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

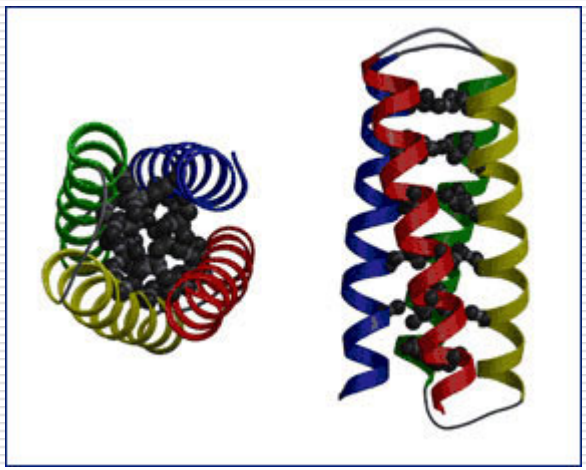
Alignments: not really understanding why, but recognizing patterns

Problems:
- degeneracy in sequences → structure mapping
- trouble with low complexity sequences (e.g. yeast)
- not "complete"
- not fully automated

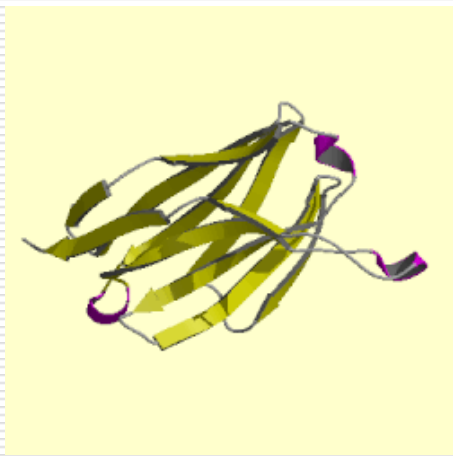
Current structure prediction methods

Energetics: based on conformational energy analysis of the protein

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



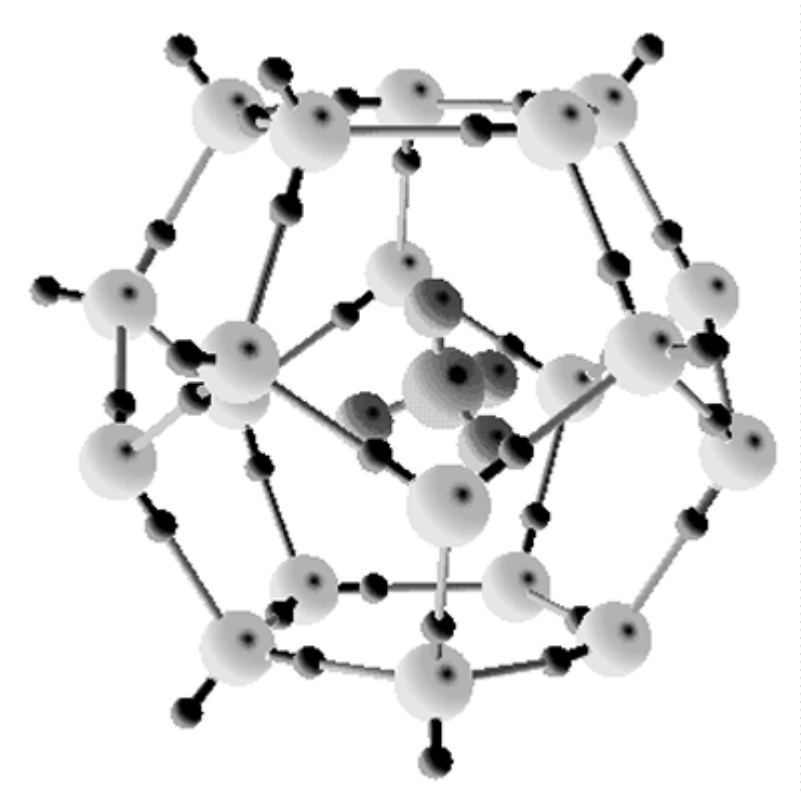
or



?

