Predicting Protein Structures through Optimization

Prashanta Shrestha
Jose Mojica
Reza Asgharzadeh
Josh Price
Brigham Young University

22 May 2012

Overview

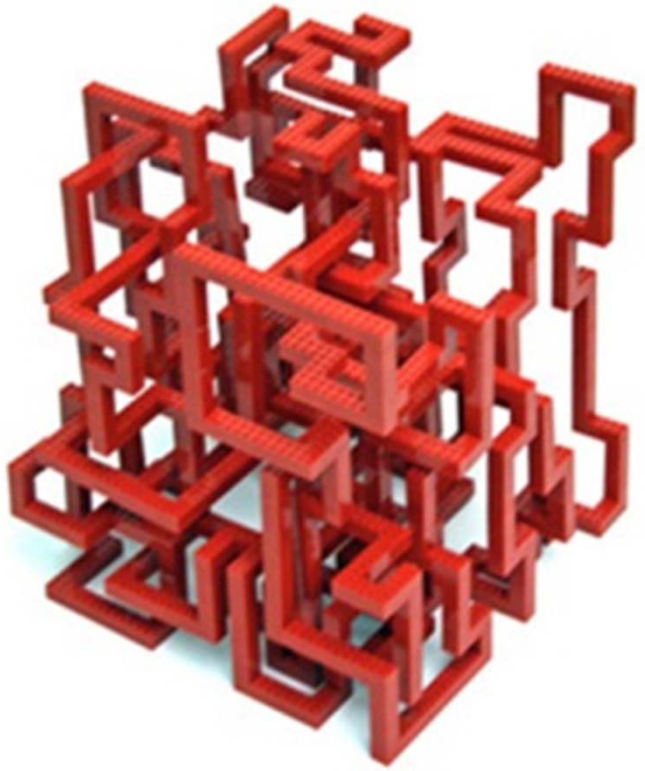
- Why predict protein structures?
- Current methods for protein folding simulation
- Case study with coarse-grain optimization
  - Genetic Algorithm
  - Global Optimization
  - Using Uncertainty Descriptions
- Experimental verification of protein structures
- Conclusions

Introduction to Protein Folding

What is an amino acid and how do they form proteins?

http://upload.wikimedia.org/wikipedia/commons/3/32/Lego_Color_Bricks.jpg

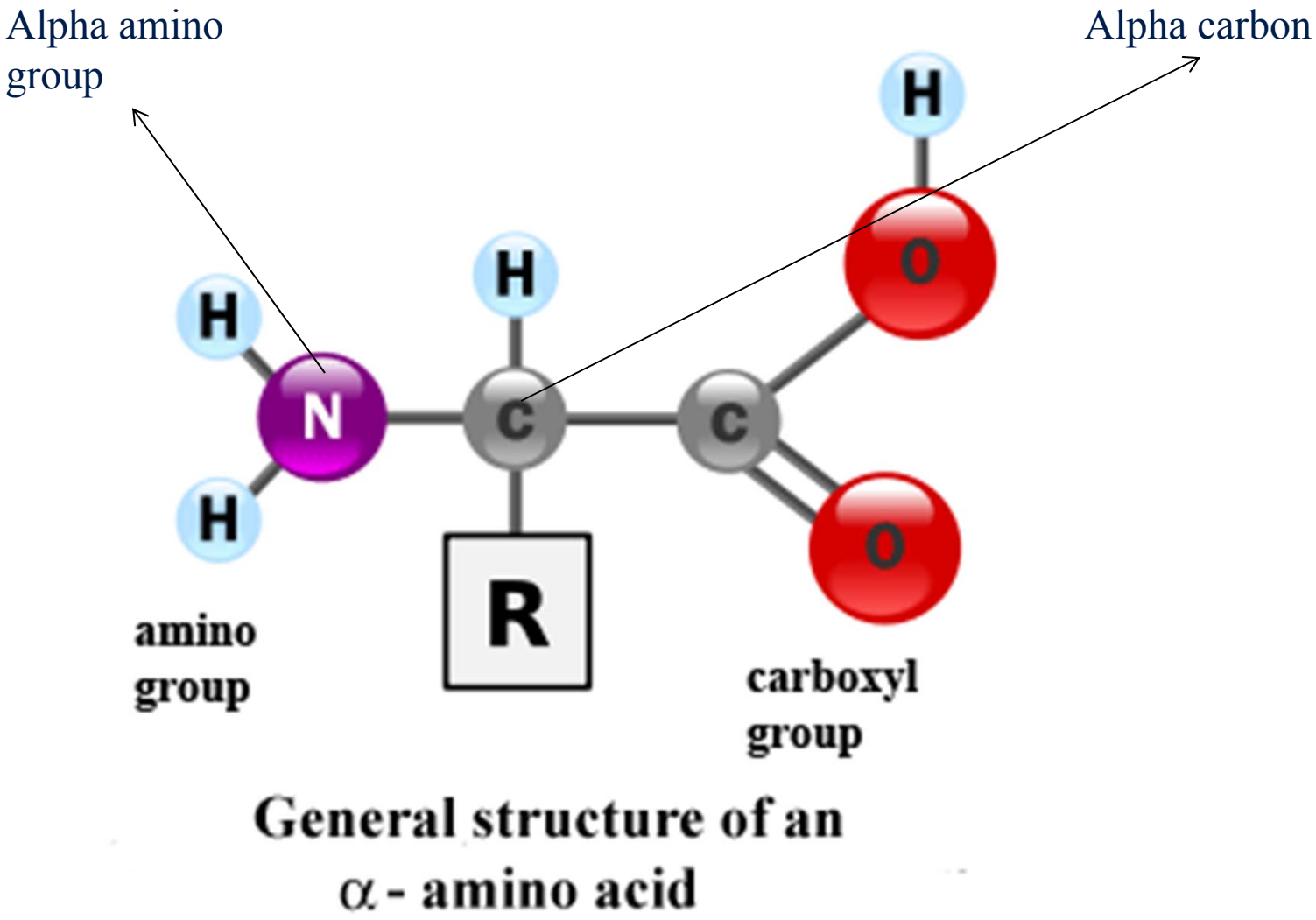
Introduction to Protein Folding



http://s3files.core77.com/blog/images/legodude.jpg

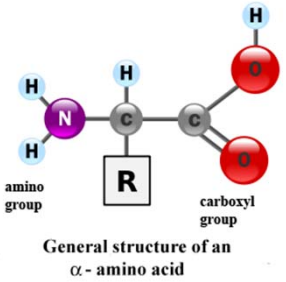
http://www.fandompost.com/wp-content/uploads/2012/02/Lego-Avengers-Header.jpg

Introduction to Protein Folding



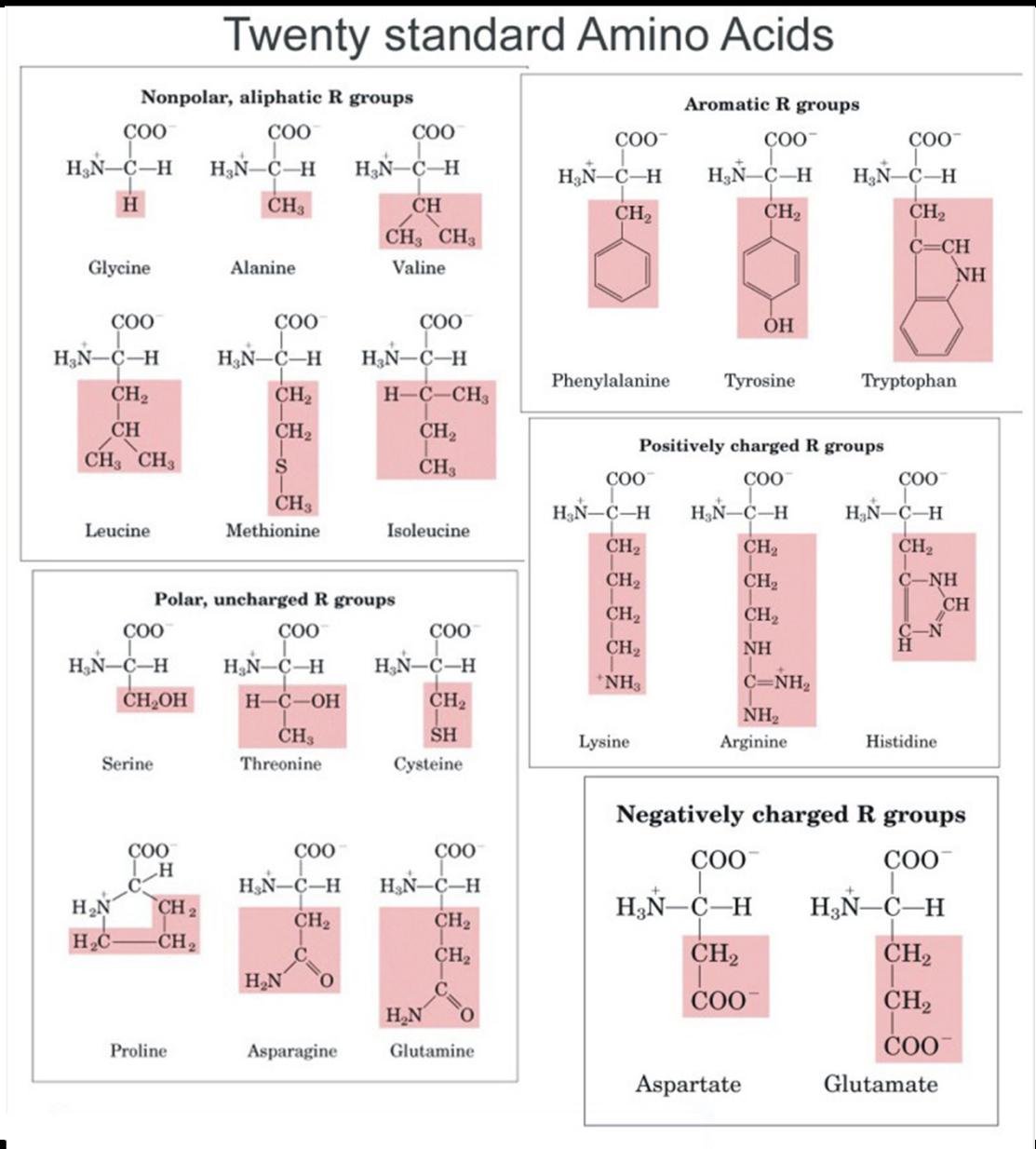
http://homepages.ius.edu/dspurloc/c122/images/aminostrc.gif

