Solvent-entropy driven searching for protein modeling examined and tested in simplified models

Rainer Koenig and Thomas Dandekar
Protein Engineering, Vol. 14 No. 5, 329, 335, May 2001

Presented by
Alexandre Ismail, Kate Stafford, Jay Shukla
Current structure prediction methods

Ab initio: too computationally expensive, poor accuracy at large scales

More successful methods: Based on alignments and energetics

Alignments: not really understanding why, but recognizing patterns

Problems: degeneracy in sequences $\rightarrow$ structure mapping
trouble with low complexity sequences (e.g. yeast)
not “complete”
not fully automated
Current structure prediction methods

Energetics: based on conformational energy analysis of the protein

Problems: computationally expensive
require parameterization
cannot predict or discriminate among structures

...Novotny 1984, 1988
Solvent ordering by hydrophobic molecules
Solvent entropy maximization!

Protein structure is the one that maximizes solvent entropy

Driving force: solvent entropy maximization
Accommodating forces: polypeptide sterics, electrostatics, etc.

Not calculating solvent entropy exactly, but representing it.

“The global minimum no longer appears as one rare state among a large number of alternative conformations, but as the protein conformation with the highest number of microstate representations of the solvent.”
Experimental Objective

Compare the performance of different algorithms under simple and demanding conditions.
The model: MC on a lattice

HP polypeptide model: H = \textcolor{dark}{\text{hydrophobic}}
P = \textcolor{light}{\text{hydrophilic}}

OL solvent model: O = \textcolor{light}{\text{ordered}}
L = \textcolor{dark}{\text{less ordered}}
N_i = \text{# of L blocks}
The model: MC on a lattice

2D model: 5 chains; lengths = 12
18
24
33
48

3D model: 4 chains; lengths = 12
14
22
28

Comparing Entropy & Energy: 3 ways to evaluate

A = energy
B = entropy
C = entropy and energy
Simulation Flowchart

- Start from random conformation
- From $C_1$ with solvent entropy $N_1$ and energy $E_1$, a single random change yields $C_2$. Reject and regenerate if $C_2$ clashes.
- $C_2$ evaluated by algorithms A, B, or C for acceptance.
- Continue until $C_{100,000}$
- Repeat 100 times
Move set and scoring

Move set: a single bead random move (non clashing)

Successive conformations = $C_1, C_2, \text{ etc}$

Solvent entropy function,

$$S_0 = 1 \text{ if O}$$
$$S_L = 1/f \text{ if L}$$

$f < 1 = \text{“entropy weight”}$

Polypeptide energy function, $\Delta E = -1$ for direct contact between two non-consecutive hydrophobic beads (i.e. hydrogen bonds)
Entropy vs. Energy

You’re a winner!
Move acceptance algorithms

Three implementations:

A = energy:
accept $C_2$ if $\text{rnd} < e^{-\Delta E/T}$

B = entropy:
accept $C_2$ if $\text{rnd} < K \cdot f^{(N_1-N_2)} = e^{(T\Delta S/T)}$

C = entropy and energy:
accept $C_2$ if $\text{rnd} < K \cdot f^{(N_1-N_2)} e^{-\Delta E/T}$

$K = \text{constant}$
Results

How many times per 100 runs did the algorithm find the native structure?

Algorithm C: best performance

Algorithm B: better performance

Algorithm A: OK performance

Computation speed was better with entropy
2D Native Fold Images

18mer  24mer  33mer
3D Native Fold Images

12mer

14mer
3D Native Fold Images

22mer

28mer
Results: 2D lattice

% Runs Finding Native Fold: 2D lattice

Algorithm

% Runs

A
B
C

0 10 20 30 40 50 60 70 80 90 100 110 120

12
18
24
33
48
Results: 3D lattice

% Runs Finding Native Fold: 3D lattice

Algorithm

% of Runs

A    B    C

0    0    0

12   14   22   28
My Research

**Summer**
MD simulation of dynamics of myoglobin-CO complex with CHARMM and GB continuum solvation. Trajectory analysis by PCA.

**Fall, Winter, Spring...and Summer again**
More detailed solvent-entropy driven lattice models of protein structure and folding.
References


Thanks!

☐ Jay Shukla & Kate Stafford
☐ Dr. Evanseck
☐ Jerome Nilmeier
☐ Rajan Munshi
☐ All BBSI characters
☐ NIH and N$F

I’m leaving you with my children.