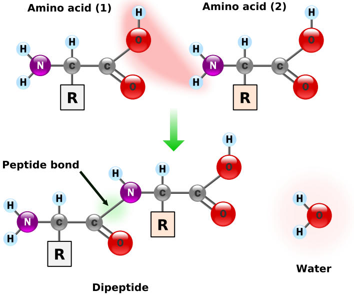
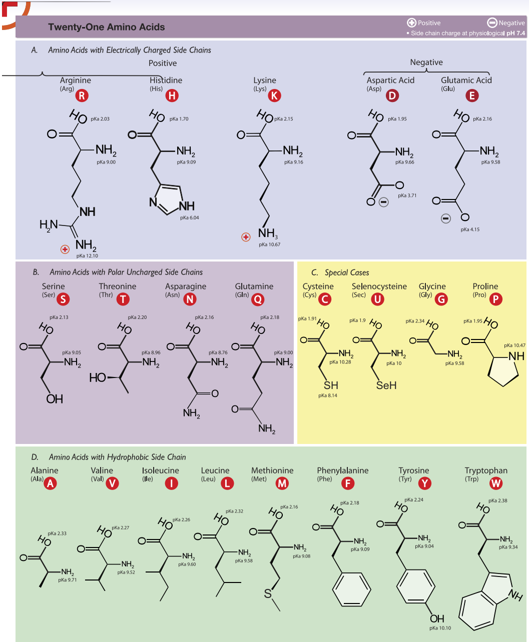
Protein Structure Prediction

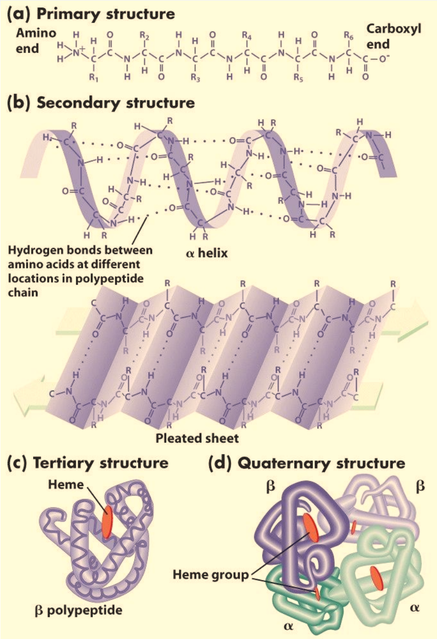
- Protein = chain of amino acids (AA)
- aa connected by peptide bonds



Amino Acids


Twenty-One Amino Acids
Positive
Negative
• Side chain charge at physiological pH 7.4
A. Amino Acids with Electrically Charged Side Chains
Positive
Negative
Arginine (Arg) R
pKa 2.03
pKa 9.00
pKa 12.10
Histidine (His) H
pKa 1.70
pKa 9.09
pKa 6.04
Lysine (Lys) K
pKa 2.15
pKa 9.16
pKa 10.67
Aspartic Acid (Asp) D
pKa 1.95
pKa 9.66
pKa 3.71
Glutamic Acid (Glu) E
pKa 2.16
pKa 9.58
pKa 4.15
B. Amino Acids with Polar Uncharged Side Chains
Serine (Ser) S
pKa 2.13
pKa 9.05
Threonine (Thr) T
pKa 2.20
pKa 8.96
Asparagine (Asn) N
pKa 2.16
pKa 8.76
Glutamine (Gln) Q
pKa 2.18
pKa 9.00
C. Special Cases
Cysteine (Cys) C
pKa 1.91
pKa 10.28
pKa 8.14
Selenocysteine (Sec) U
pKa 1.9
pKa 10
Glycine (Gly) G
pKa 2.34
pKa 9.58
Proline (Pro) P
pKa 1.95
pKa 10.47
D. Amino Acids with Hydrophobic Side Chain
Alanine (Ala) A
pKa 2.33
pKa 9.71
Valine (Val) V
pKa 2.27
pKa 9.52
Isoleucine (Ile) I
pKa 2.26
pKa 9.60
Leucine (Leu) L
pKa 2.32
pKa 9.58
Methionine (Met) M
pKa 2.16
pKa 9.08
Phenylalanine (Phe) F
pKa 2.18
pKa 9.09
Tyrosine (Tyr) Y
pKa 2.24
pKa 9.04
pKa 10.10
Tryptophan (Trp) W
pKa 2.38
pKa 9.34

Levels of structure



Protein Structure Prediction

Christian Anfinsen, 1961:

denatured RNase refolds into functional state (in vitro)

⇒ no external folding machinery

⇒ Anfinsen's dogma/thermodynamic hypthesis:
all information about native structure is in the sequence
(at least for small globular proteins)

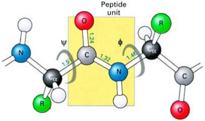
native structure = minimum of the free energy

- unique
- stable
- kinetically accessible

Levinthal's Paradox, 1969

Cyrus Levinthal: protein folding is not trial-and-error
Thought experiment:

- protein with 100 peptide bonds (101 aa)
- assume 3 states for each of the 200 phi and psi bond angles

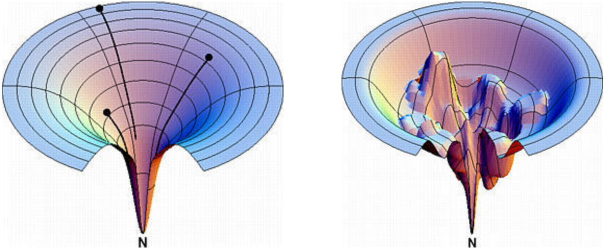


- $\Rightarrow 3^{200} \approx 10^{95}$ conformations
- assuming one quadrillion samples per secon, still over 60 orders of magnitude longer than the age of the universe

BUT: proteins fold in milliseconds to seconds

PARADOX

Principles of Folding 'Essentially' Understood



Folding Funnel

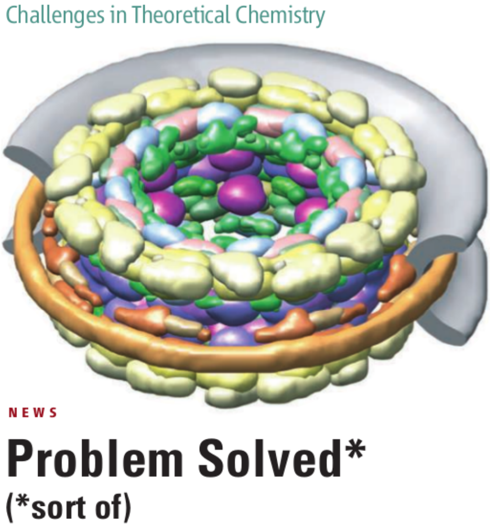
resolves Levinthal's Paradox

Driving forces:

- hiding of non-polar groups away from water
- close, nearly void-free packing of buried groups and atoms
- formation of intramolecular hydrogen bonds by nearly all buried polar atoms

Hydrophobic effect · Van-der-Waals · Electrostatic

August 8th, Science: problem solved?