Introduction to Protein Folding



General structure of an α - amino acid

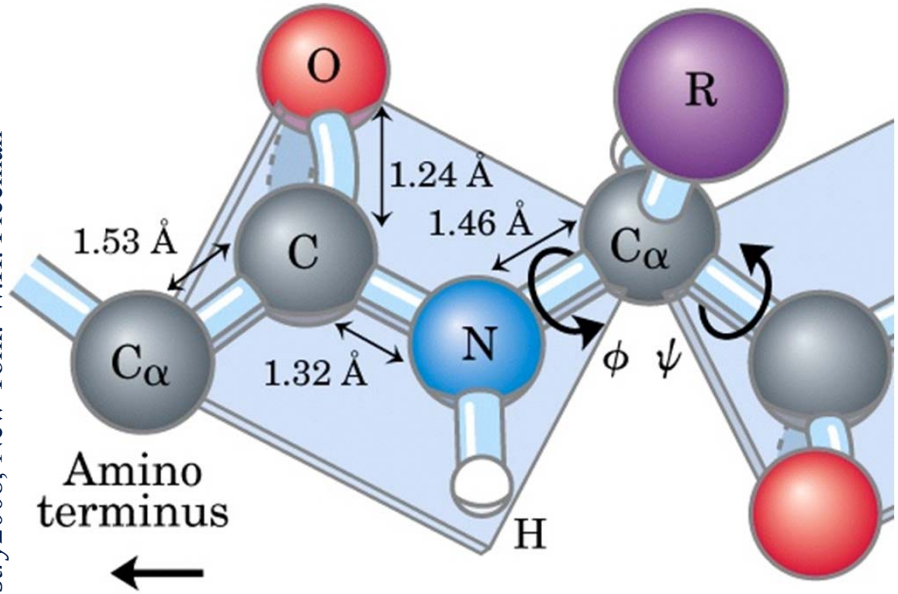
http://homepages.ius.edu/dspurloc/c122/images/aminostrc.gif



http://www.geneinfinity.org/images/aminoacids.jpg

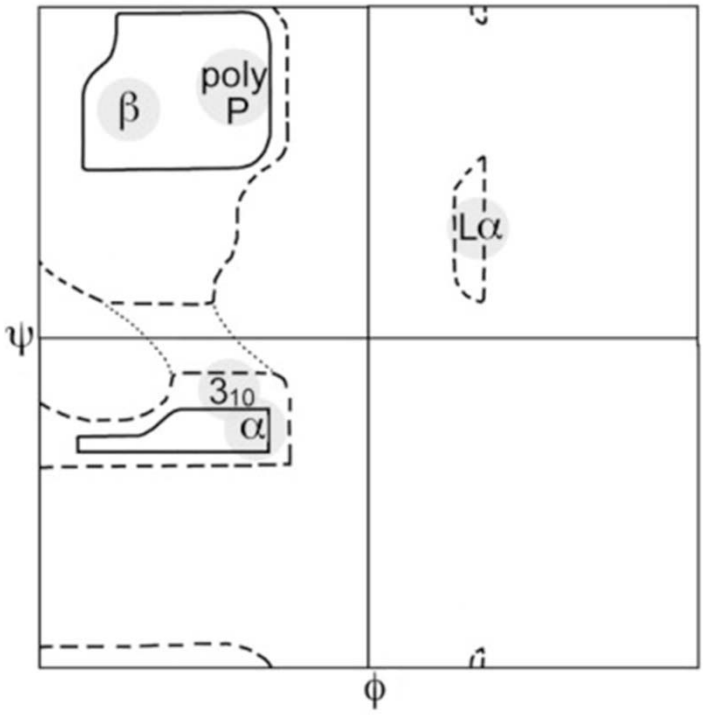
Introduction to Protein Folding

Φ and ψ angles



Nelson, D.L., A.L. Lehninger, and M.M. Cox, *Lehninger principles of biochemistry*2008, New York: W.H. Freeman

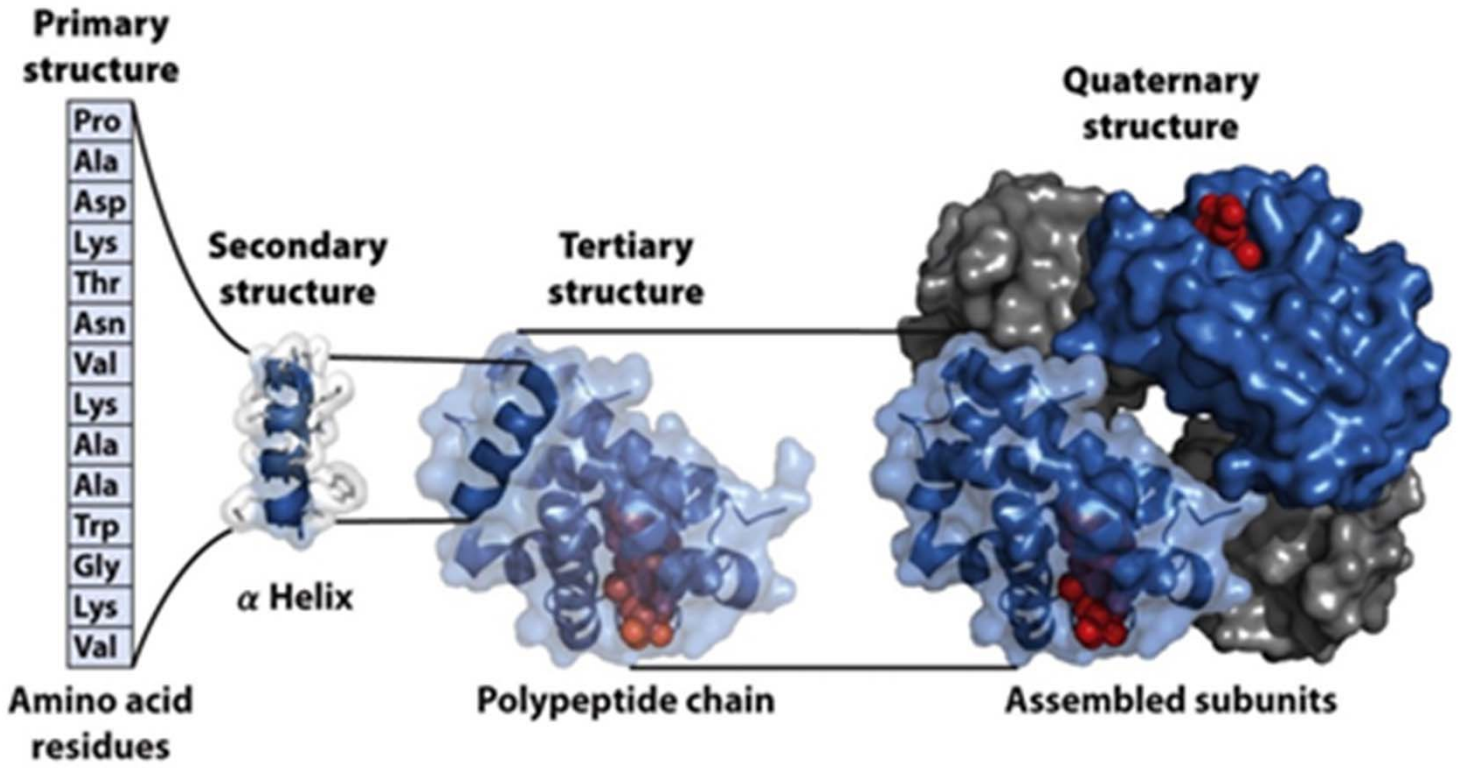
Ramachandran plot



http://upload.wikimedia.org/wikipedia/commons/8/85/Ramachandran_plot_original_outlines.jpg

Why Predict Protein Structures?

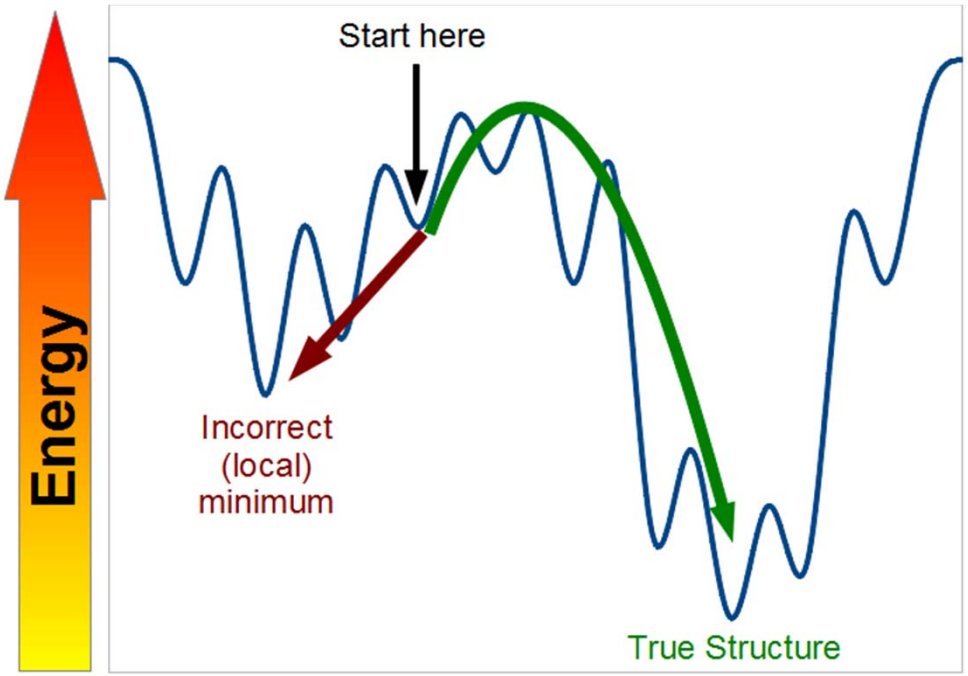
What gives a protein it's structure, and how do they do it so efficiently?