…Novotny 1984, 1988

Solvent ordering by hydrophobic molecules

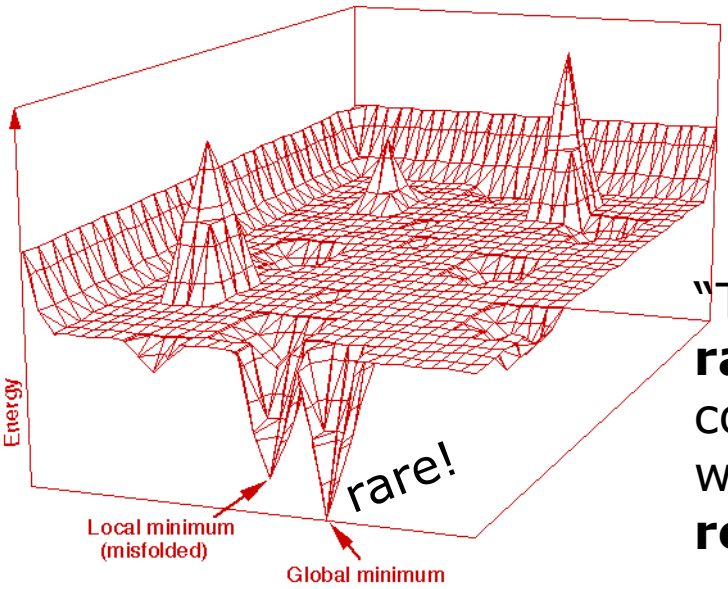


Solvent entropy maximization!

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**."

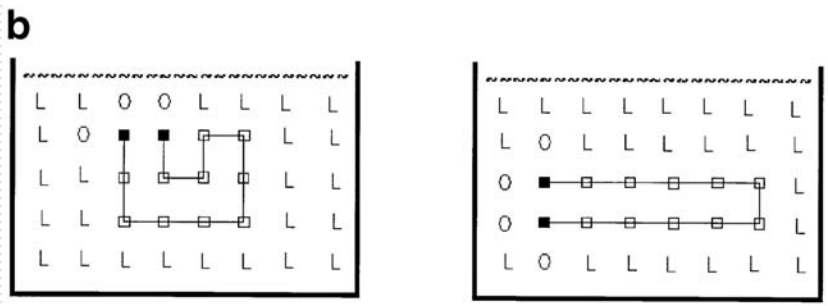
Experimental Objective

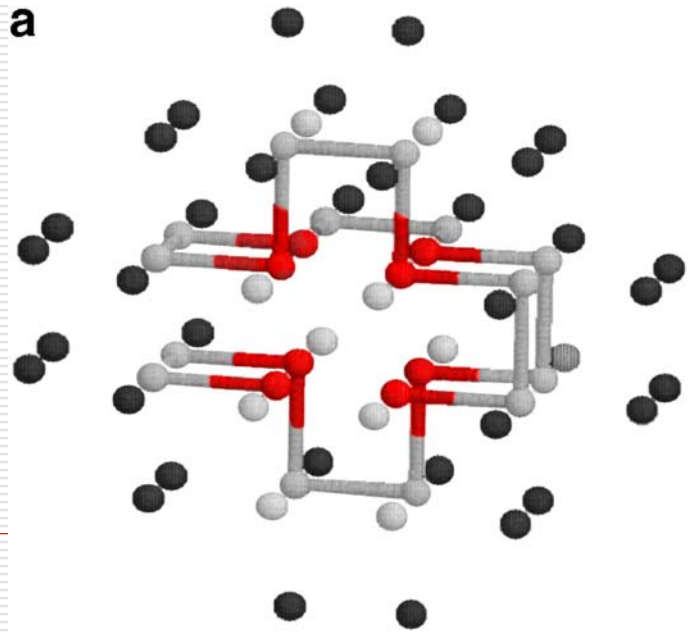
Compare the performance of different algorithms under simple and demanding conditions.