Challenges in Theoretical Chemistry

NEWS

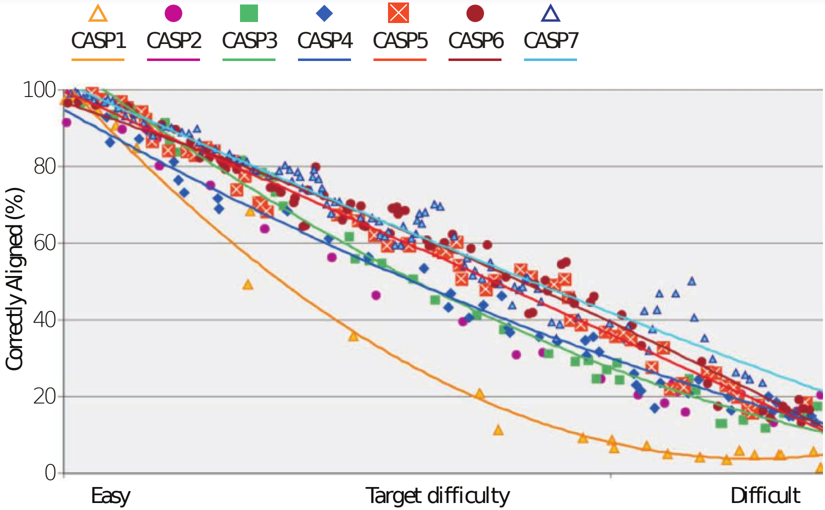
Problem Solved*

(*sort of)

Robert F. Service. Problem solved* (*sort of). Science, 2008.

[this and some following slides inspired by Jinbo Xu, Jerome Waldispühl]

Increasing Accuracy of Predictions: Slowly but Steadily



Steady rise. Computer modelers have slowly but steadily improved the accuracy of the protein-folding models.

Distance between 3D structures

RMSD = Root Mean Square Deviation

Compares two vectors of coordinates (here, coordinates of atoms in protein conformations). Yields distance between conformations.

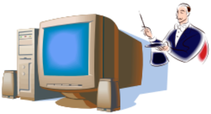
$$\begin{aligned}\mathrm{RMSD}(v,w) &= \sqrt{\frac{1}{n}\sum \|v_i - w_i\|^2} \\ &= \sqrt{\frac{1}{n}\sum (v_{ix} - w_{ix})^2 + (v_{iy} - w_{iy})^2 + (v_{iz} - w_{iz})^2}\end{aligned}$$

RMSD depends on orientation;
it is applied to superimposed structures, or after minimizing over rotations/translations (Kabsch algorithm)

CASP/CAFASP

- CASP:

  Critical
  Assessment of
  Structure
  Prediction



CASP
Predictor

- CAFASP:

  Critical
  Assessment of
  Fully
  Automated
  Structure
  Prediction

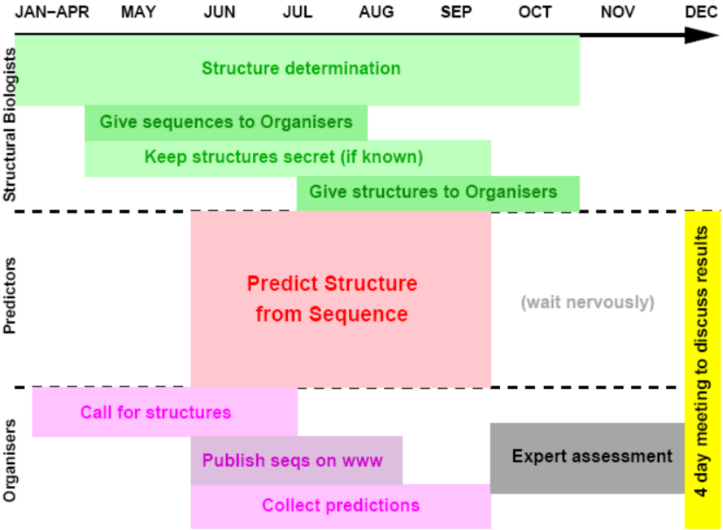
CAFASP
Predictor

1. Won't get tired
2. High-throughput

CASP/CAFASP

- Public
  - Organized by structure community
  - Evaluated by the unbiased third-party
  - Held every two years
- Blind:
  - Experimental structures to be determined by structure centers after competition
- Drawback: <100 targets
  - Blindness
  - Some centers are reluctant to release their structures

CASP/CAFASP Schedule



Test Protein Category

- New Fold (NF) targets
    - No similar fold in PDB
- Homology
    - Modeling (HM) targets
    - Easy HM: has a homologous protein in PDB
    - Hard HM: has a distant homologous protein in PDB
    - Also called Comparative Modeling (CM) targets
- Fold Recognition (FR) targets
    - Has a similar fold in PDB

Protein Structure Prediction

- Stage 1: Backbone Prediction
  - Ab initio prediction
  - Homology modeling
  - Protein threading
- Stage 2: Loop Modeling
- Stage 3: Side-Chain Packing
- Stage 4: Structure Refinement

Protein Structure Prediction

- Stage 1: Backbone Prediction
  - Ab initio prediction
  - Homology modeling
  - Protein threading
- Stage 2: Loop Modeling
- Stage 3: Side-Chain Packing
- Stage 4: Structure Refinement

Ab-initio Prediction:
Sampling the global conformation space

- Lattice models / Discrete-state models
- Molecular Dynamics
- Fragment assembly
  from pre-set library of 3D motifs (=fragments)

Ab-initio Prediction: Sampling the global conformation space

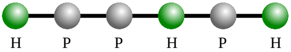
- Lattice models / Discrete-state models
- Molecular Dynamics
- Fragment assembly
  from pre-set library of 3D motifs (=fragments)

Lattice Models: The Simplest Protein Model

The HP-Model (Lau & Dill, 1989)

- model only hydrophobic interaction
  - alphabet $\{H, P\}$; H/P = hydrophobic/polar
  - energy function favors HH-contacts
- structures are discrete, simple, and 2D
  - model only backbone (C-α) positions
  - structures are drawn on a square lattice $\mathbb{Z}^2$ without overlaps: Self-Avoiding Walk

Example

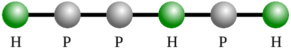


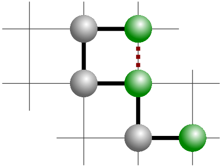
Lattice Models: The Simplest Protein Model

The HP-Model (Lau & Dill, 1989)