Nelson, D.L., A.L. Lehninger, and M.M. Cox, *Lehninger principles of biochemistry*2008, New York: W.H. Freeman

Genetic Algorithm for Structure Prediction

- GA's can adapt to the environment in an efficient manner.
- Maintain a population of solutions.
- GA increases the probability of getting to energy minima and finding global optima.



http://conflux.mwclarkson.com/wp-content/uploads/2011/09/eldscp2.png

A genetic algorithm will search for the design that will give the lowest energy structure

Models- 2D & 3D lattice structures

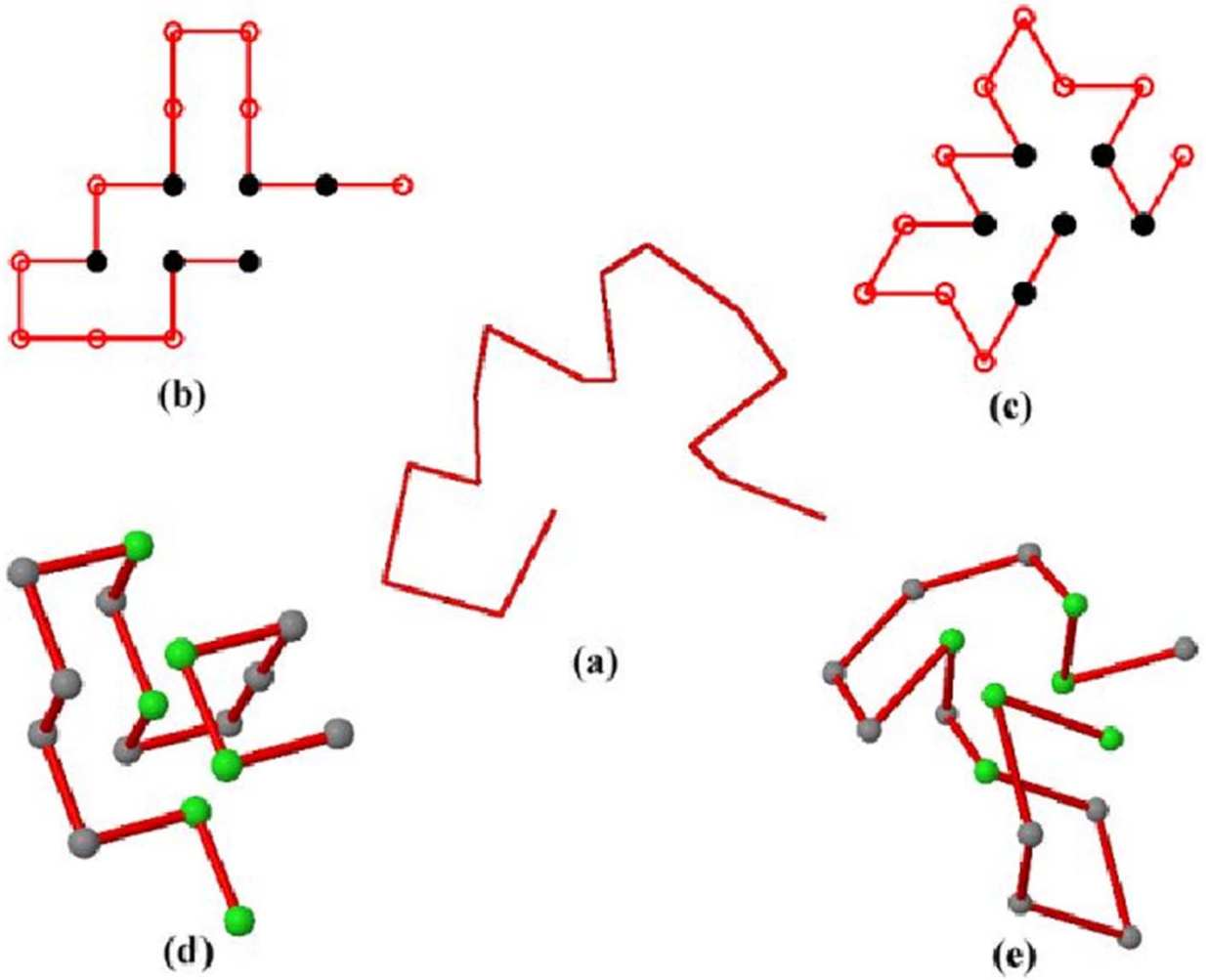
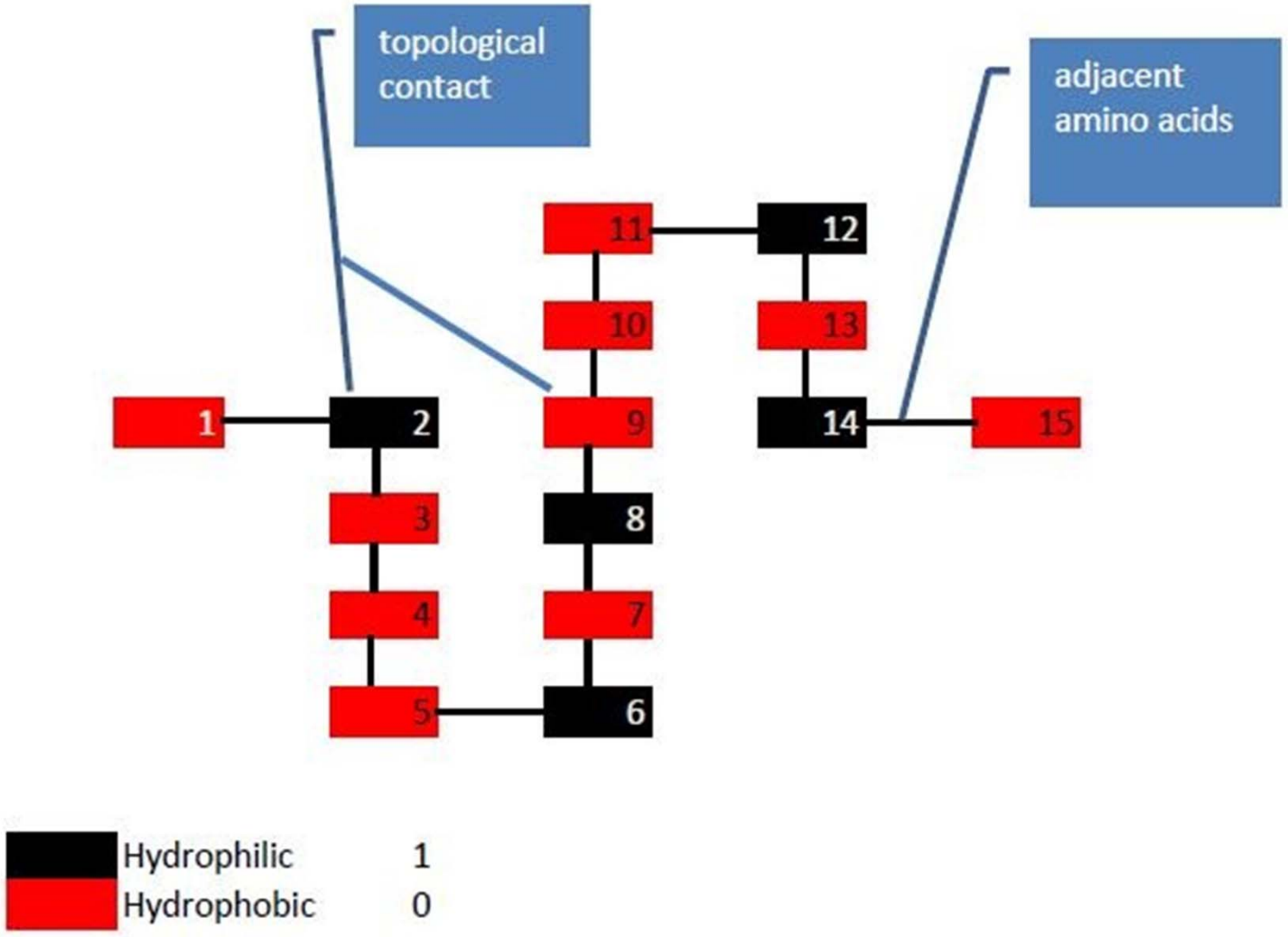


Figure 1 Four different types of lattice model for visual comparison taking the protein with the PDB id: 1A0Ma as the example. Visual comparison for PDB id: 1A0Ma. (a) Real protein structure; (b) and (c) are 2D square and triangular lattice model simulation results. Black-filled dots indicate Hydrophobic amino acids and white dots denote hydrophilic amino acid. (d) and (e) are 3D square and face-centered cubic (FCC) lattice model simulation results from CPSP-tools [3]. In (d) and (e), green balls indicate hydrophobic amino acids while the gray balls indicate the hydrophilic amino acids.

Model: 2D Lattice, Amino Acid Sequence



Design Equations

- Fitness and Objective:

minimize Total energy Potential (E)

- Design variables:
  - Shape
  - Amino acid sequence
- s.t:
  - self-avoiding path (no self-crossing and not occupying a lattice space more than once)
  - right angle vertices only
  - adjacent amino acids cannot be separated by more than one lattice space

Energy potential for each amino acid *i* in protein chain

$$E_{p,i} = -\frac{\sum_j^3 \mid H_j - H_i \mid}{N_H}$$

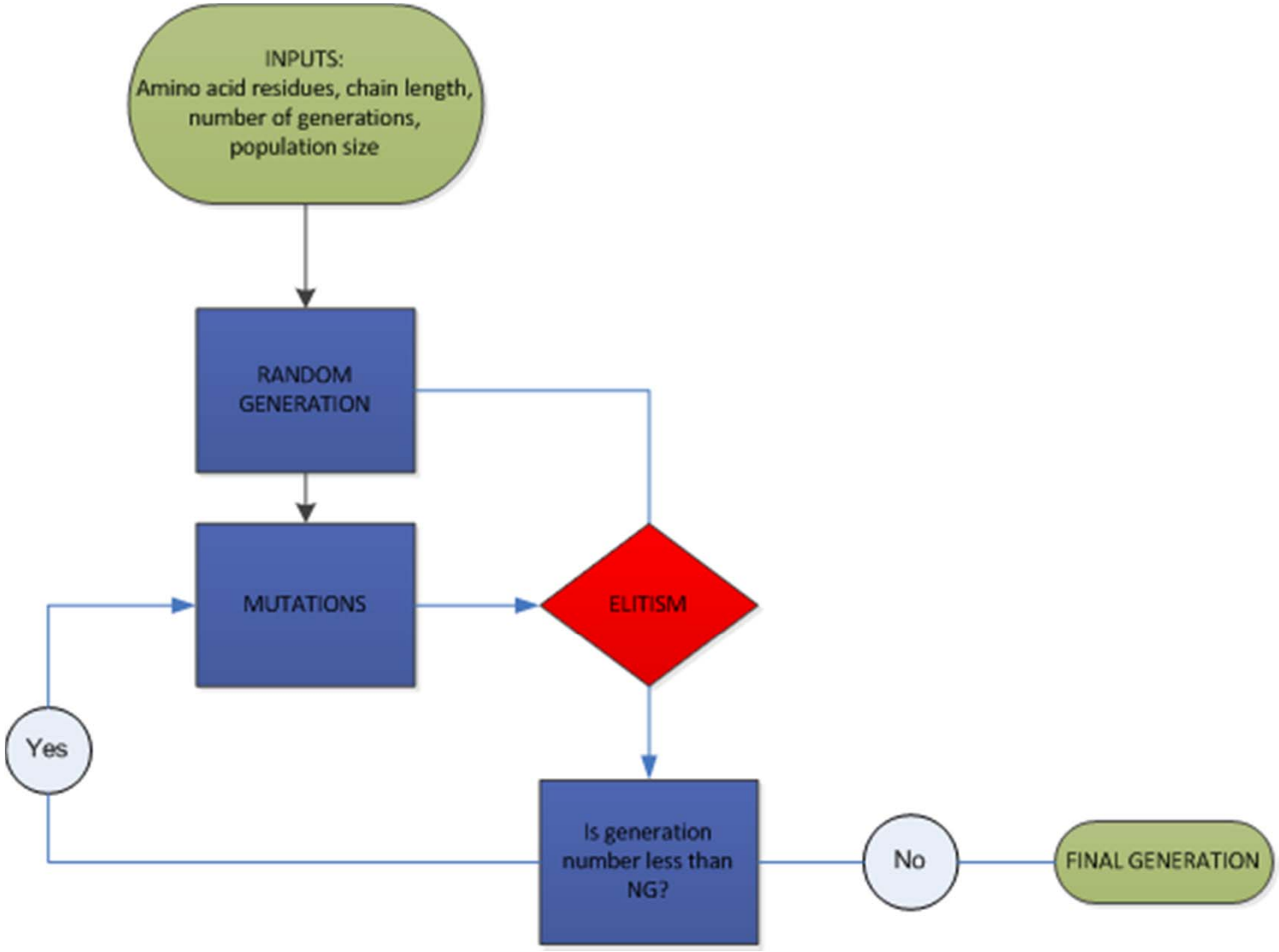
Penalty function for each amino acid That contains a hydrophobic-hydrophobic Topological contact

