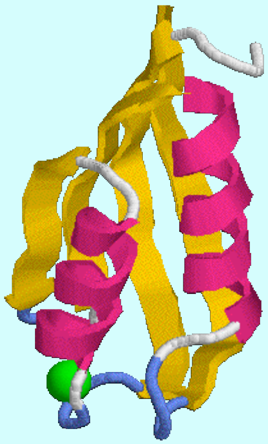
Protein Design Based on the Designability of Protein Structures

Chen Zeng, PI (GWU)
Chao Tang, Co-PI (NEC)
Ned Wingreen (Princeton)
Naigong Zhang (GWU)
Luhua Lai (PKU)

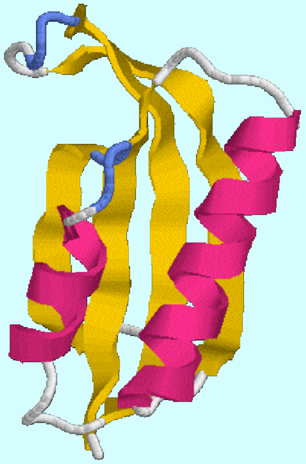
NSF Support:

DMR-0094176, 0313129

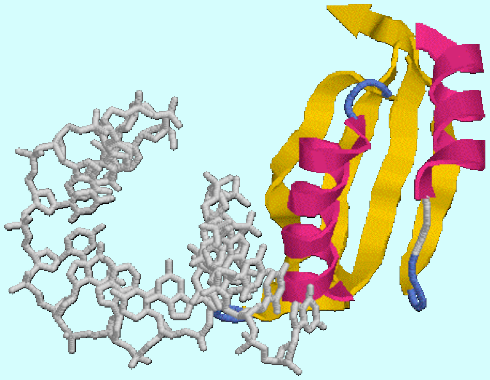
Protein Folds
(~1000! Why so small?)



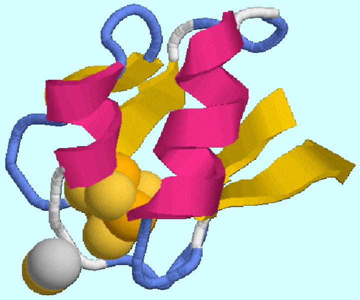
Acylphosphatase



Ribosomal protein S6



Viral DNA-binding domain



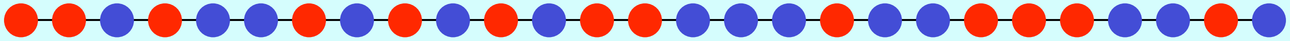
Ferredoxin

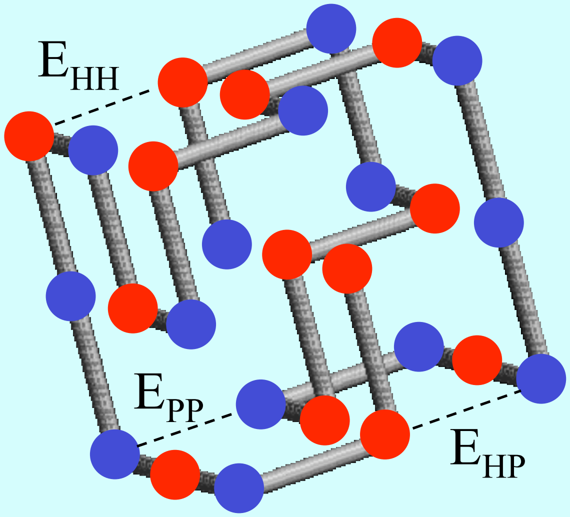
HP Lattice Model

● H--Hydrophobic

● P--Polar

(Dill)