- model only hydrophobic interaction
  - alphabet $\{H, P\}$; H/P = hydrophobic/polar
  - energy function favors HH-contacts
- structures are discrete, simple, and 2D
  - model only backbone (C-α) positions
  - structures are drawn on a square lattice $\mathbb{Z}^2$ without overlaps: Self-Avoiding Walk

Example



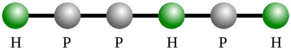


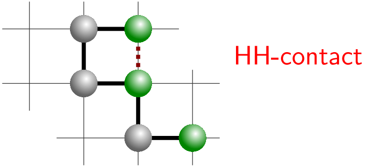
Lattice Models: The Simplest Protein Model

The HP-Model (Lau & Dill, 1989)

- model only hydrophobic interaction
  - alphabet $\{H, P\}$; H/P = hydrophobic/polar
  - energy function favors HH-contacts
- structures are discrete, simple, and 2D
  - model only backbone (C-α) positions
  - structures are drawn on a square lattice $\mathbb{Z}^2$ without overlaps: Self-Avoiding Walk

Example





Lattice Models: Discrete Structure Space

Structure space of a sequence = set of possible structures

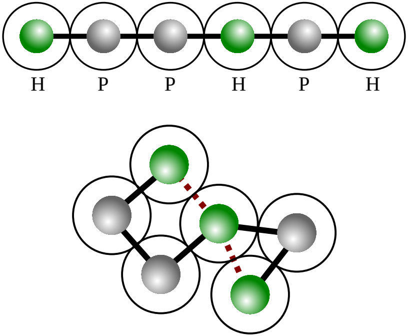
Lattices

- Lattice discretizes the structure space
- Structures can be enumerated
- Structure prediction gets combinatorial problem

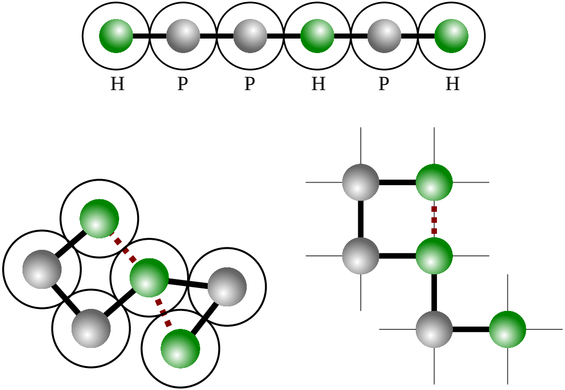
Discrete Structure Space Without Lattice: Off-lattice models

- discrete rotational ϕ/ψ-angles of the backbone
- fragment library
- related idea: Tangent Sphere Model

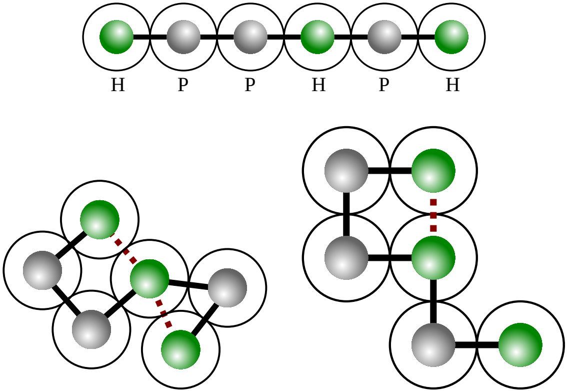
Tangent Sphere Model



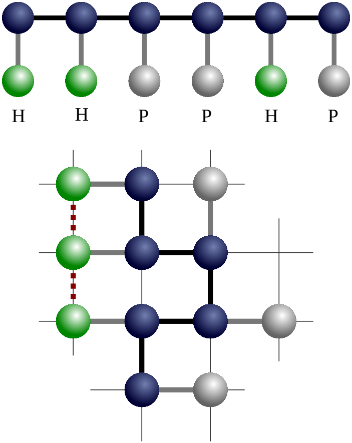
Tangent Sphere Model



Tangent Sphere Model



Side chain models



Lattices

Definition

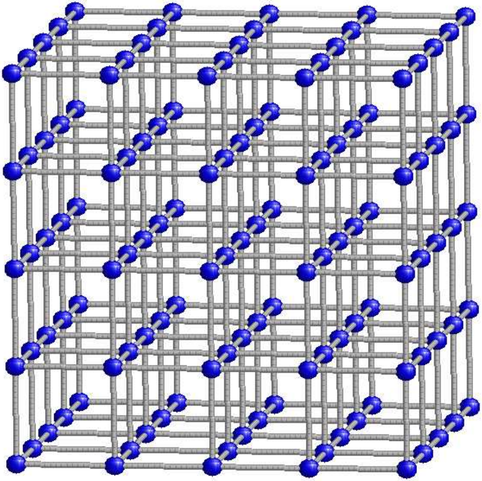
A *lattice* is a set L of *lattice points* such that

$$\vec{0} \in L$$

$$\vec{u}, \vec{v} \in L \text{ implies } \vec{u} + \vec{v}, \vec{u} - \vec{v} \in L$$

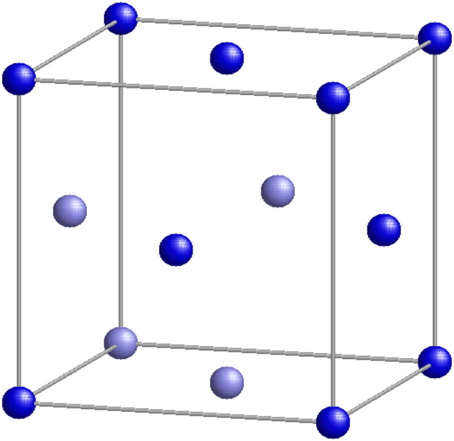
Cubic Lattice

Cubic Lattice $= \mathbb{Z}^3$



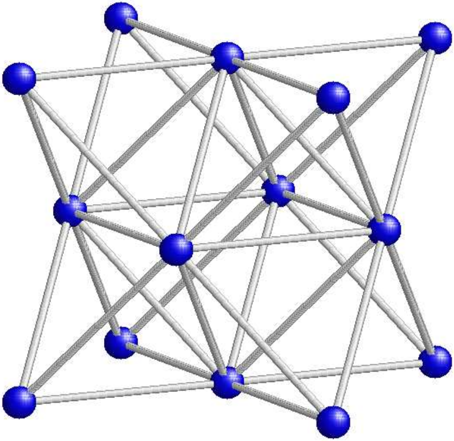
Face-Centered Cubic Lattice (FCC)

$$\mathsf{FCC} = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{Z}^3 \mid x + y + z \text{ even} \right\}$$



Face-Centered Cubic Lattice (FCC)

$$\mathrm{FCC} = \{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{Z}^3 \mid x + y + z \text{ even} \}$$



The Best Lattice?