The model: MC on a lattice

HP polypeptide model: H = ***h***ydrophobic [dark]
P = hydro***p***hilic [light]

OL solvent model: O = ***o***rdered [light]
L = ***l***ess ordered [dark]
N_i = #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)

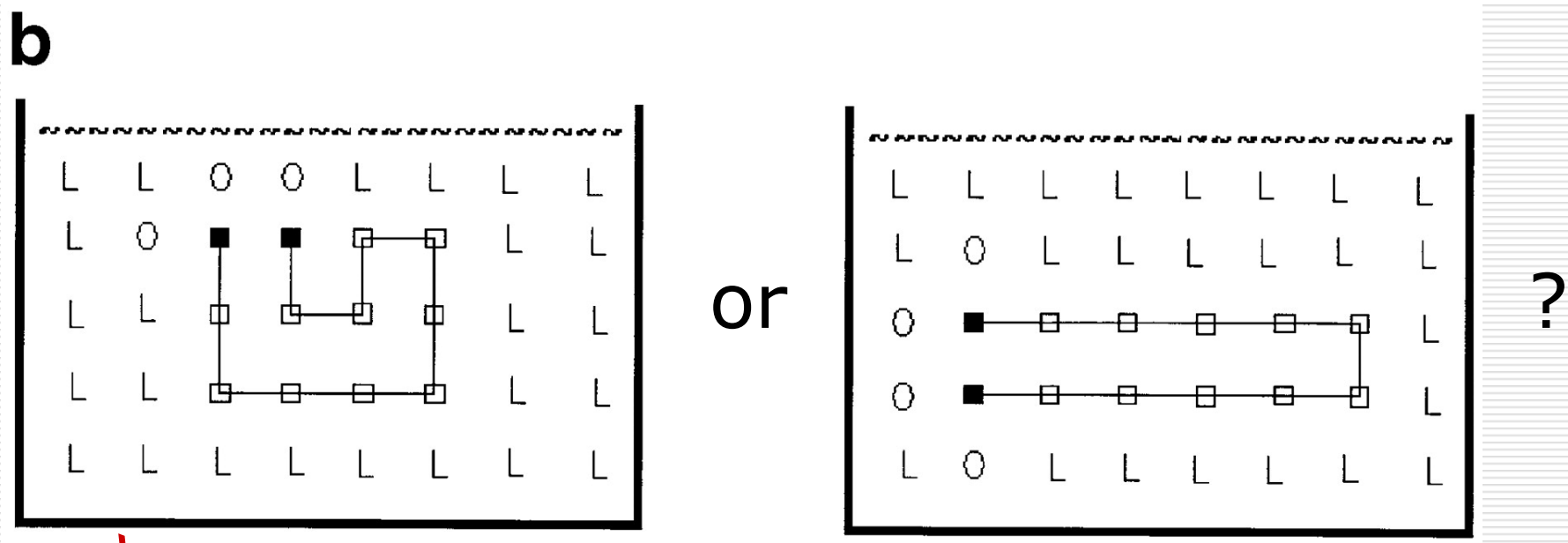
Successive conformations = C_1, C_2, etc

Solvent entropy function, $S_O = 1$ if O
$S_L = 1/f$ if L

$f < 1$ = "entropy weight"

Polypeptide energy function, ΔE = **-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 rnd < $e^{-\Delta E/T}$

Boltzmann style!