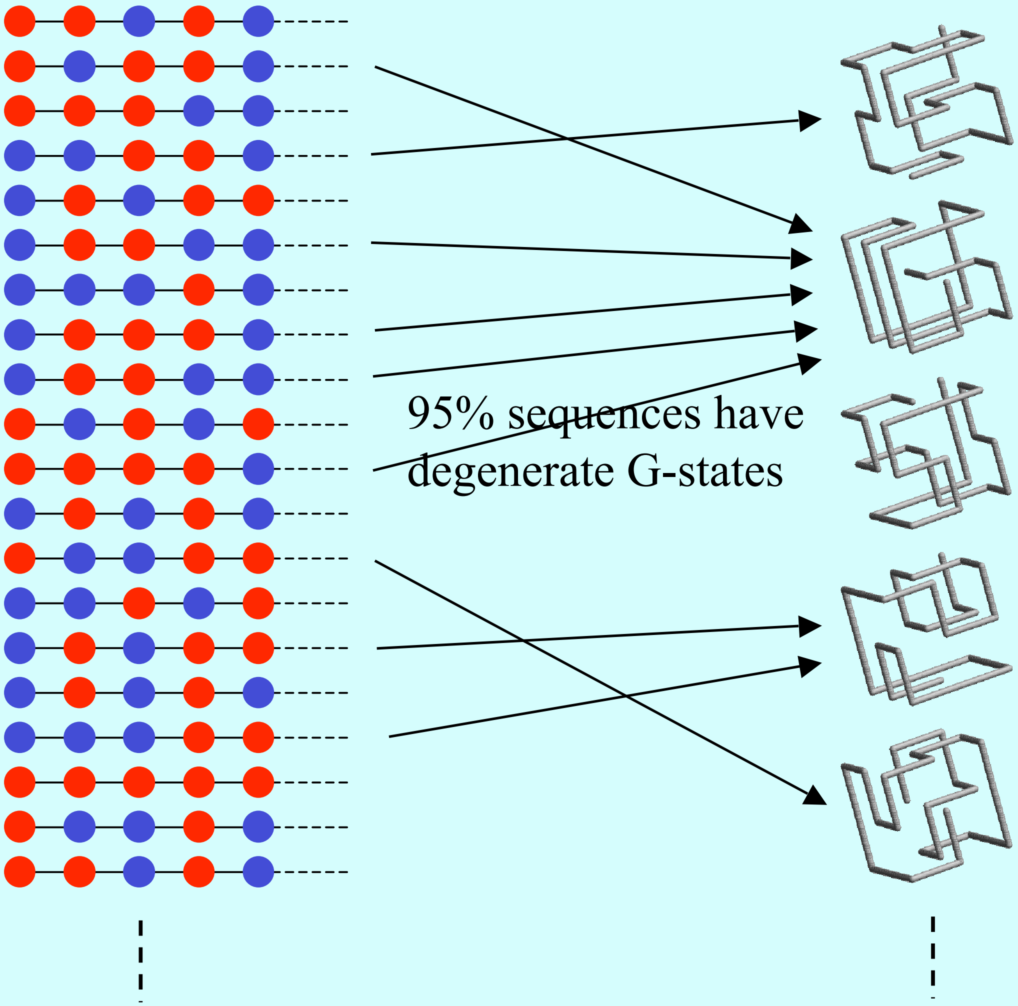


Hydrophobicity
Segregation

Enumeration (HP Lattice Model)