- Use protein structures from database PDB
- Generate best approximation on lattice
- Compare off-lattice and on-lattice structure

Measures

$$cRMSD(\omega,\omega') = \sqrt{\frac{1}{n}\sum_{1\leq i\leq n} \|\omega(i)-\omega'(i)\|^2}$$

$$dRMSD(\omega,\omega') = \sqrt{\frac{1}{n(n-1)/2}\sum_{1\leq i<j\leq n} (D_{ij}-D'_{ij})^2}$$

$$D_{ij} = \|\omega(i)-\omega(j)\|$$

$$D'_{ij} = \|\omega'(i)-\omega'(j)\|$$

Lattice Approximation - Some Results

Study by Park and Levitt

Lattice	dRMSD	cRMSD
cubic	2.84	2.34
body-centered cubic (BCC)	2.59	2.14
face-centered cubic (FCC)	1.78	1.46

Conclusion

Approximation depends almost only on complexity of the model

Britt H. Park, Michael Levitt. The complexity and accuracy of discrete state models of protein structure Journal of Molecular Biology, 1995

Lattice Approximation - Some Results

Study by Park and Levitt

Lattice	dRMSD	cRMSD
cubic	2.84	2.34
body-centered cubic (BCC)	2.59	2.14
face-centered cubic (FCC)	1.78	1.46

Conclusion

Approximation depends almost only on complexity of the model

Britt H. Park, Michael Levitt. The complexity and accuracy of discrete state models of protein structure Journal of Molecular Biology, 1995

Lattice/Discrete Models: Pairwise Potentials

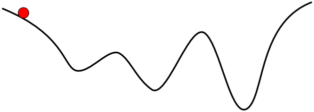
- Ab-initio Potentials
  - HP
  - HPNX
    (H=Hydrophobic, P=Postive, N=Negative, X=Neutral)
- Statistical Potentials: 20× 20 amino acids
  - quasi-chemical approximation (Myiazawa-Jernigan)
  - potential of mean force (Sippl)

Miyazawa S, Jernigan R (1985) Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. Macromolecules

Sippl MJ (1990) Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. J Mol Biol.

Stochastic Local Search

Simulated Annealing & Genetic Algorithms

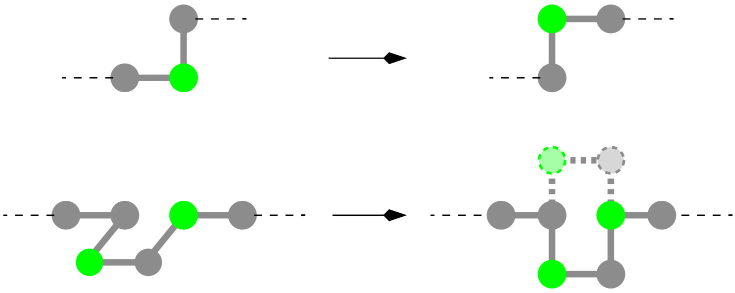


- Applicable to simple or complex protein models
- Heuristic search methods
- Find local optima in energy landscape
- Even for simple models: cannot prove optimality

Move Sets: Local Moves and Pivot Moves

- Stochastic search systematically generates new structures from existing structures
- Idea: new structures are neighbors in the structure space
- New structures generated by applying moves from a *move set*
  - local moves
  - pivot moves

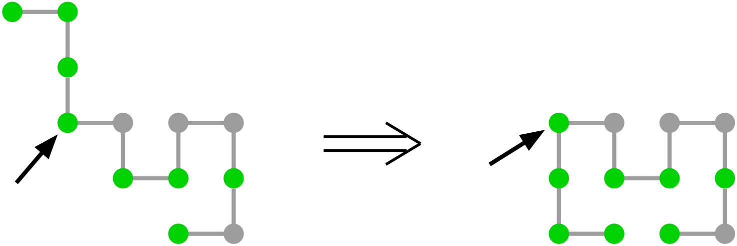
Local Moves



Explanation

A local move changes the positions of a bounded number of monomers at a time.

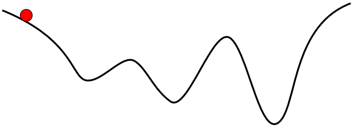
Pivot Moves



Explanation

A pivot move rotates (and/or reflects) a prefix structure $\omega(1)..\omega(i)$ around $\omega(i)$.

Simulated Annealing — Idea



- Perform a random walk through the structure space by repeatedly applying random moves
- Prefer going to better structures
- Sometimes allow going to worse structures
  depends on temperature T
  high T: accept almost all structures
  low T: accept almost only better structures

Simulated Annealing — Algorithm

Find an optimal structure for sequence s (temperature T)

- Start with random structure ω
- Perform simulation steps
    - apply a random local move to $\omega \to \omega'$
    - only accept new structure, i.e. $\omega := \omega'$
        - either if $E(s, \omega') < E(s, \omega)$
        - or with probability

$$\exp(-\frac{(E(s,\omega') - E(s,\omega))}{T})$$

- (Cool the temperature down)

Remarks

- Acceptance rule = Metropolis criterion
- Guarantee for finding the global optimum only for exponentially slow cooling. Otherwise: we don't know.

(Hybrid) Genetic Algorithm — Idea

- Extend the idea of simulated annealing to population of structures
- New structures are generated from existing by
  - Mutation = random pivot move
  - Crossover = random merging two structures

The (Hybrid) Genetic Algorithm [Unger& Moult]

Find an optimal structure for sequence *s*

- Generate random start population (e.g. 200 structures)
- Repeat
  - Mutate all structures
  - Generate offspring population by crossover
  - Accept offspring only due to Metropolis criterion
    (Here: the energy of each offspring is compared to average energy in population.)

R Unger and J Moult. Local interactions dominate folding in a simple protein model. Journal of Molecular Biology, 1996.