B = entropy:

accept C_2 if rnd < $K*f^{(N1-N2)} = e^{(T\Delta S/T)}$

C = entropy and energy:

accept C_2 if rnd < $K*f^{(N1-N2)}e^{-\Delta E/T}$

K = constant

Results

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

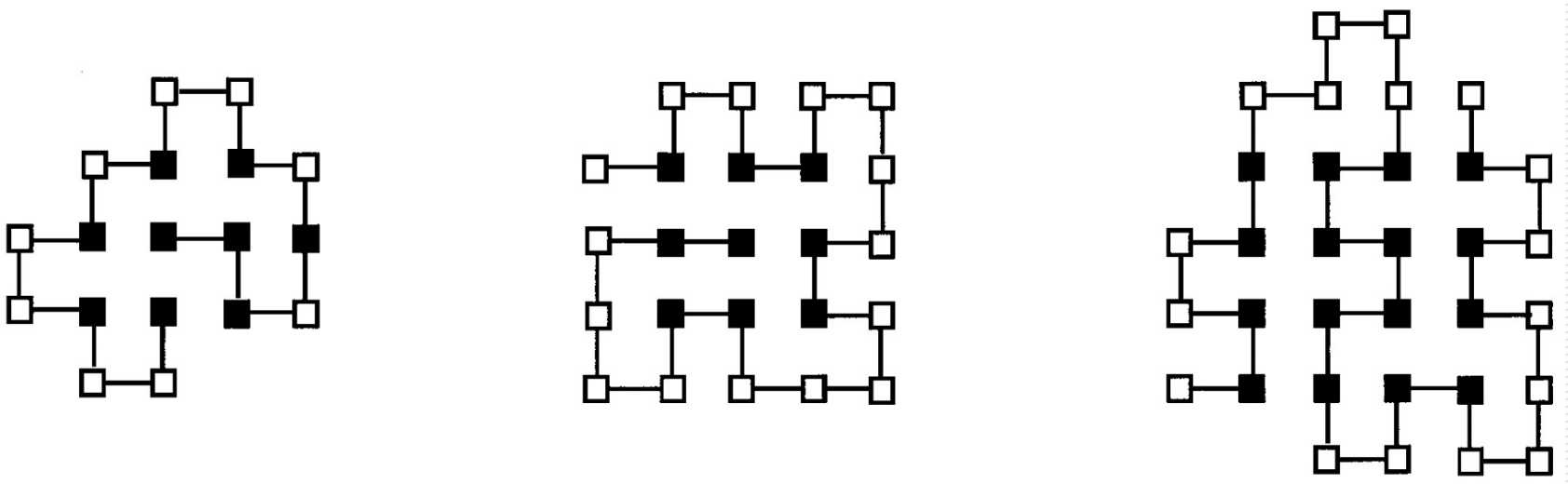
Algorithm C: best performance

Algorithm B: better performance

Algorithm A: OK performance