$$P_k = \sum_k k * (2 \times N_H + 2)$$

Total energy: objective function

$$\min \quad E = \sum_i E_{p,i} + \sum_k P_k$$

Genetic Algorithm

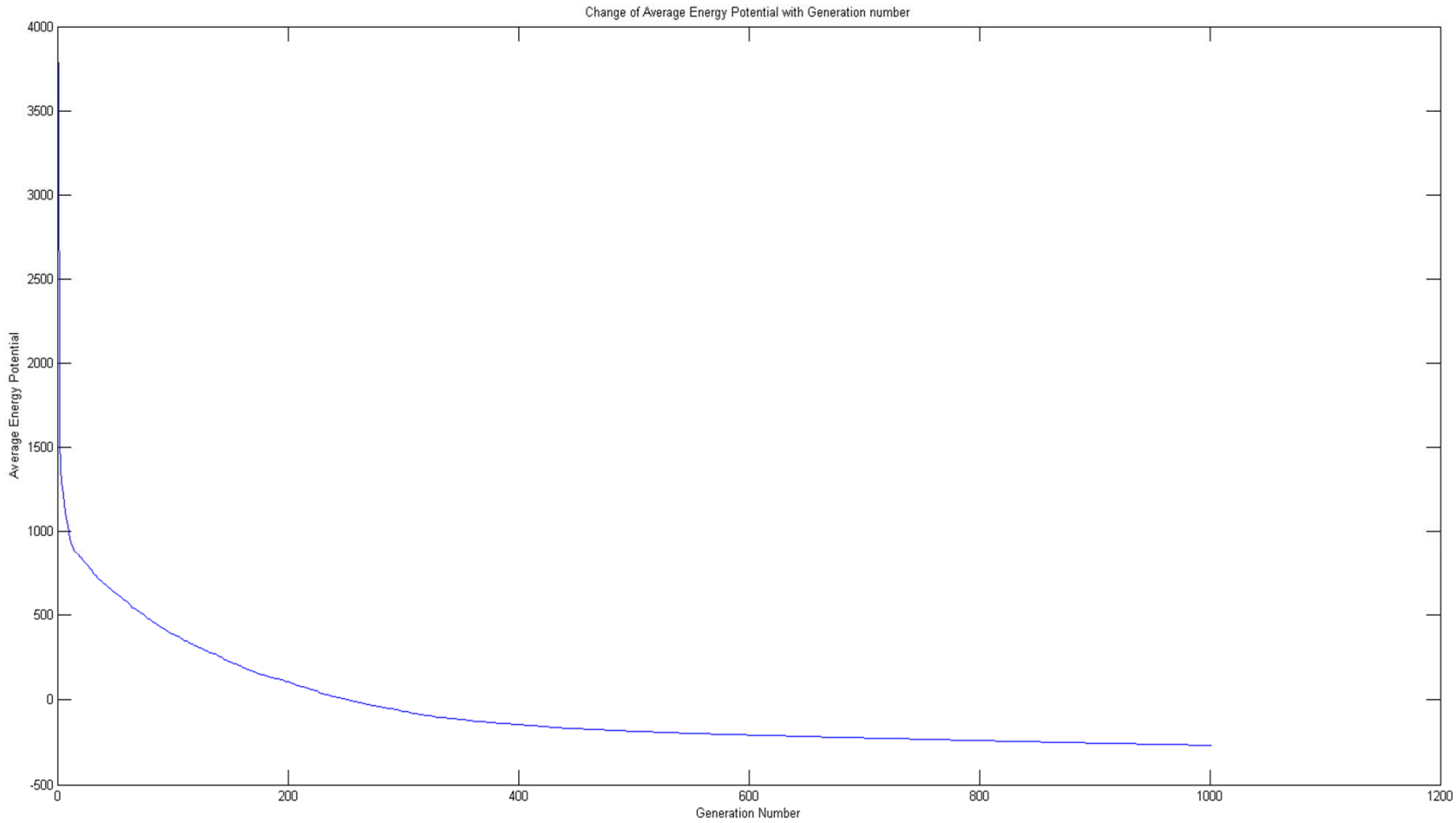


Optimization Procedure

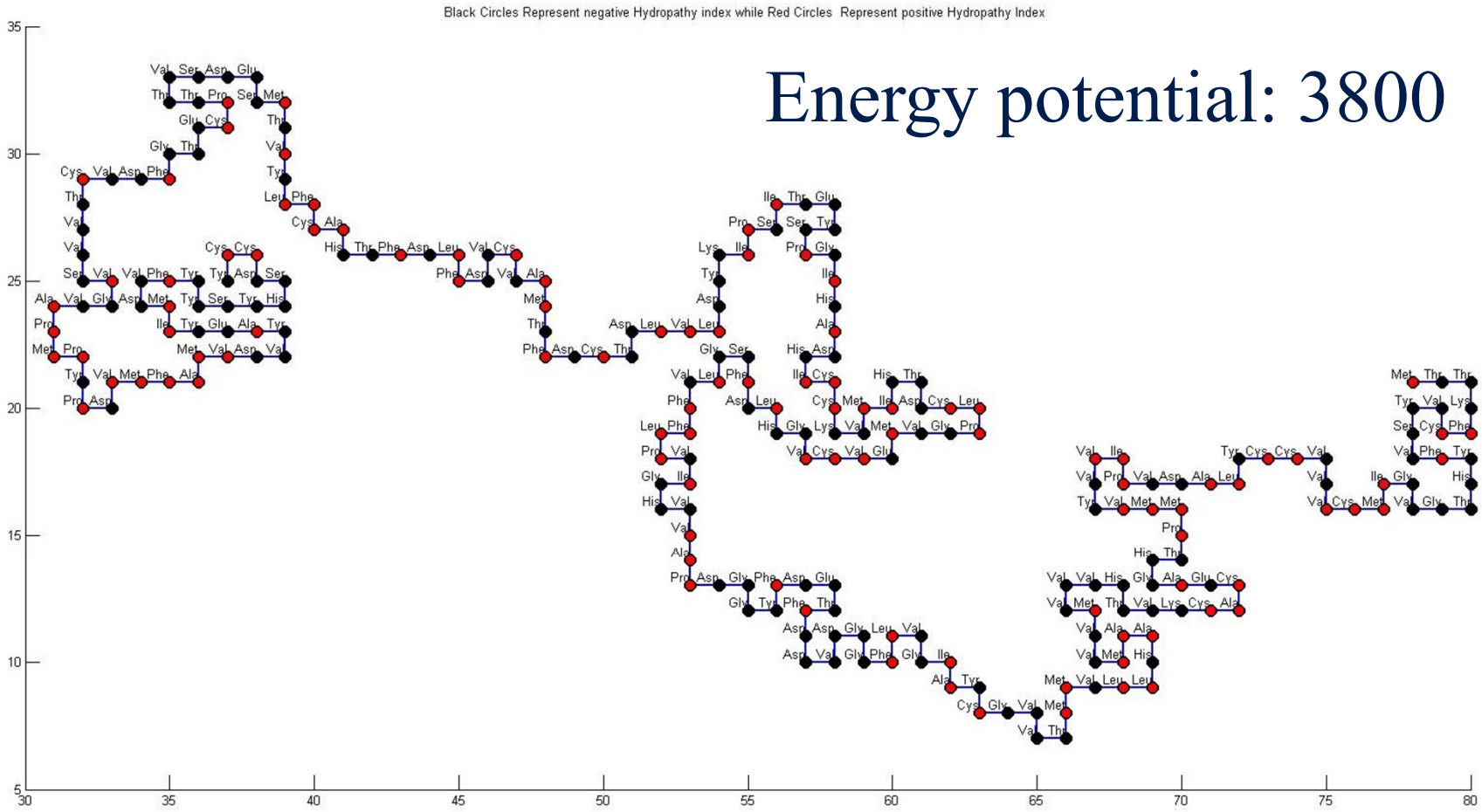
- 20 types of amino acid
- 238 amino acids in protein chain
- Population size: 250
- Generation number: 1000
- Algorithm decides their position in lattice

Amino acid	Hydropathy
Glycine	-0.4
Alanine	1.8
Proline	1.6
Valine	4.2
Leucine	3.8
Isoleucine	4.5
Methionine	1.9
Phenylalanine	2.8
Tyrosine	-1.3
Tryptophan	-0.9
Serine	-0.8
Threonine	-0.7
Cysteine	2.5
Asparagine	-3.5
Glutamine	-3.5
Lysine	-3.9
Histidine	-3.2
Arginine	-4.5
Aspartate	-3.5
Glutamate	-3.5