Molecular Dynamics

- Simulates the motion of a protein
  considering forces between atoms;
  sounds like the ultimate solution
- Uses force field potentials (e.g. AMBER, CHARMM)

$$\begin{aligned}
E_{total} &= E_{bonded} + E_{nonbonded} \\
E_{bonded} &= E_{bond-stretch} + E_{angle-bend} + E_{rotation-along-bond} \\
E_{nonbonded} &= E_{electrostatic} + E_{van-der-Waals}
\end{aligned}$$

- Applies Newton's laws of motion
- Changes are calculated for small time steps
  - small enough to avoid discretization error
    smaller than vibration of system
    $\Rightarrow$ in order of femtoseconds $= 10^{-15}$ seconds!
  - computationally intensive
  - critical for simulation time

Molecular Dynamics: Limits

- Simulation gap
  Assume one billion steps: $10^{-15} \times 10^{9}$ is still 10^{-6}
  For folding small proteins, we need at least millisecond
- force fields empirical (from comparably small molecules)
  valid for protein folding case?
  (*"embarrassment of molecular mechanics"*)
- Newton's equations solved numerically (instabilities)
- explicit/implicit solvent
- Quantum MD
- Pair potential/many-body potentials

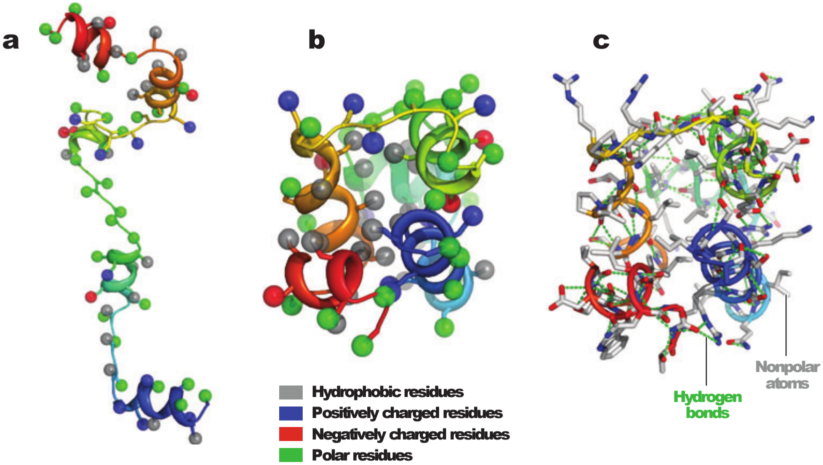
Limitations of MD are not exclusively a matter of computational resources

Fragment Assembly: Rosetta

- Monte Carlo search in coarse grained model
- Limit conformational search space by using 9mer motifs
- Rationale
  - Local structures often fold independently of full protein
  - Can predict large areas of protein by matching sequence to motifs

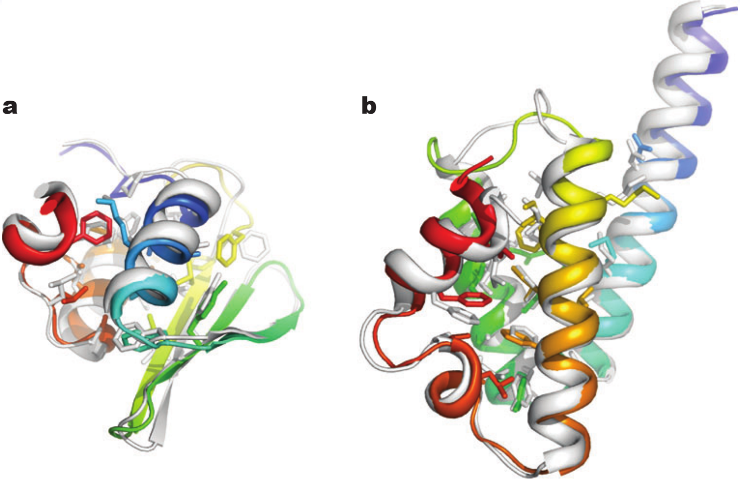
- New structures generated by swapping compatible fragments
- Select candidates for refinement
  - Accepted structures are clustered based on energy and structural size
  - Best cluster is one with the greatest number of conformations within N- rms deviation structure of the center
  - Representative structures taken from each of the best five clusters and returned to the user as predictions

Rosetta: Fragment Assembly and Refinement



Rhiju Das and David Baker. Macromolecular Modeling with Rosetta. Annu. Rev. Biochem, 2008.

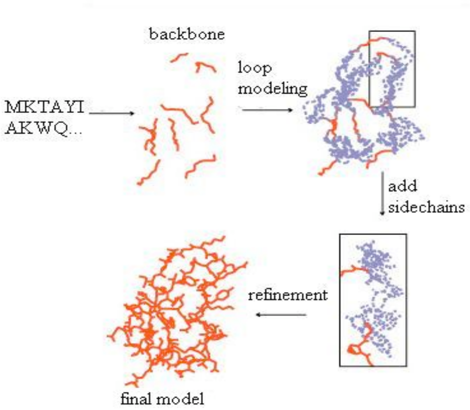
Rosetta *de-novo* Blind Prediction Results (CASP6)



atomic level prediction, < 2 Å; a/b: 70/90 residues, 1.6/1.4 Å

More of Rosetta: Rosetta@home Protein Folding, Design, and Docking Foldit

Protein Structure Prediction



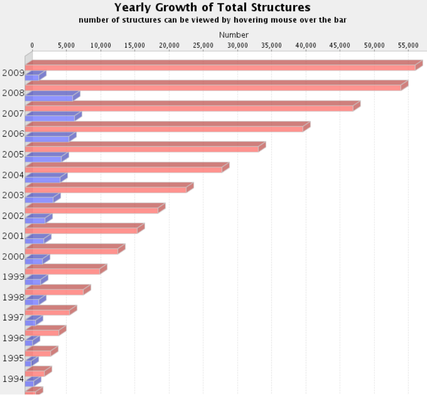
- Stage 1: Backbone Prediction
  - Ab initio folding
  - **Homology modeling**
  - Protein threading
- Stage 2: Loop Modeling
- Stage 3: Side-Chain Packing
- Stage 4: Structure Refinement

The picture is adapted from http://www.cs.ucdavis.edu/~koehl/ProModel/fillgap.html

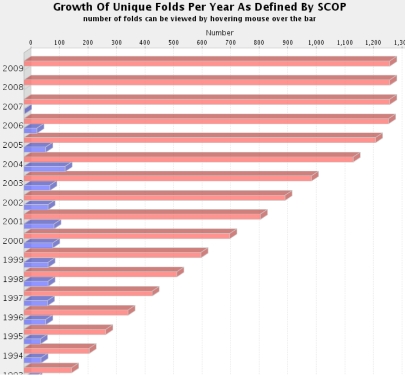
Sometimes grouped “Comparative Modeling”

- Homology modeling
  - identification of homologous proteins through sequence alignment
  - structure prediction through placing residues into “corresponding” positions of homologous structure models
- Protein threading
  - make structure prediction through identification of “good” sequence-structure fit