Computation speed was better with entropy

2D Native Fold Images

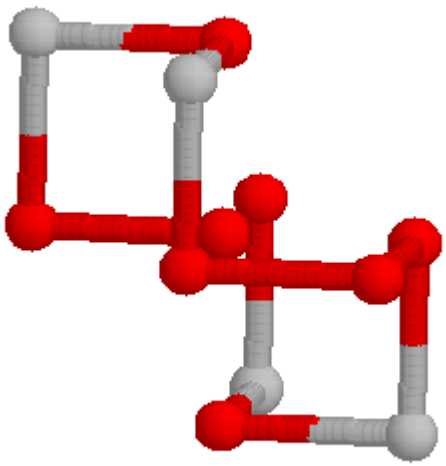


18mer

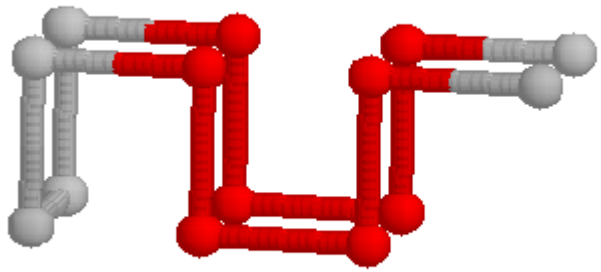
24mer

33mer

3D Native Fold Images

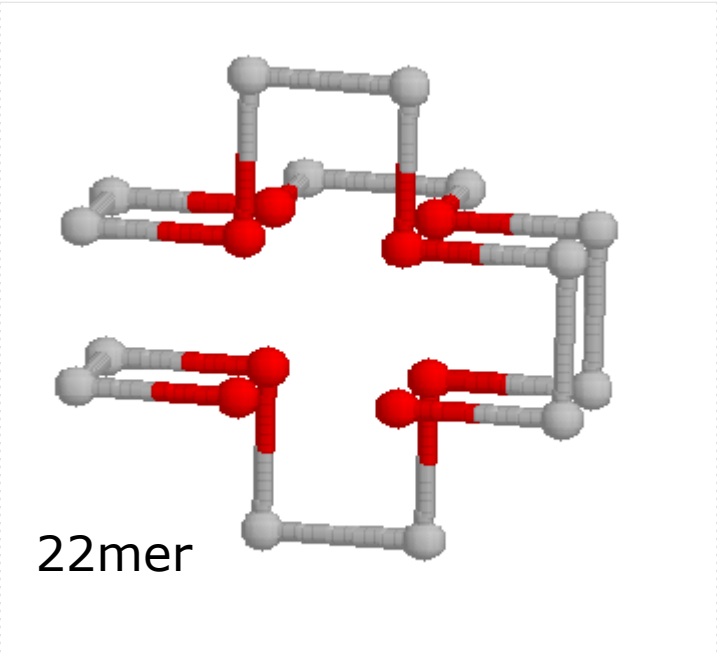


12mer

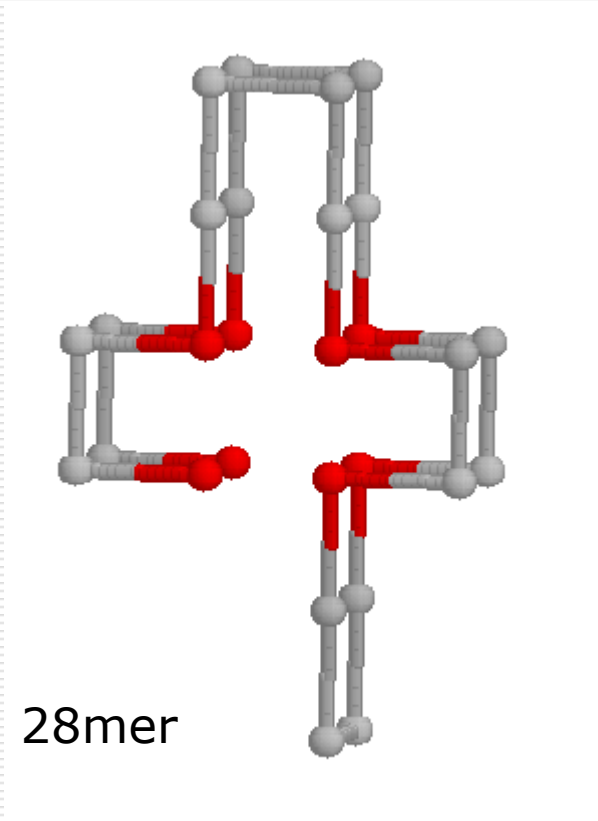


14mer

3D Native Fold Images

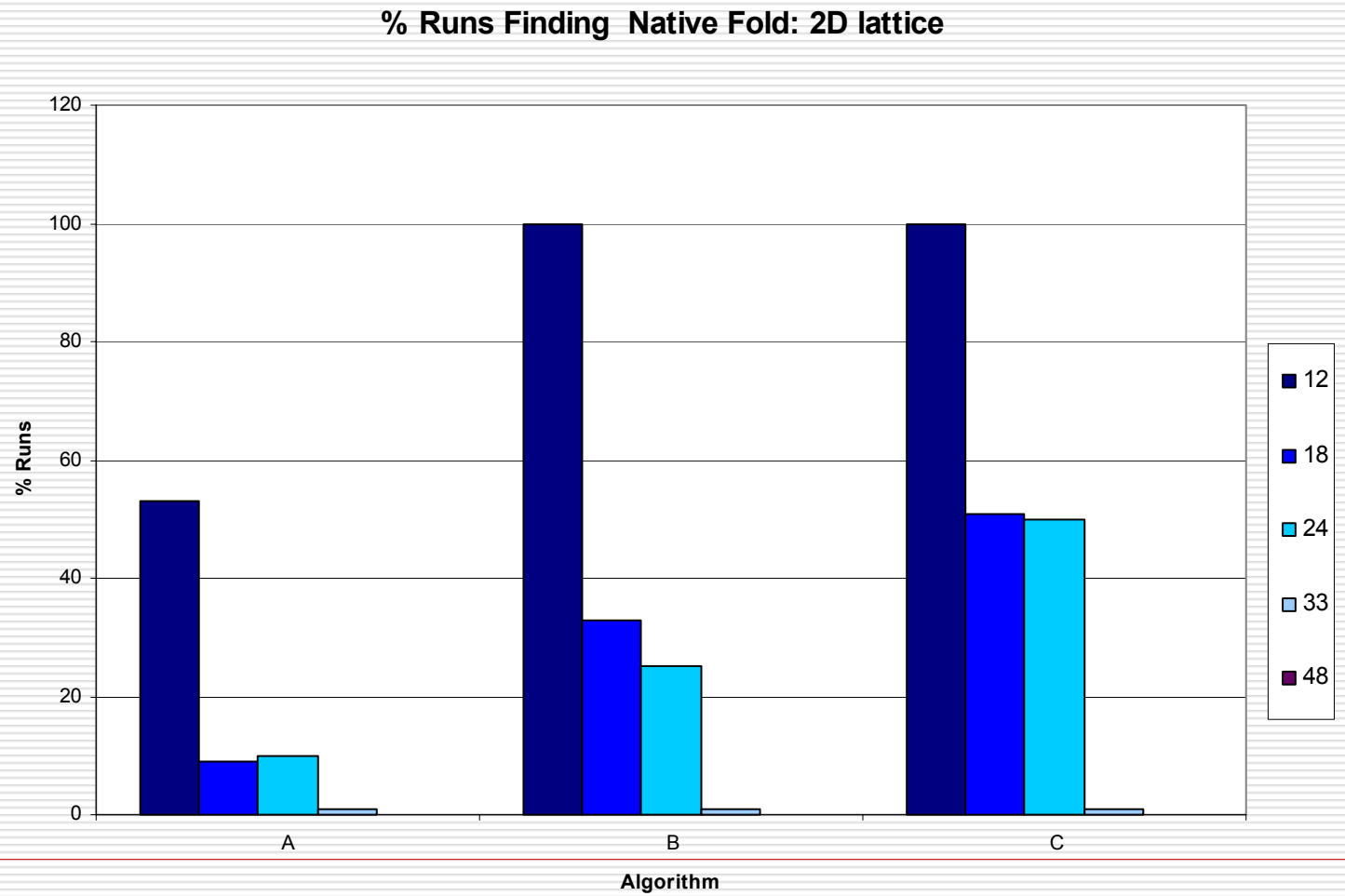