Results – Energy minimization

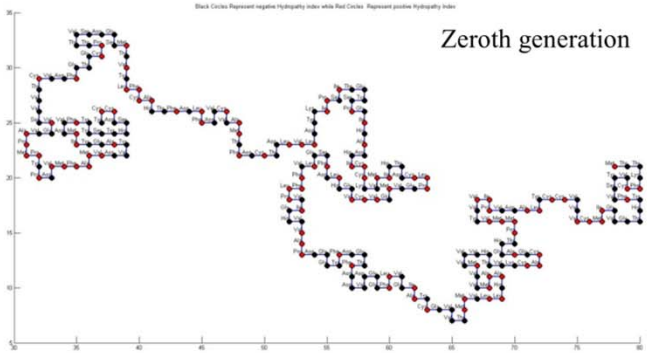


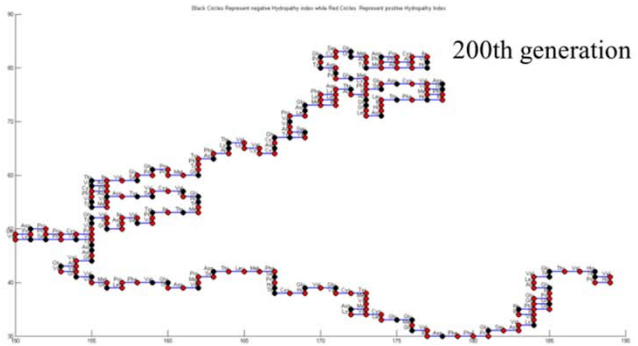
2D structure, 1st random child

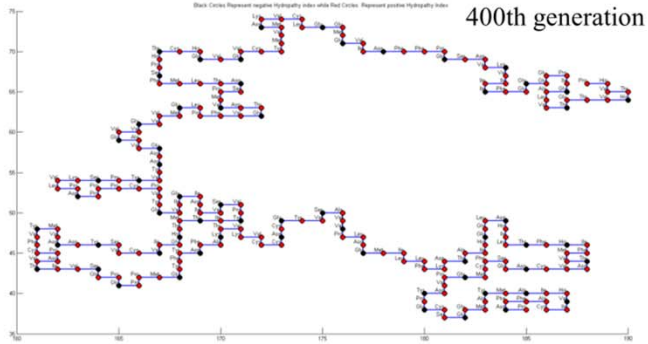


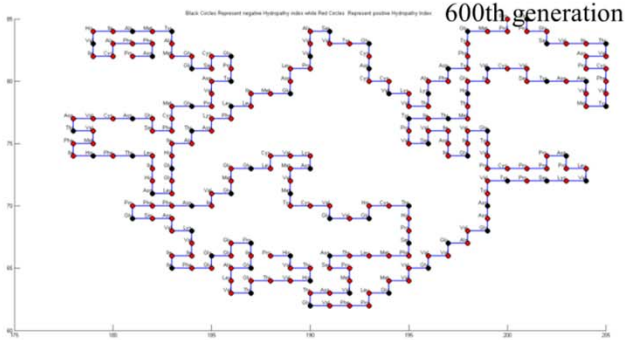
Energy potential: 3800

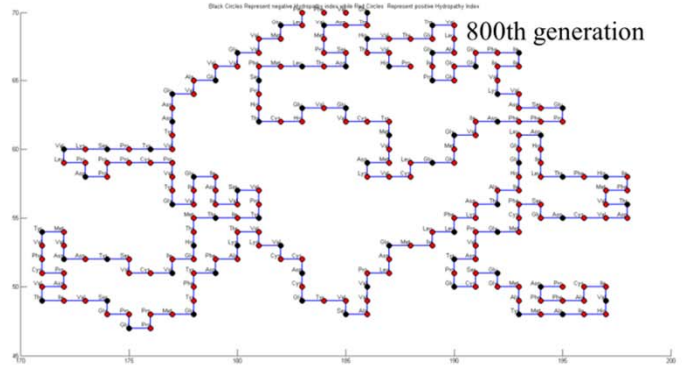
Iterations on 2D Structure

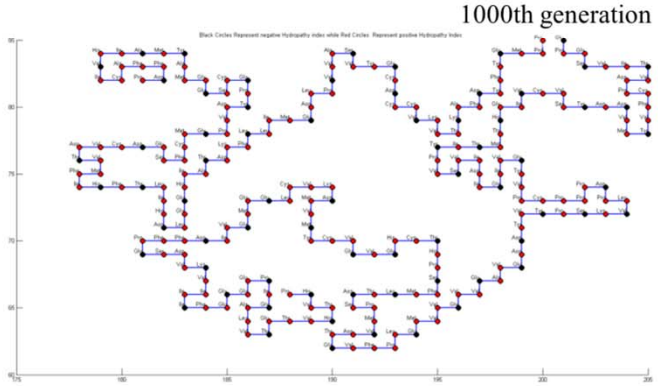






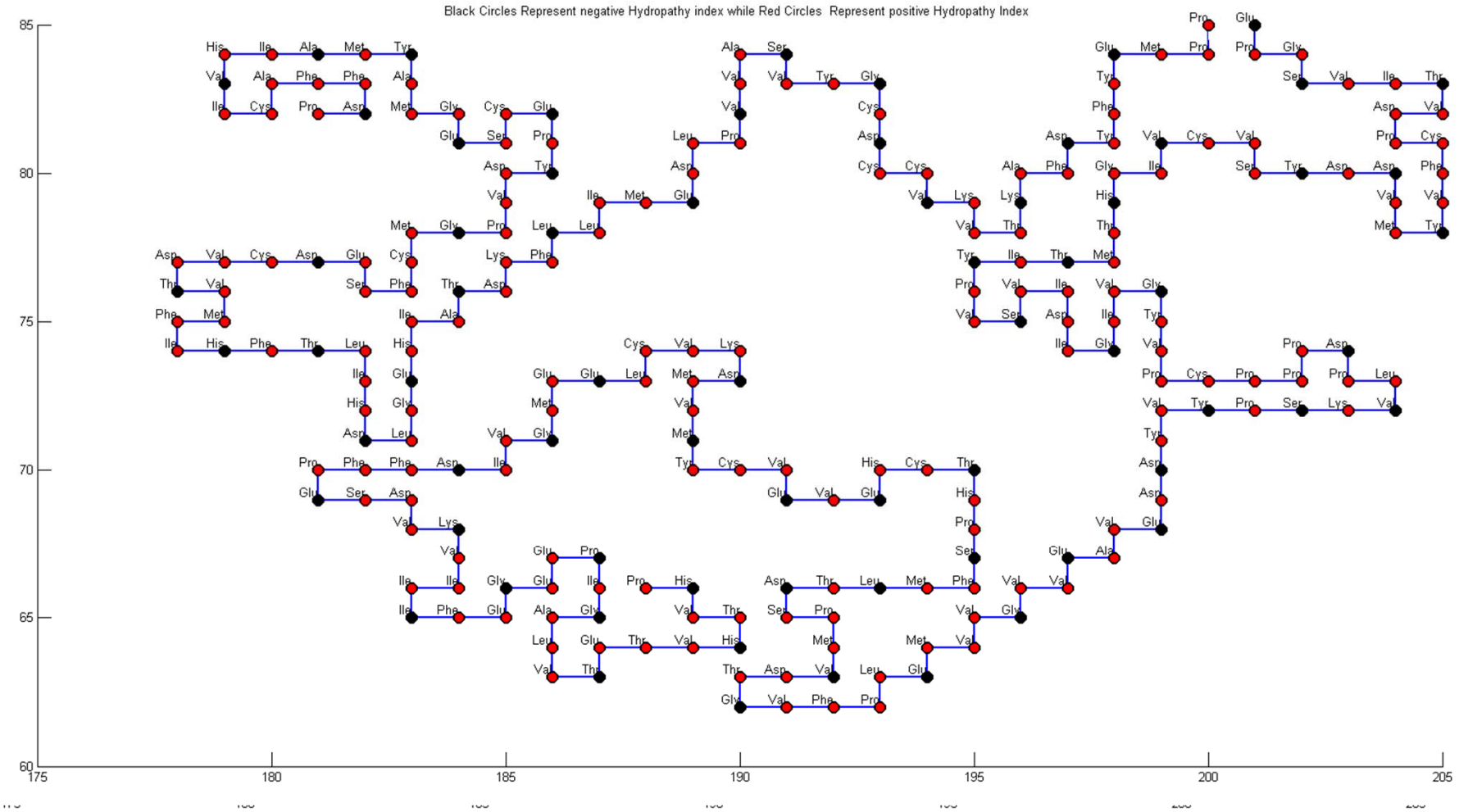




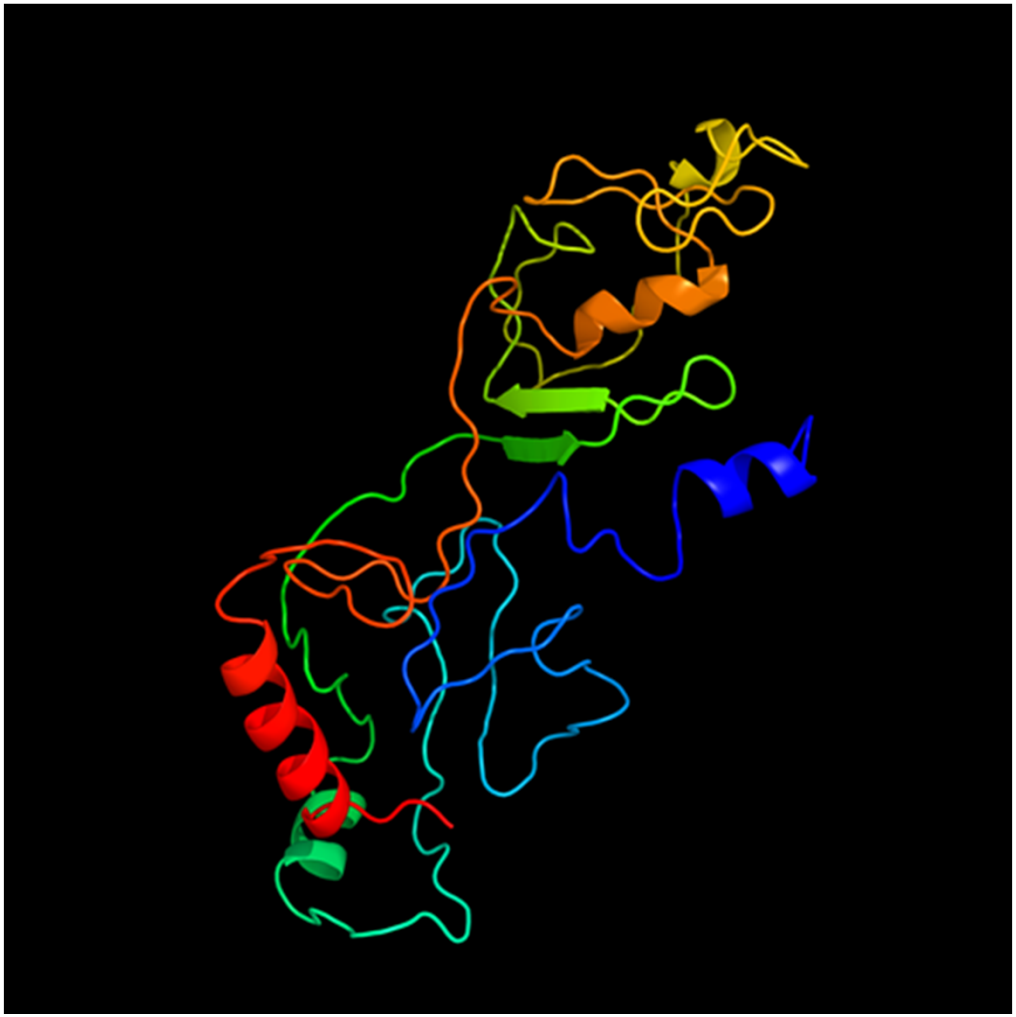


Results: 2D Structure, 1000th Generation

Energy potential: -267



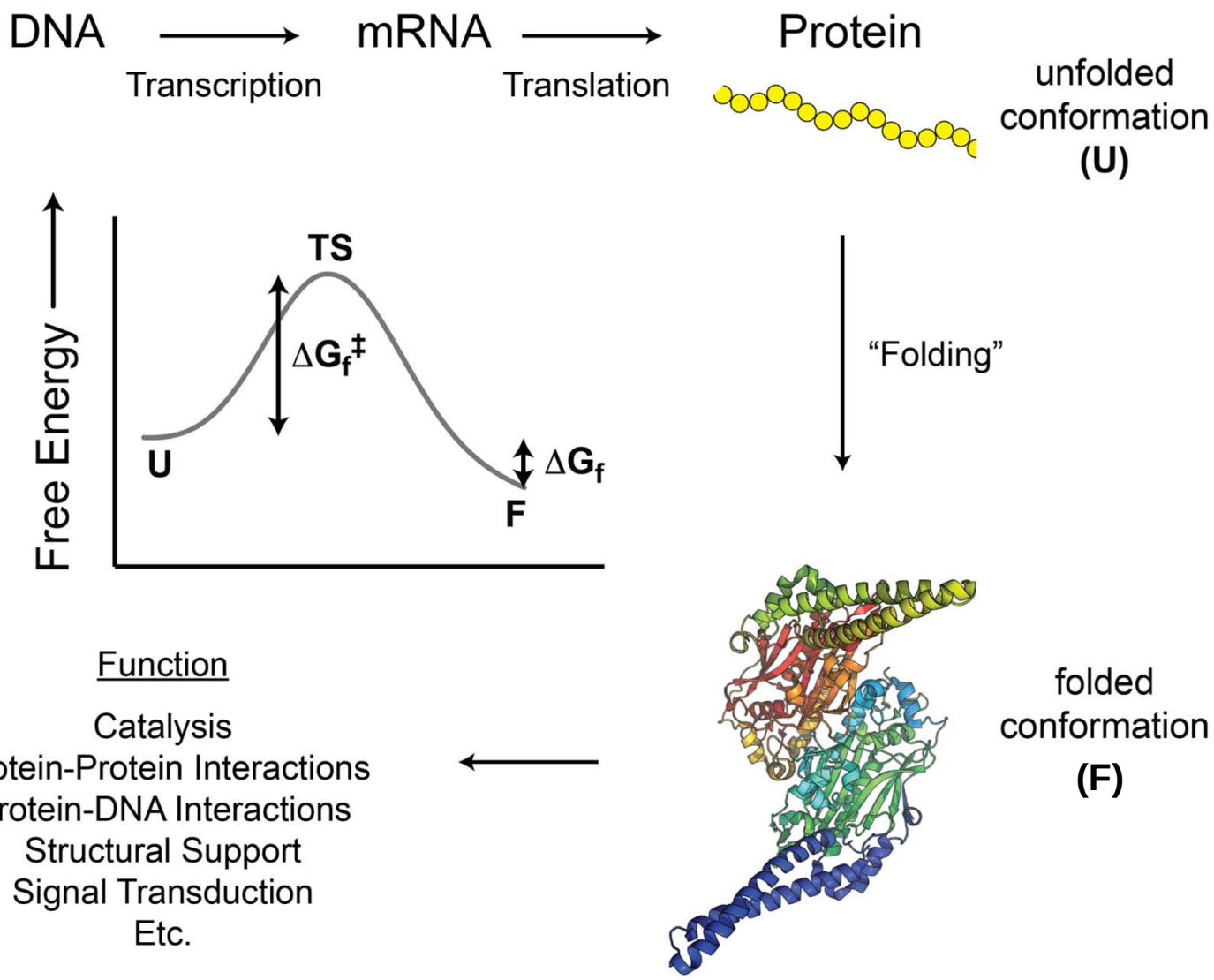
Results: 3 D structure 1000th generation



Kelley, L.A. and M.J.E. Sternberg, *Protein structure prediction on the Web: a case study using the Phyre server.* Nat. Protocols, 2009

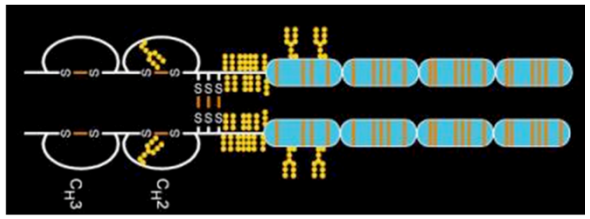
- 4 alpha helices.
- 2 beta sheets.
- The sequence shows 34% similarity to centaurin-alpha-1.

Protein Folding



Protein Stability: Who Cares?

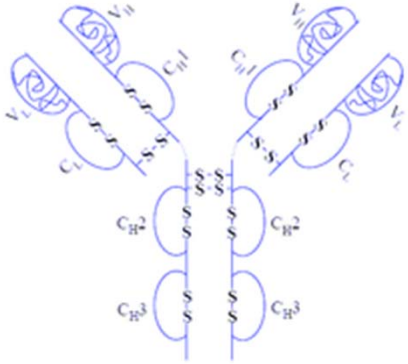