H. Li, C. Tang, N. Wingreen (NEC)

Sequences (10^8) Structures (10^5) Designability



95% sequences have degenerate G-states

$N_S=1$

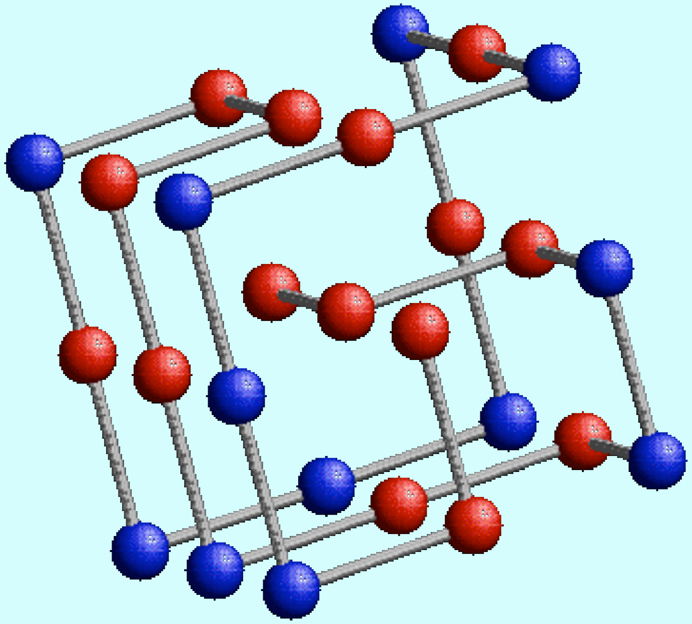
$N_S=5$

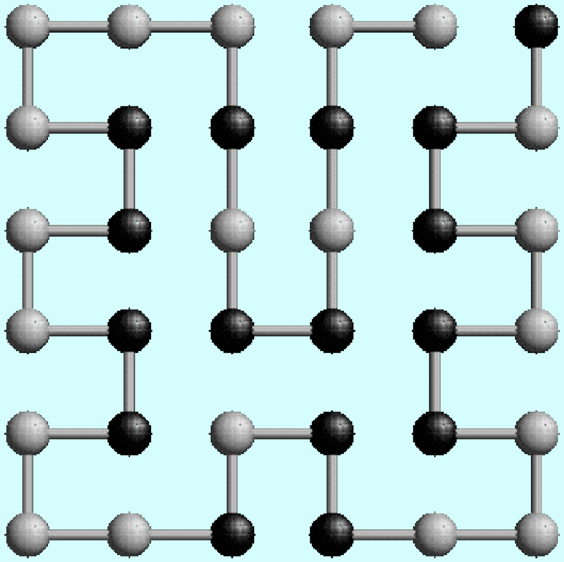
$N_S=0$

$N_S=2$

$N_S=1$

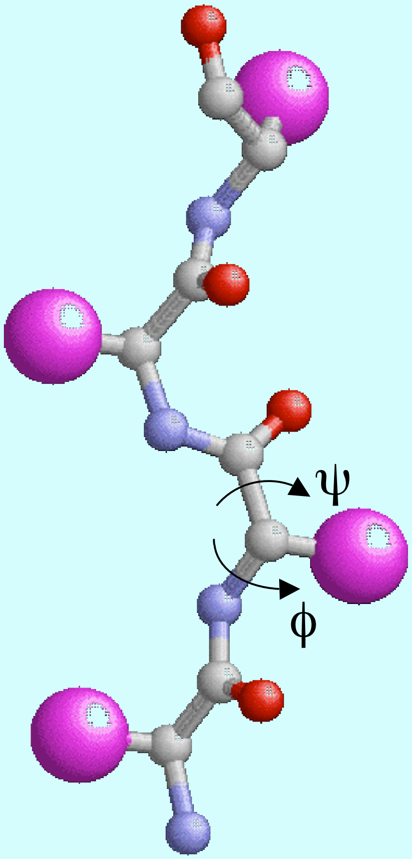
Top Lattice Structures





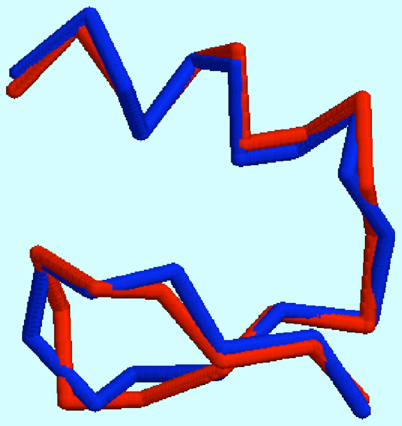
Off-Lattice *m*-State Model

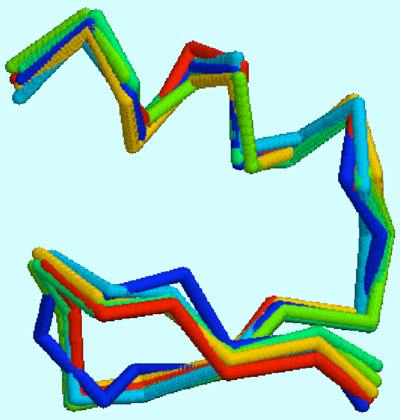
Park & Levitt (1995)