PDB New Fold Growth

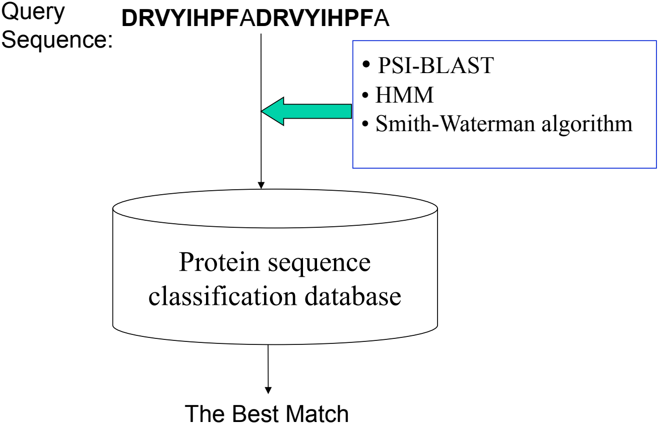

Yearly Growth of Total Structures

number of structures can be viewed by hovering mouse over the bar


Growth Of Unique Folds Per Year As Defined By SCOP

number of folds can be viewed by hovering mouse over the bar

Homology Modeling



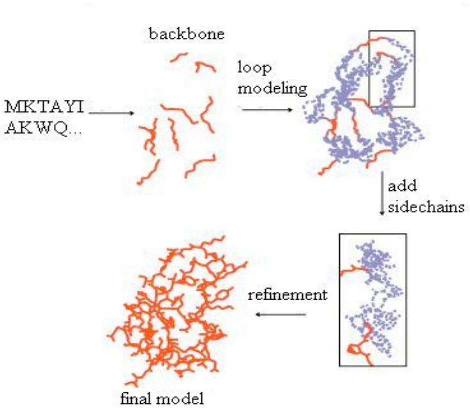
Query Sequence: **DRVYIHPF**A**DRVYIHPF**A

- PSI-BLAST
- HMM
- Smith-Waterman algorithm

Protein sequence classification database

The Best Match

Protein Structure Prediction



- Stage 1: Backbone Prediction
  - Ab initio folding
  - Homology modeling
  - **Protein threading**
- Stage 2: Loop Modeling
- Stage 3: Side-Chain Packing
- Stage 4: Structure Refinement

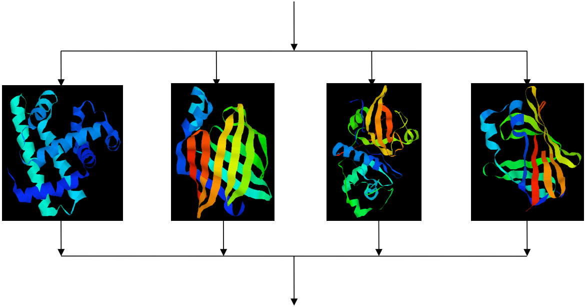
The picture is adapted from http://www.cs.ucdavis.edu/~koehl/ProModel/fillgap.html

Protein Threading

- Make a structure prediction through finding an optimal alignment (placement) of a protein sequence onto each known structure (structural template)
  - “alignment” quality is measured by some statistics-based scoring function
  - best overall “alignment” among all templates may give a structure prediction
- Step 1: Construction of Template Library
- Step 2: Design of Scoring Function
- Step 3: Alignment
- Step 4: Template Selection and Model Construction

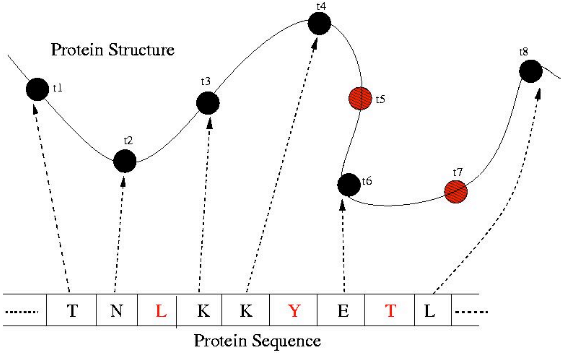
Protein Threading

Query Sequence: **DRVYIHPF**A**DRVYIHPF**A



The Best Match

Protein Threading



Positions or residues in red are gaps

Threading Model

- Each template is parsed as a chain of cores. Two adjacent cores are connected by a loop. Cores are the most conserved segments in a protein.
- No gap allowed within a core.
- Only the pairwise contact between two core residues are considered because contacts involved with loop residues are not conserved well.
- Global alignment employed

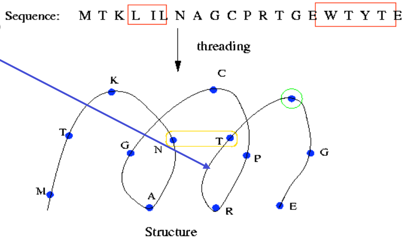
Scoring Function

how preferable to put two particular residues nearby: E_p

(**Pairwise potential**)

alignment gap penalty: E_g

(**gap score**)



how well a residue fits a structural environment: E_s

(**Fitness score**)

sequence similarity between query and template proteins: E_m

(**Mutation score**)

How consistent of the secondary structures: E_ss

E= E_p +E_s +E_m +E_g +E_ss

Minimize E to find a sequence-template alignment

Scoring: Fitness Score

occurring probability of amino acid a with s

$$FitnessScore(a,s) = -\log \frac{P(a,s)}{P(a)P(s)}$$

occurring probability of amino acid a

occurring probability of solvent accessibility s

Scoring: Pairwise Potential

occurring probability of a and b with distance < cutoff

$$PairwisePotential(a,b) = -\log \frac{P(a,b)}{P(a)P(b)}$$

occurring probability of amino acid a

occurring probability of amino acid b

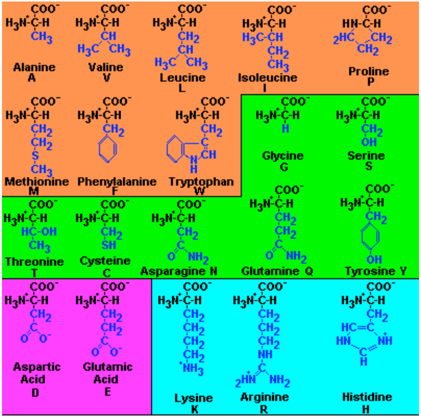
Scoring: Secondary Structure

1. Difference between predicted secondary structure and template secondary structure

2. PSIPRED for secondary structure prediction

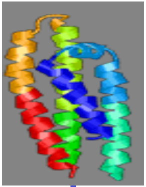
Scoring: Mutational Score

Could be based on chemical similarity, etc, etc.