Many of the recent advances in cancer and autoimmune disease therapy have come from biotechnology with the development of protein therapeutics



Etanercept

A fusion of TNFa and the Fc domain of IgG

Autoimmune disease



Monoclonal antibodies

Infliximab
Adalimumab

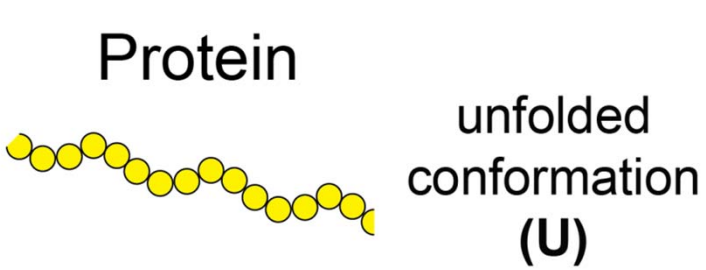
Autoimmune diseases

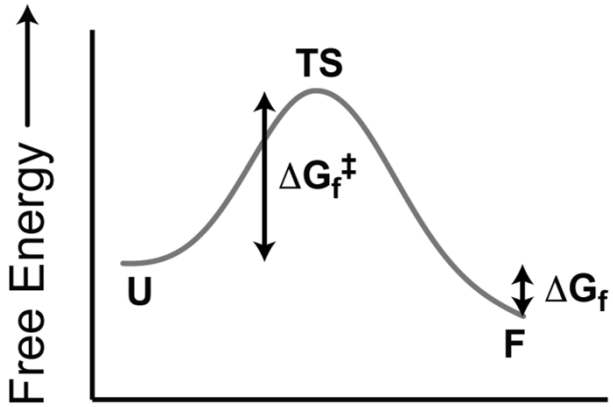
Bevacizumab
Rituximab
Trastuzumab

Cancer

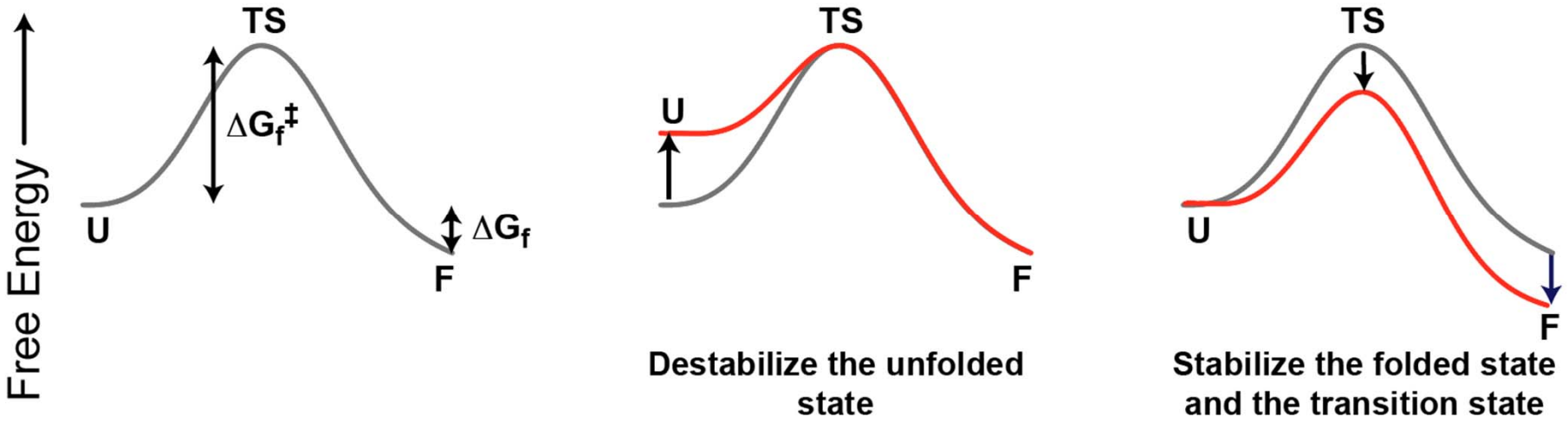
Development of Protein Drugs Is Challenging!

- Misfolding and aggregation problems
- Degradation by proteases in the body
- Rapid clearance by kidney filtration
- Immune Responses
- All of these problems relate to protein stability



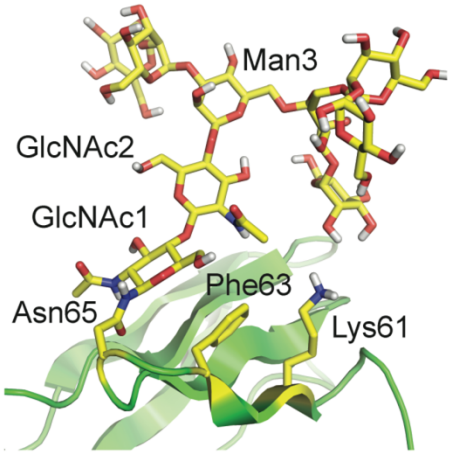


General Ways to Increase Protein Stability?

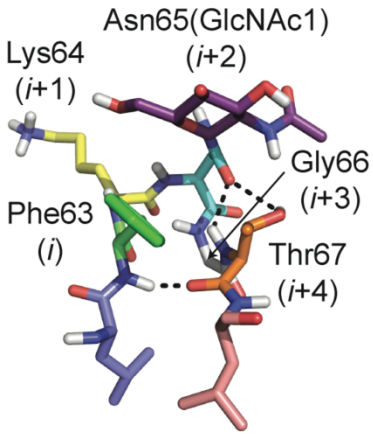


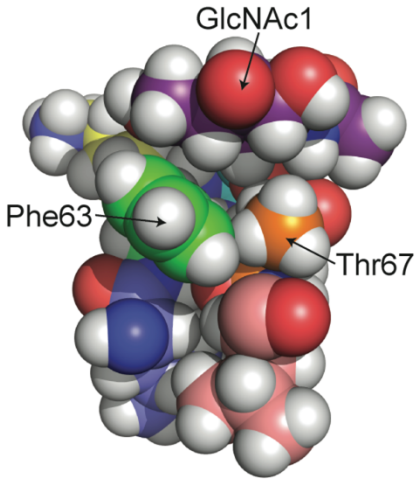
Can we develop general strategies for increasing protein thermodynamic stability and accelerating protein folding rate?