$$(\phi,\psi) = (\phi_1,\psi_1), (\phi_2,\psi_2), \ldots, (\phi_m,\psi_m)$$

Fit to 1PSV By 3-State Model



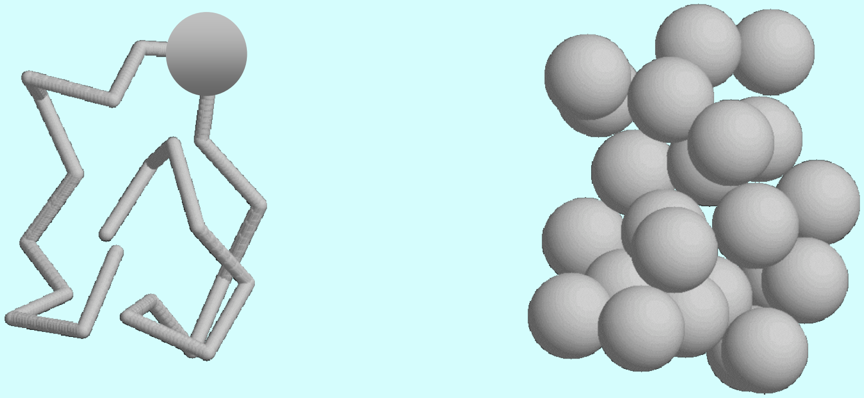


▬ 1PSV

crms ~ 0.8 Å

Enumeration of Configurations

(off-lattice HP model)

