle



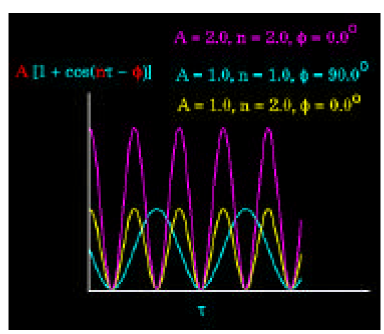

Bond length and angle (parameters)



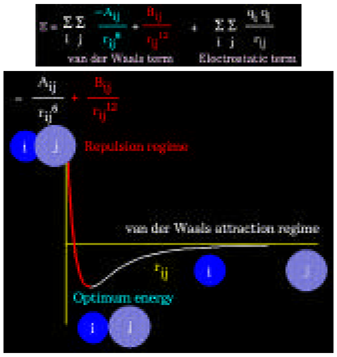

Torsional angle



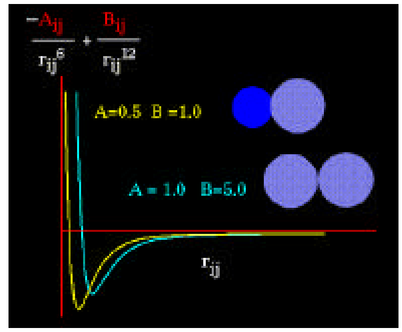

Torsional angle (parameters)



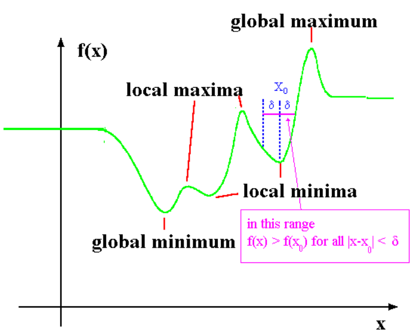
## Non-bonded terms



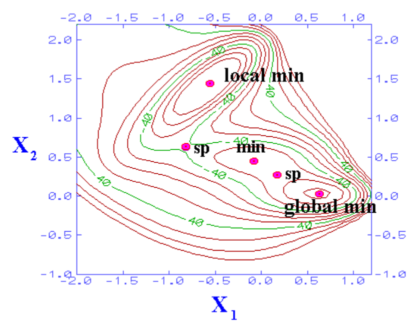

## Non-bonded terms (parameters)




## Energy Minimazation




## Energy Minimazation




## Molecular Dynamics

- Model system
- Initial conditions
- Boundary conditions
- Integration algorithm
- Constraints
- Ensemble
- Results

## Molecular Dynamics

$$F_i = m_i a_i$$

$$a_i = dv_i / dt$$

$$v_i = dr_i / dt$$

$$- dE / dr_i = F_i$$

$$- dE / dr_i = m_i d^2r_i / dt^2$$

## Molecular Dynamics

1. solve for $a_i$ at t using: $-\frac{dE}{dr_i} = F_i = m_i a_i(t)$
2. update $v_i$ at $t + \Delta t/2$ using: $v_i(t + \Delta t/2) = v_i(t - \Delta t/2) + a_i(t)\,\Delta t$
3. update $r_i$ at $t + \Delta t$ using: $r_i(t + \Delta t) = r_i(t) + v_i(t + \Delta t/2)\,\Delta t$

## Protein Modeling Methods

- ***Ab initio* methods:**
  solution of a protein folding problem
  search in conformational space
- **Energy-based methods:**
  energy minimization
  molecular simulation
- **Knowledge-based methods:**
  homology modeling
  fold recogniion

## Knowledge-based methods

Finding patterns in known structures

Deriving rules (usually in the form of PMF)

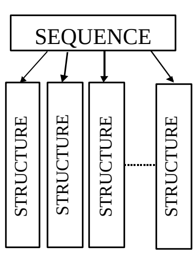
Applying the rules

## Fold Recognition

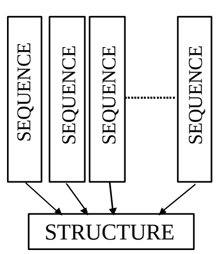
Pattern searching
- sequence patterns
- structure patterns
- residue composition patterns

Threading
- sequence-structure compatibility
- structure-sequence compatibility

## Threading




Sequence-structure compatibility




Structure-sequence compatibility

## Threading

Only the local environment is taken into account
Non-local contacts are assumed with generic peptide
No gaps are allowed in the alignment

## Homology Modeling

- **Identification of structurally conserved regions (using multiple alignment)**
- **Backbone construction (based on SCR)**
- **Loop construction (KB or conformational search)**
- **Side-chain restoration (KB, rotamer, or MM)**
- **Structure verification and evaluation**
- **Structure refinement (energy minimization)**

## Homology Modeling Programs

Modeller
(http://guitar.rockefeller.edu/modeller)

Swiss-Model
(http://www.expasy.ch/swissmod)

Whatif
(http://www.cmbi.kun.nl/whatif)

## Swiss-Model

- Method:
  Knowledge-based approach.
- Requirements:
  At least one known 3D-structure of a related protein.
  Good quality sequence alignements.
- Procedures:
  Superposition of related 3D-structures.
  Generation of a multiple a alignement.
  Generation of a framework for the new sequence.
  Rebuild lacking loops.
  Complete and correct backbone.
  Correct and rebuild side chains.
  Verify model structure quality and check packing.
  Refine structure by energy minimisation and molecular dynamics.

## Methods and Programs used by Swiss-Model

- **Sequence Alignment**
  BLAST (Altschul S.F., *J. Mol. Biol.* **215**:403, 1990)
  SIM (Huang, X., Miller, M. *Adv. Appl. Math*.**12**:337, 1991)
  ProModII (Peitsch, M.C. *Unpublished*, Server-specific tool)
- **Knowledge Based Protein Modelling**
  ProMod (Peitsch M.C. *Biochem Soc Trans* **24**:274, 1996)
- **Energy Minimisation**
  Gromos96 (van Gunsteren W.F. *http://igc.ethz.ch/gromos/*)
- **Model evaluation**
  Swiss-PdbViewer
  (*http://www.expasy.ch/spdbv/mainpage.html*)

## Swiss-Model Request Types

- First Approach mode.
- Optimise mode.
- Combine mode.
- GPCR mode.

## Model Confidence Factors

### The Model B-factors are determined as follows:

- The number of template structures used for model building.
- The deviation of the model from the template structures.
- The Distance trap value used for framework building.

### The Model B-factor is computed as:

85.0 * (1/ # selected template str.) * (Distance trap / 2.5)
and
99.9 for all atoms added during loop and side-chain building