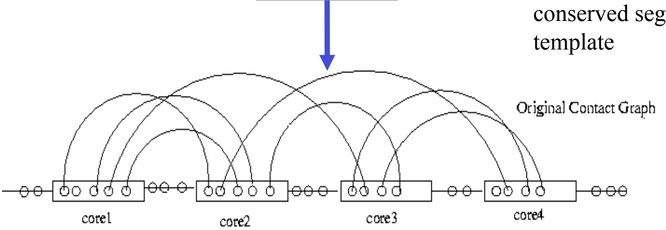


Contact Graph

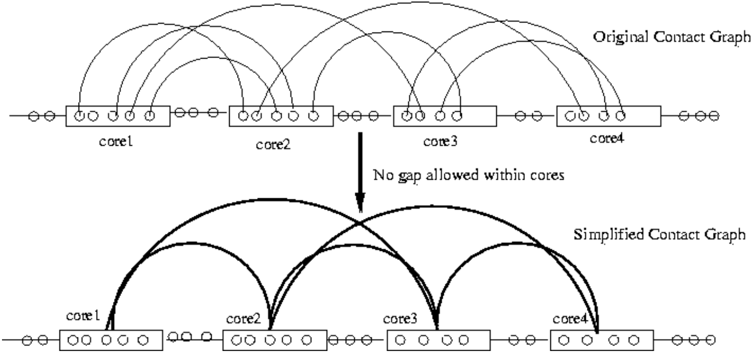
template



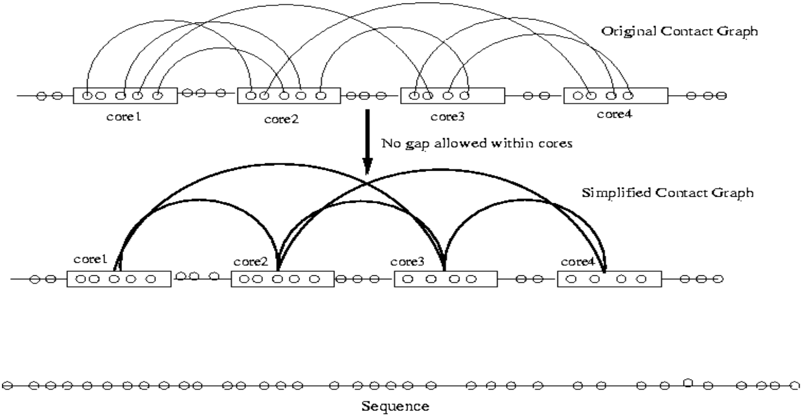
1. Each residue as a vertex
2. One edge between two residues if their spatial distance is within given cutoff.
3. Cores are the most conserved segments in the template



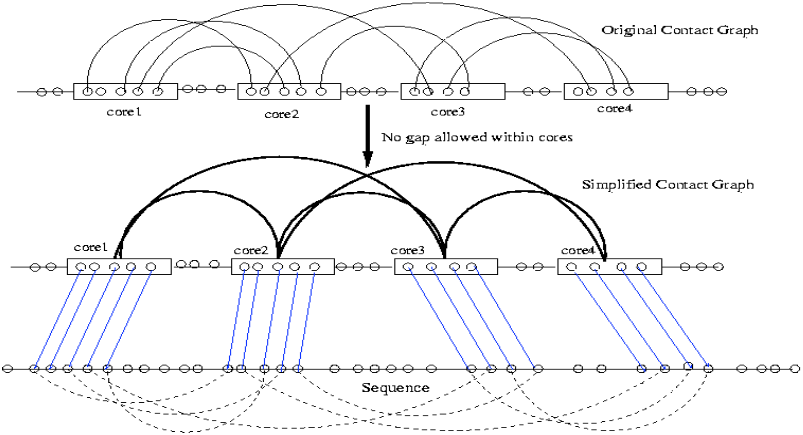
Simplified Contact Graph



Alignment Example



Alignment Example

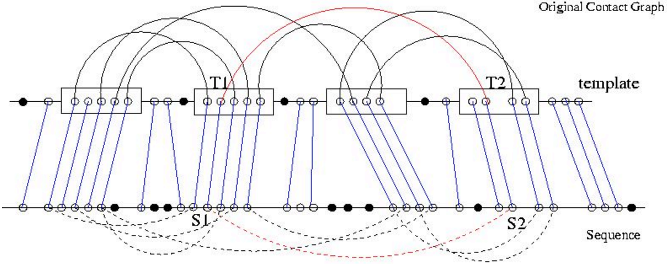


Calculation of Alignment Score

Alignment Score=Singleton Score + Pairwise Score+Gap Penalty

Singleton Score (S1,T1)=Mutation Score(S1,T1)+Fitness Score(S1,T1) + SS(S1,T1)

Pairwise Score = Pairwise Score (S1, S2, dist(T1,T2))+....

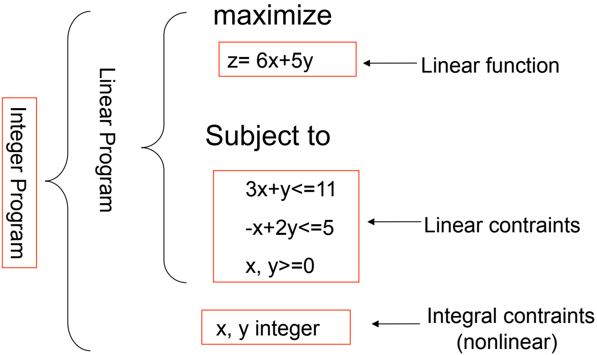


Filled small circles are unaligned template positions or sequence residues

Threading Algorithms

- NP-Hard problem
  - Can be reduced to MAX-CUT
- Approximation Algorithm
  - Interaction-frozen algorithm (A. Godzik et al.)
  - Monte Carlo sampling (S.H. Bryant et al.)
  - Double dynamic programming (D. Jones et al.)
- Exact Algorithm
  - Branch-and-bound (R.H. Lathrop and T.F. Smith)
  - PROSPECT-I uses divide-and-conquer (Y. Xu et al.)
  - Linear programming by RAPTOR (J. Xu et al.)

Linear & Integer Program



Integer Program

Linear Program

maximize

z= 6x+5y

Linear function

Subject to

3x+y<=11

-x+2y<=5

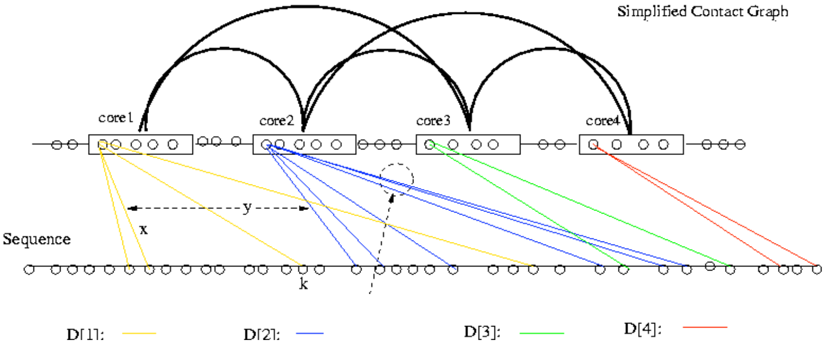
x, y>=0

Linear contraints

x, y integer

Integral contraints (nonlinear)

Variables



- x(i,l) denotes core i is aligned to sequence position l
- y(i,l,j,k) denotes that core i is aligned to position l and core j is aligned to position k at the same time.