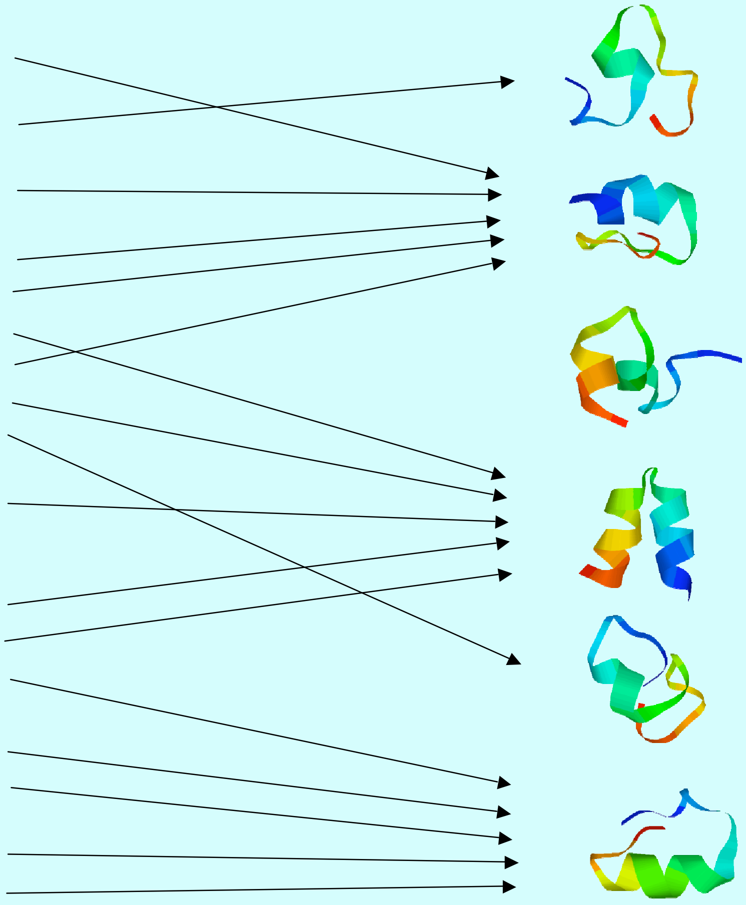


a_i – accessible surface area of residue i

$$A=\Sigma\, a_i < A_c$$

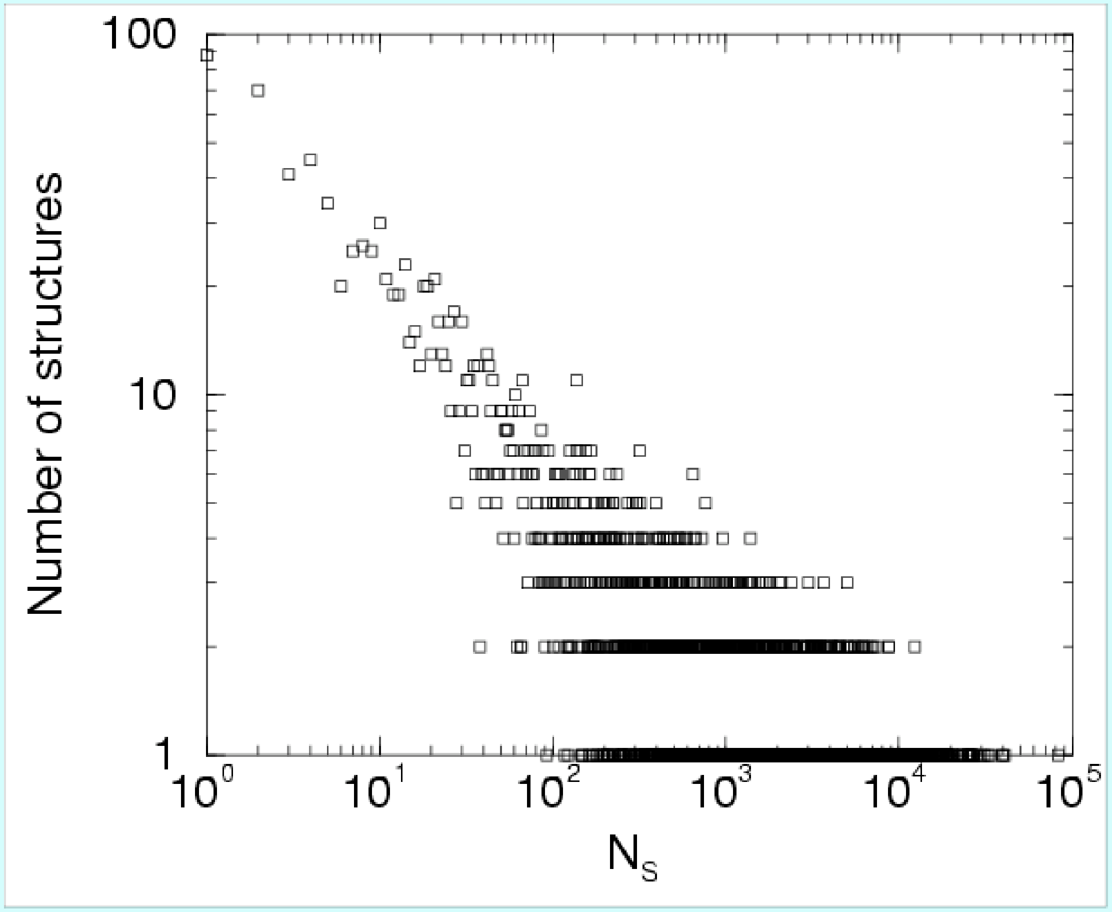
$$\boxed{H = \Sigma\, h_i\, a'_i} \qquad (a'_i=a_i/A)$$