22mer

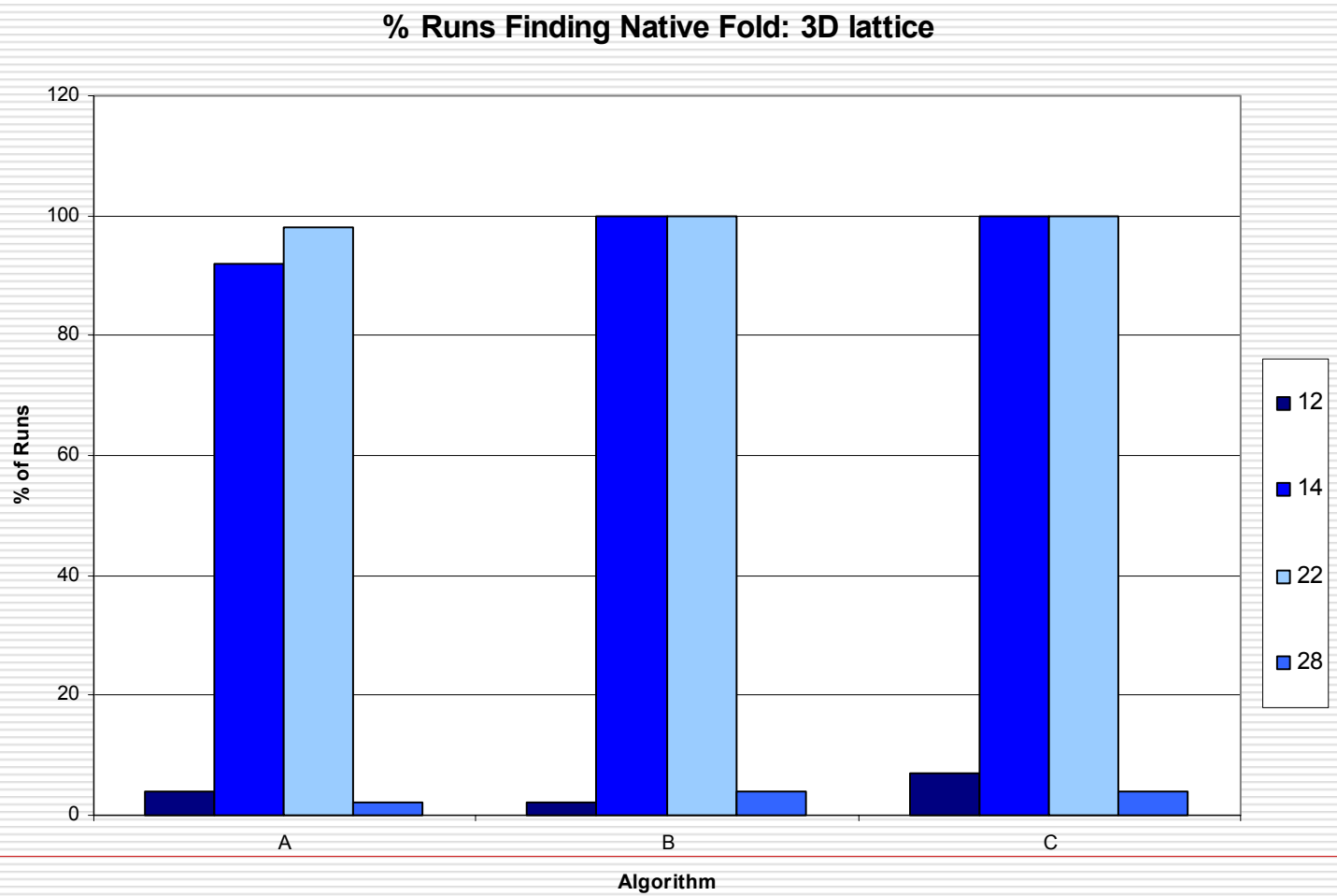


28mer

Results: 2D lattice



Results: 3D lattice



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

- Novotny, J., Bruccoleri, R.E. (1984). An analysis of incorrectly folded protein models; implications for structure prediction. *J. Mol. Biol. 177,* 787-818
- Novotny, J., Rashin, A.A., & Bruccoleri, R.E. (1988). Criteria that discriminate between native proteins and incorrectly folded models. *Proteins Struct. Funct. Genet. 4,* 19-30
- Lau, K.F. and Dill, K.A. (1989) *Macromolecules*, **22**, 3986-3997
- Lau, K.F. and Dill, K.A. (1990) *Proc. Natl Acad. Sci. USA* **87**, 638-642

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