LP Formulation

$$\mathit{Minimize}$$

$$E = \sum a_{i,l} x_{i,l} + \sum b_{(i,l)(j,k)} y_{(i,l)(j,k)}$$

$$s.t.$$

$$\left.\begin{aligned} x_{i,l} &= \sum_{k \in R[i,j,l]} y_{(i,l)(j,k)}, \forall l \in D[i] \\ x_{j,k} &= \sum_{l \in R[j,k,i]} y_{(i,l)(j,k)}, \forall k \in D[j] \end{aligned}\right\}$$

$$\sum_{l \in D[i]} x_{i,l} = 1$$

$$x_{i,l}, y_{(i,l)(j,k)} \in \{0,1\}$$

a: singleton score parameter

b: pairwise score parameter

Each y variable is 1 if and only if its two x variable are 1

Each core has only one alignment position

Online Servers

RAPTOR

http://www.bioinformatics.uwaterloo.ca/~j3xu/raptor/index.php

www.bakerlab.org

ROBETTA

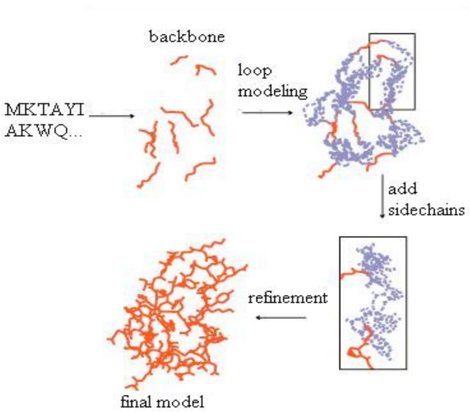
Full-chain Protein Structure Prediction Server

http://robetta.bakerlab.org/index.html

phyre

http://www.sbg.bio.ic.ac.uk/~phyre/

Protein Structure Prediction

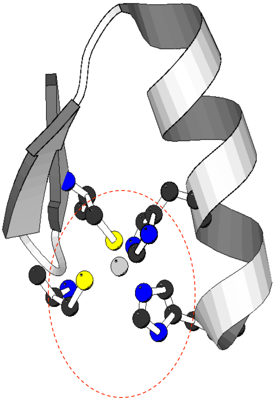


The picture is adapted from http://www.cs.ucdavis.edu/~koehl/ProModel/fillgap.html

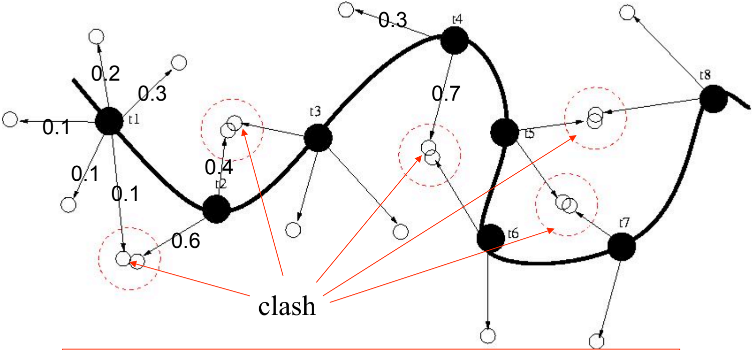
- Stage 1: Backbone Prediction
  - Ab initio folding
  - Homology modeling
  - Protein threading
- Stage 2: Loop Modeling
- Stage 3: Side-Chain Packing
- Stage 4: Structure Refinement

Protein Side-Chain Packing

- Problem: given the backbone coordinates of a protein, predict the coordinates of the side-chain atoms
- Insight: a protein structure is a geometric object with special features
- Method: decompose a protein structure into some very small blocks



Side-Chain Packing



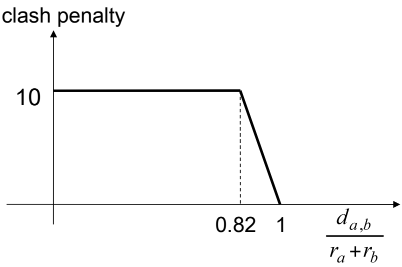
Each residue has many possible side-chain positions.
Each possible position is called a rotamer.
Need to avoid atomic clashes.

Energy Function

Assume rotamer A(i) is assigned to residue i. The side-chain packing quality is measured by

$$\underbrace{\sum_i S(i, A(i))}_{\text{occurring preference}} + \underbrace{\sum P(i, j, A(i), A(j))}_{\text{clash penalty}}$$

occurring preference
The higher the occurring probability, the smaller the value



$d_{a,b}$: distance between two atoms

r_a, r_b :atom radii

Minimize the energy function to obtain the best side-chain packing.

Many Methods

- NP-hard [Akutsu, 1997; Pierce et al., 2002] and NP-complete to achieve an approximation ratio O(N) [Chazelle et al, 2004]
- Dead-End Elimination: eliminate rotamers one-by-one
- SCWRL: biconnected decomposition of a protein structure [Dunbrack et al., 2003]
  - One of the most popular side-chain packing programs
- Linear integer programming [Althaus et al, 2000; Eriksson et al, 2001; Kingsford et al, 2004]
  - The formulation similar to that used in RAPTOR

Dead-end elimination

- Conformation consists of N residues, each with a set of r possible rotomers

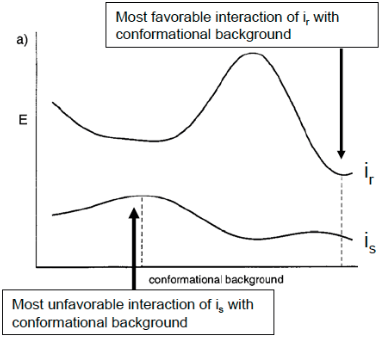
- Simplification:
  Global conformation energy formulated as 2 parts:
  - Sum of all interactions between backbone and N residues
  - Sum of all pairwise interactions between i*i residues

    (residues i, j, rotatmers r, s)

$$E_{total} = \sum_{i=1}^{N} E(i_r) + \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} E(i_r, j_s)$$

Dead-end elimination

- If two rotamers r, s at residue position i
- can eliminate rotamer s, if pairwise energy between i_r and all other sideschains is **always** higher than pairwise energy between i_s and all other sidechains



Eliminate i_r iff:

$$E(i_r) - E(i_s) + \sum_{j \neq i} \min E(i_r, j) + \sum_{j \neq i} \min E(i_s, j) > 0$$

http://www.ch.embnet.org/CoursEMBnet/Pages3D08/slides/SIB-PhD-Day2_p.pdf

Dead-end elimination

- Apply iteratively to all rotamer pairs

- After each elimination, energy landscape changes so could cause new elimination that couldn’t have happened before