Searching for New Folds



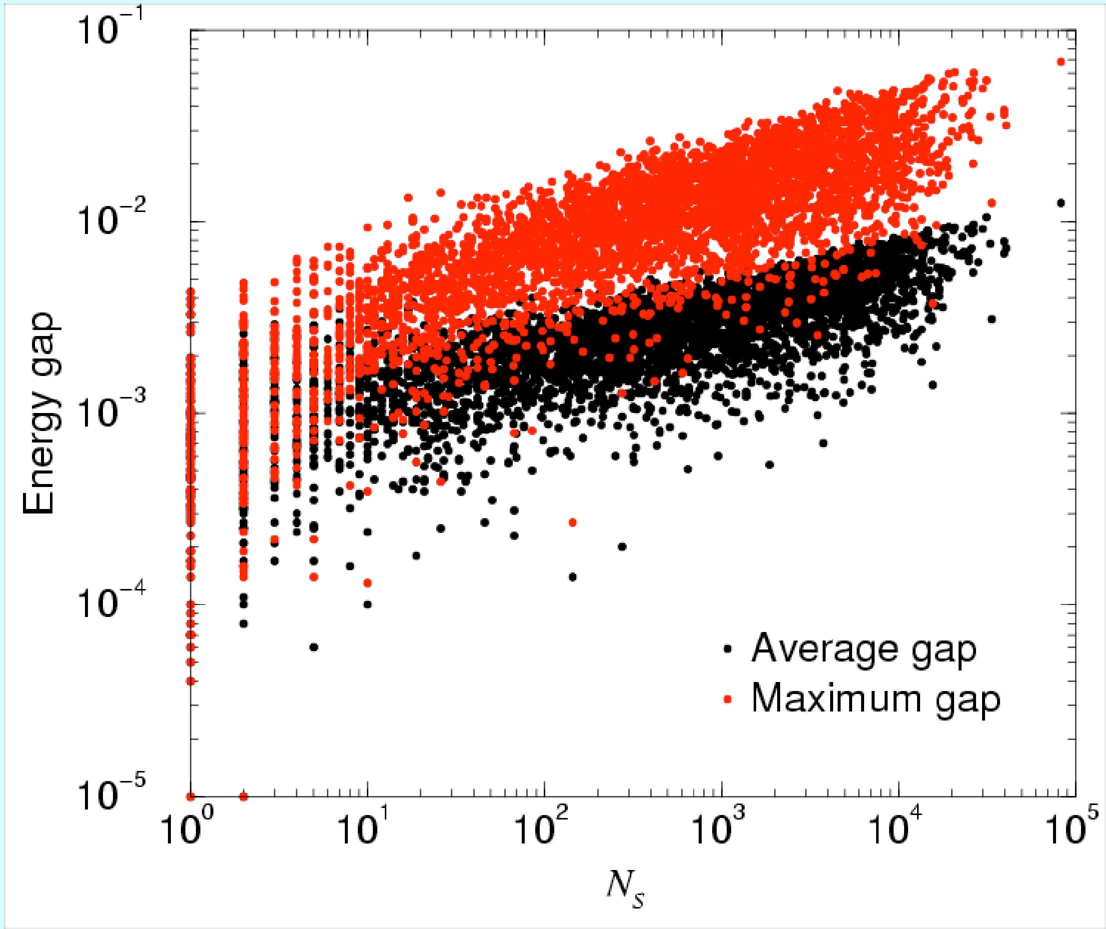
MAQTDVILCPDTHQKAVCLEKIR
EYFDCGLPSAQWCVIKNALLTFR
PNMDKILYVVACQGGARASLTLR
GAWETRKLHCMQPIVYTTLNGLT
MNSAKPLVVTIYSQYNVHLRFDD
GFDSKLACVNPMRTYEIWLETFR
MLKSPQCNYIAVRIHGRYLDFGS
CVNKMYHGFDAALLTRQVLPSLT
QIVLTAGNYIGRGPNIPCLDIGS
EFIINCAQLVRENHWGVSGLRAN
LAGTRVNIMPCDEWSILSLMKIH
FHDISAQVYTERPQMVKRLAFRA
TCNMRWDPSIVYTWQFGHLCVHE
WMTVINEDSAPILCWHGGLMFGN
VVEERPAADIMNWGLRCSLKELT
ILKNETVGGAPQWYIVHNQFNAK
NQKDIETRYPMKSLVSCILHIKM
LMKIHYDTFREWQVNSCKLDDVS
MSKALLVPQWIVRCSYTPLKWPS
TCNMRWDPSIVYTWQFGHLSLVT
CAKVINEDSAPILCTRSGLRLTD
VNYIPPAADIAQREMRCSLTNQL
VDGHETVGGAPQWYFAKRLCRGA
DEYLIETRYPMKSLILSTAEEKL
ATDIHYDAGREWQVNSFELITSV
ETLDLLVPQWIVRCSHTFDIPQS

Histogram of N_s



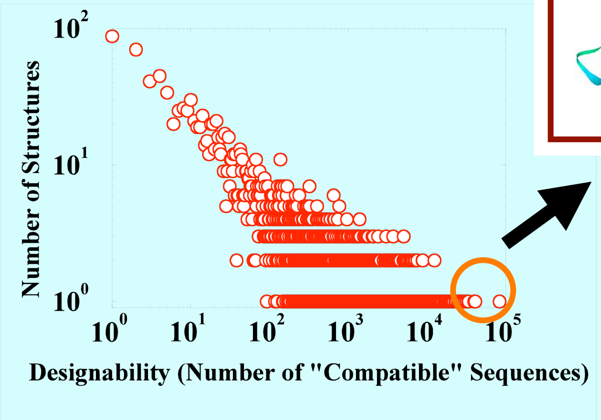
N=23, r_β=1.9 $\overset{o}{A}$, λ=0.4 $\overset{o}{A}$,
10,000 configurations, 4688 clusters