Glycosylation and Increased Protein Stability



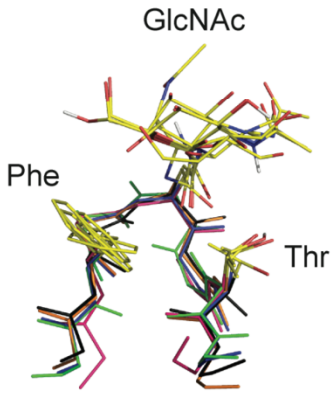
A single N-glycan stabilizes HsCD2ad by -3.1 kcal/mol





A substantial amount of this stabilization comes from an interaction between Phe, the first GlcNAc of the N-glycan, and likely Thr, in the context of a type I β-bulge turn

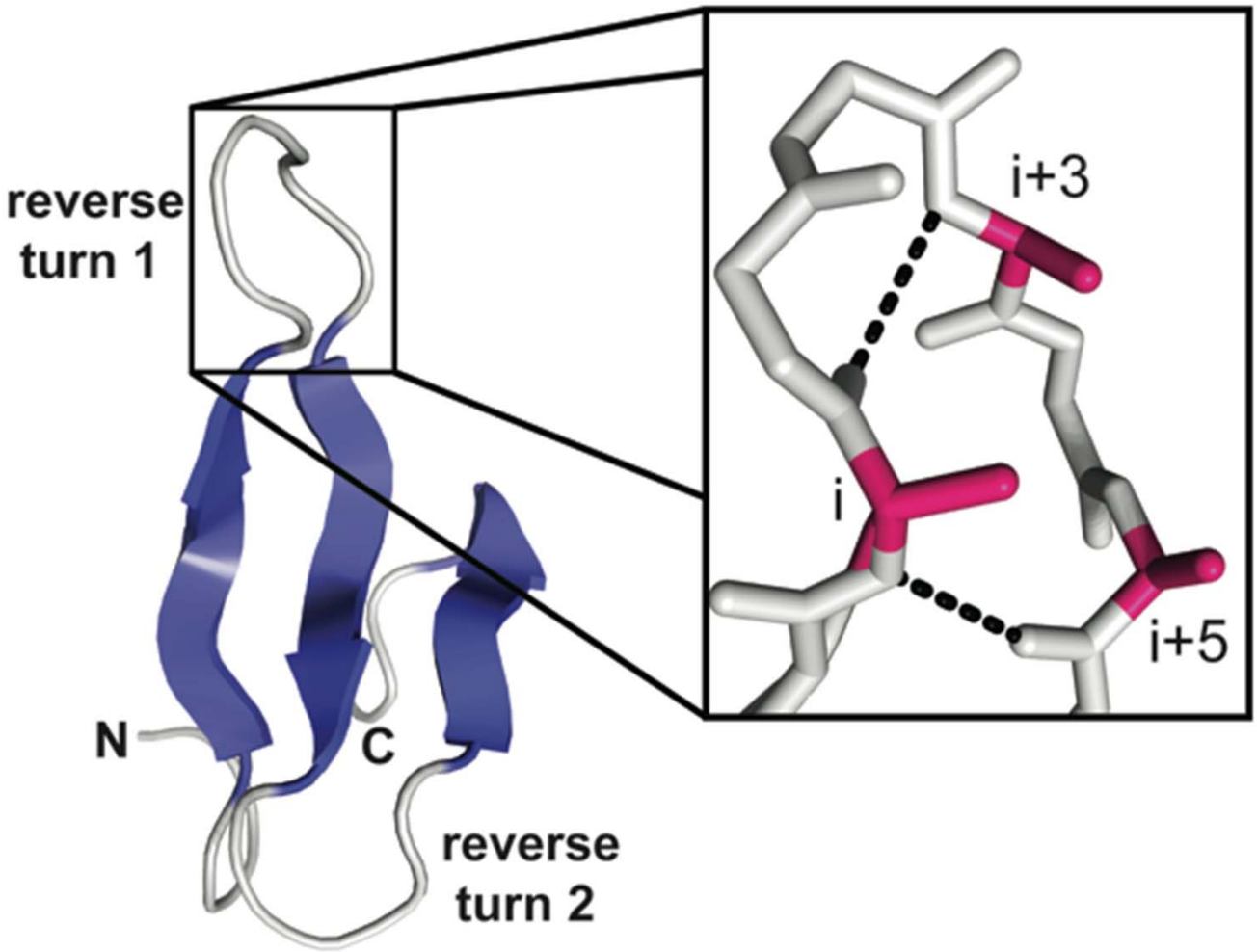
Similar motifs, which we call "enhanced aromatic sequons" are present in other proteins, suggesting that the stabilizing effect observed in HsCD2ad might be portable



But, is this effect generally applicable?

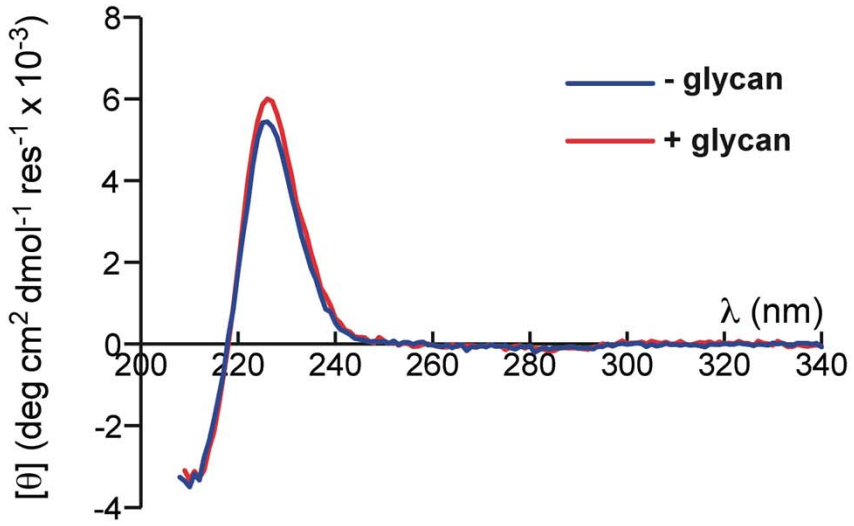
Wyss, D.F., et al. *Science* **1995**, *269*, 1273–1278
Culyba, E.K., et al. *Science* **2011**, *331*, 571–575

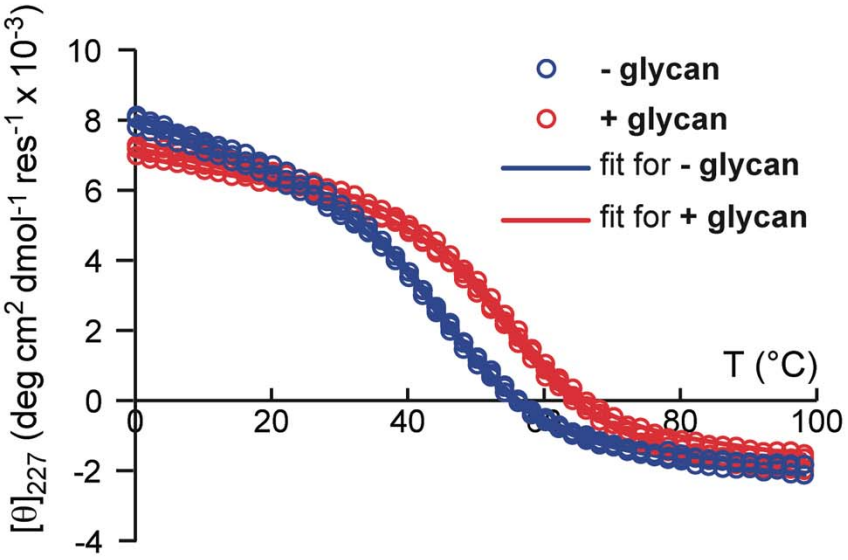
Model System: The WW domain of Pin1



Ranganathan, R.; Lu, K.P.; Hunter, T.; Noel, J.P. *Cell* **1997**, *89*, 875-886

Characterizing the Folding Energetics of WW





Circular dichroism (CD) spectra provide a qualitative indicator for the secondary structure of the protein of interest

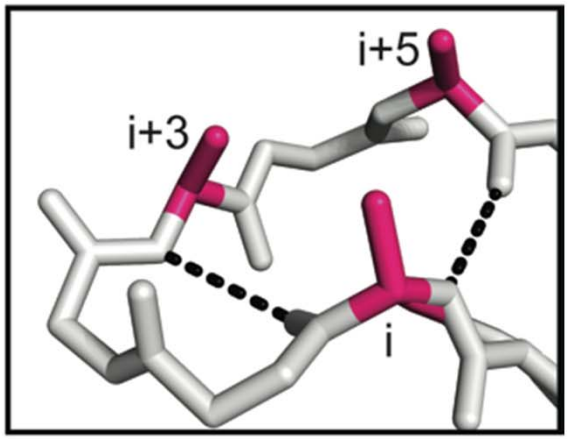
The spectra shown here are characteristic of a protein composed of β-sheets

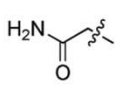
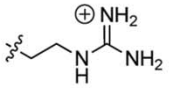
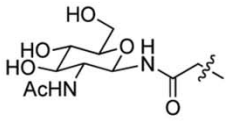
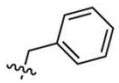
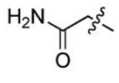
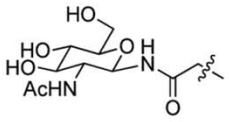
Variable temperature CD data provide quantitative information about the thermodynamic stability of a protein's folded state, relative to its unfolded state.