Energy gap vs. N_s

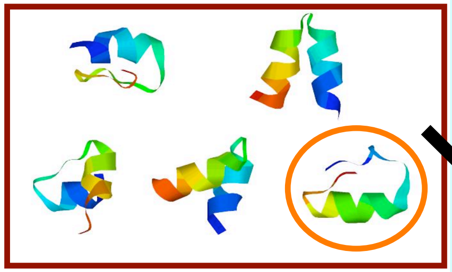


Design Procedure

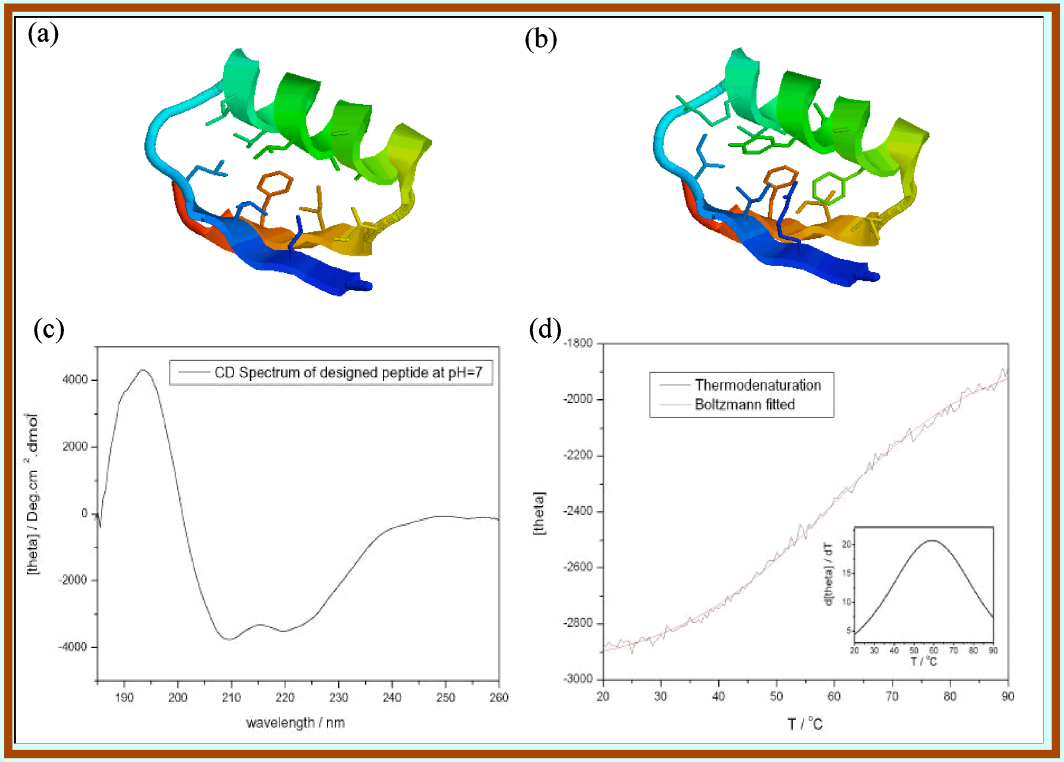
I. Model Computation

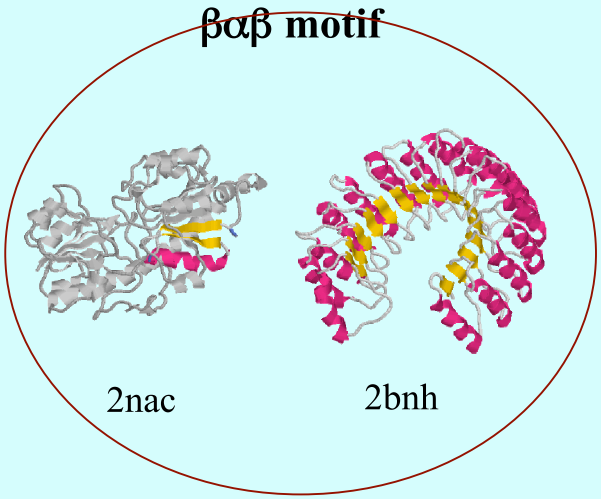


II. Pick Top Folds

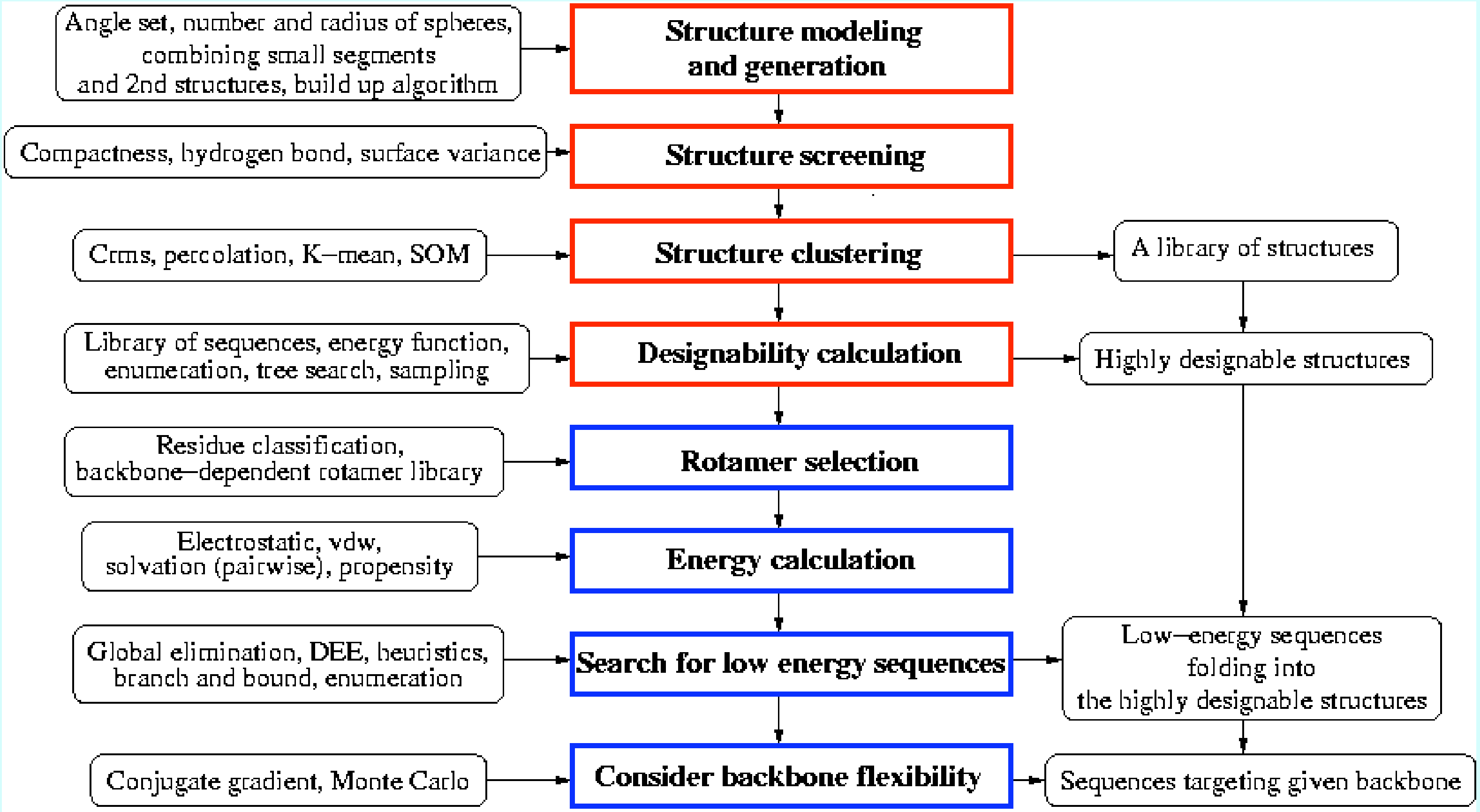


III. Sequence Design and Verification





Design Protocol



Angle set, number and radius of spheres, combining small segments and 2nd structures, build up algorithm

Structure modeling and generation

Compactness, hydrogen bond, surface variance

Structure screening

Crms, percolation, K–mean, SOM

Structure clustering

A library of structures

Library of sequences, energy function, enumeration, tree search, sampling

Designability calculation

Highly designable structures

Residue classification, backbone–dependent rotamer library

Rotamer selection

Electrostatic, vdw, solvation (pairwise), propensity

Energy calculation

Global elimination, DEE, heuristics, branch and bound, enumeration

Search for low energy sequences

Low–energy sequences folding into the highly designable structures

Conjugate gradient, Monte Carlo

Consider backbone flexibility

Sequences targeting given backbone