Price, J.L., et al. *J. Am. Chem. Soc.* **2010**, *132*, 15239–15367

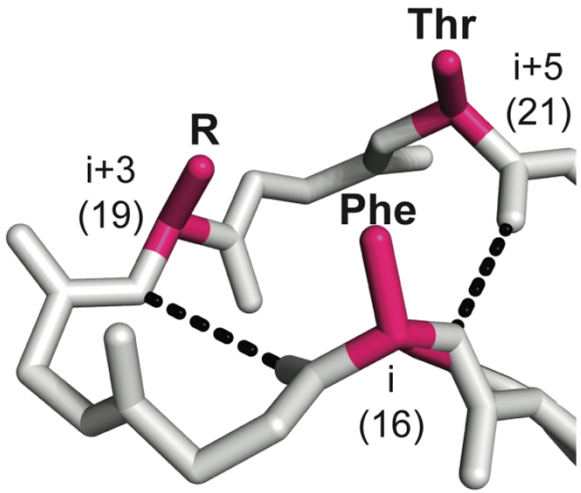
Culyba, E.K., et al. *Science* **2011**, *331*, 571–575

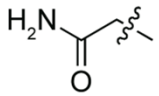
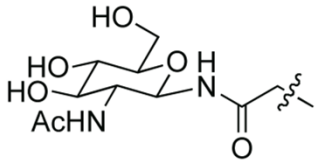
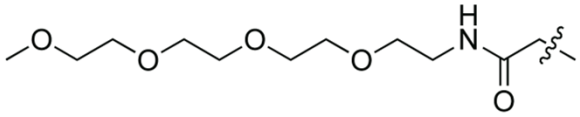
Can Glycosylation Stabilize WW?



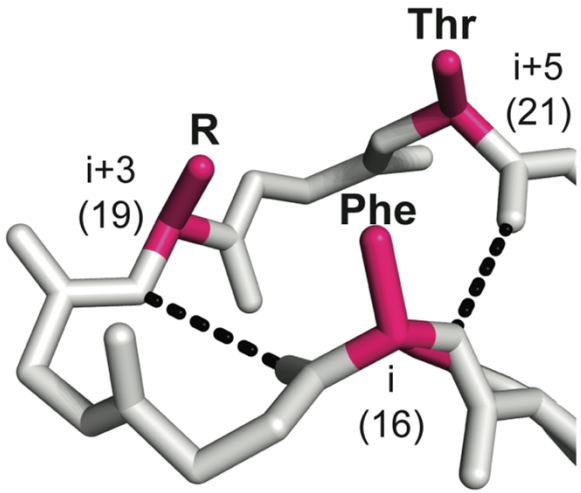
i	i+3	i+5	T_m (°C)	ΔT_m (°C)	$\Delta\Delta G_f$ (kcal/mol) at 65 °C
			56.2 ± 0.3	-2.6 ± 0.4	0.21 ± 0.06
”		”	53.6 ± 0.2		
			47.4 ± 0.4	7.6 ± 0.5	-0.70 ± 0.08
”		”	55.0 ± 0.3		

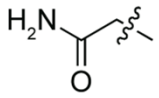
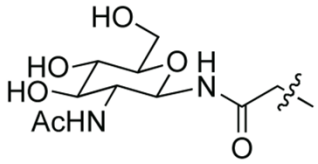
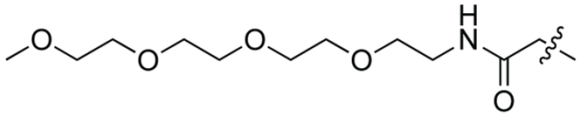
Can PEGylation Stabilize WW?

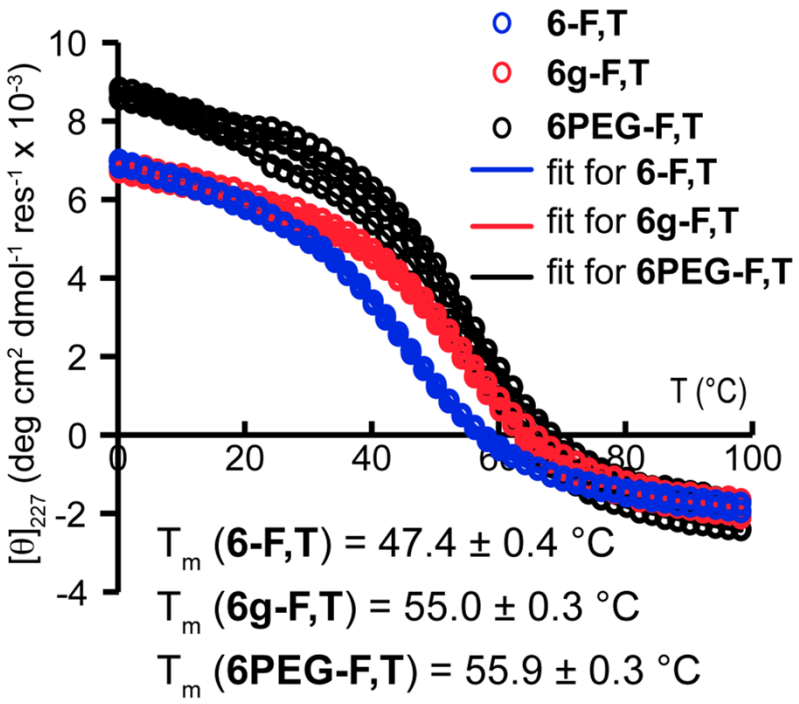


Protein	R =
6-F,T	
6g-F,T	
6PEG-F,T	

PEGylation Stabilizes WW like Glycosylation!



Protein	R =
6-F,T	
6g-F,T	
6PEG-F,T	



Do PEGylation and Glycosylation Stabilize WW by the Same Mechanism?

Protein	Sequence‡	T_m (°C)	ΔT_m (°C)	ΔG_f (kcal/mol)	$\Delta\Delta G_f$ (kcal/mol)
	15 21				
6	--M**S**RS**N**G**R**--	56.2 ± 0.3	-2.6 ± 0.4	0.95 ± 0.04	0.21 ± 0.06
6g	--M**S**RS**N**G**R**--	53.6 ± 0.3		1.16 ± 0.04	
6-F	--M**F**RS**N**G**R**--	51.0 ± 0.3	0.7 ± 0.4	1.45 ± 0.06	-0.17 ± 0.08
6g-F	--M**F**RS**N**G**R**--	51.7 ± 0.3		1.28 ± 0.04	
6-T	--M**S**RS**N**G**T**--	52.5 ± 0.3	-0.2 ± 0.5	1.22 ± 0.05	0.04 ± 0.07
6g-T	--M**S**RS**N**G**T**--	52.3 ± 0.3		1.26 ± 0.05	
6-F,T	--M**F**RS**N**G**T**--	47.4 ± 0.4	7.6 ± 0.5	1.72 ± 0.09	-0.70 ± 0.10
6g-F,T	--M**F**RS**N**G**T**--	55.0 ± 0.3		1.02 ± 0.04	

† Tabulated data are given as mean ± standard error at 65 °C for WW variants at 10 μM in 20 mM aqueous sodium phosphate, pH 7.

‡ **N** = Asn(glycan).

Glycosylation requires specific side-chains at the i and i+5 positions for stabilization to occur

Do PEGylation and Glycosylation Stabilize WW by the Same Mechanism?

Protein	Sequence‡	T_m (°C)	ΔT_m (°C)	ΔG_f (kcal/mol)	$\Delta\Delta G_f$ (kcal/mol)
	15 21				
6[b]	--M**S**RS**N**G**R**--	56.2 ± 0.3	8.0 ± 0.4	0.95 ± 0.04	-0.86 ± 0.05
6PEG	--M**S**RS**N**G**R**--	64.3 ± 0.2		0.09 ± 0.03	
6-F[b]	--M**F**RS**N**G**R**--	51.0 ± 0.3	10.9 ± 0.4	1.45 ± 0.06	-1.13 ± 0.07
6PEG-F	--M**F**RS**N**G**R**--	61.9 ± 0.3		0.32 ± 0.03	
6-T[b]	--M**S**RS**N**G**T**--	52.5 ± 0.3	5.3 ± 0.4	1.22 ± 0.05	-0.55 ± 0.06
6PEG-T	--M**S**RS**N**G**T**--	57.8 ± 0.2		0.68 ± 0.03	
6-F,T[b]	--M**F**RS**N**G**T**--	47.4 ± 0.4	8.5 ± 0.6	1.72 ± 0.09	-0.81 ± 0.10
6PEG-F,T	--M**F**RS**N**G**T**--	55.9 ± 0.3		0.91 ± 0.05	

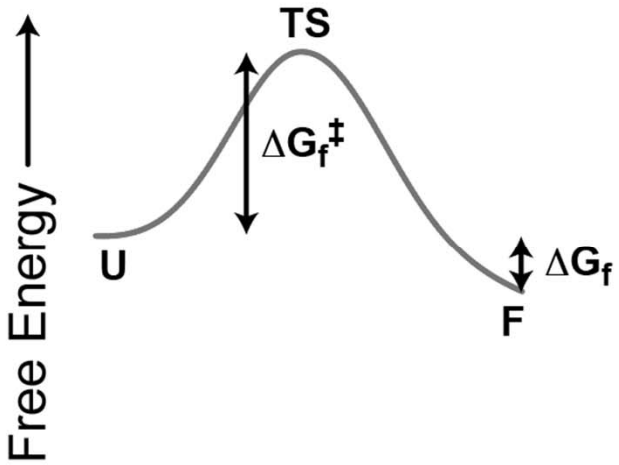
[a]Tabulated data are given as mean ± standard error at 65 °C for WW variants at 10 μM in 20 mM aqueous sodium phosphate, pH 7.

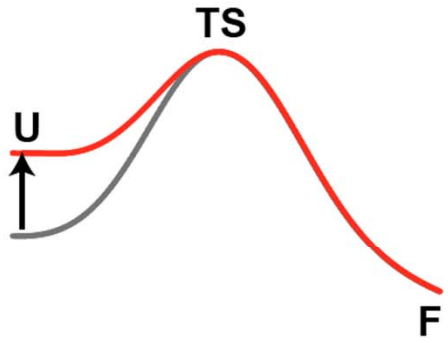
[b]Data for these proteins is from reference 29.

PEGylation is stabilizing regardless of the idenities of the i and i+5 position side chains

[c]**N** = Asn(PEG).

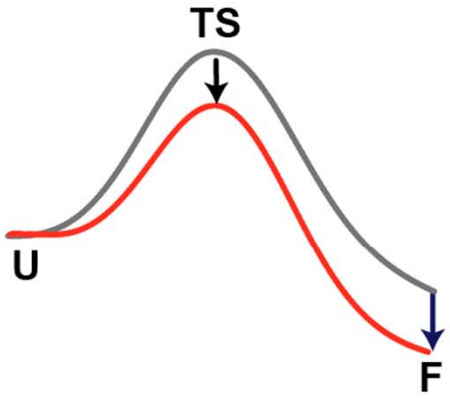
Do PEGylation and Glycosylation Stabilize WW by the Same Mechanism?





Destabilize the unfolded state

PEGylation?



Stabilize the folded state and the transition state

Glycsosylation?

Conclusion

- Genetic algorithms are robust for minimizing energy potential for protein structure prediction
- Genetic algorithms predict actual protein structure within accepted tolerance
- 30-60% accuracy good enough to correlate certain experimental results
- Higher accuracy can be achieved by incorporating other local interactions among amino acids

Conclusion

- Strategies for increasing protein thermodynamic stability
- Methods to accelerate protein folding rate
- Can we combine the two approaches?
  - Experimental results
  - Numerical simulations