GeoFold: A mechanistic model to study the effect of topology on protein unfolding pathways and kinetics

Vibin Ramakrishnan*, Saeed Salem‡, Saipraveen Srinivasan¶, Suzanne Matthews‡, Mohammed J. Zaki‡, Wilfredo Colón¶, **Christopher Bystroff** *‡

CENTER FOR BIOTECHNOLOGY AND INTERDISCIPLINARY STUDIES, DEPTS OF BIOLOGY, CHEMISTRY¶ & COMPUTER SCIENCE‡*
RENSSELAER POLYTECHNIC INSTITUTE, TROY, NEW YORK 12180

Funding from NSF DBI-0448072

Abstract:

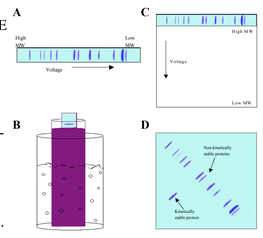
Protein unfolding is modeled as an ensemble of pathways, where each is a tree of intermediate states and branches in the tree represent additional degrees of conformational freedom. Using a known protein structure, the program (GeoFold) generates a directed acyclic graph of linked elemental subsystems, each modeling a partitioning of a substructure into two at topologically allowed positions in the chain. The graph begins at the native state and ends at small fragments representing the fully unfolded state. Each substructure is assigned a free energy based on its buried solvent accessible surface area, sidechain entropy, and backbone configurational entropy. Each bifurcating edge, representing the transition state of a single partitioning, is assigned a free energy barrier height based on the principle that exposed surfaces are solvated before the configurational entropy is expressed. To simulate unfolding on the graph, rates are calculated for each elemental subsystem at each time step using transition state theory. The model exhibits two-state behavior with respect to temperature or denaturant and shows the expected linear relationship of overall folding rate with denaturant. Predicted unfolding rates are compared with experimental values for fifteen well-studied proteins. Strengths and deficiencies of the model are discussed.

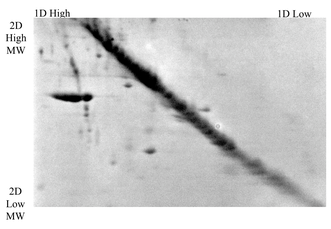
URL: http://www.bioinfo.rpi.edu/matths3/geofold.html

1. Kinetically stable (KS)proteins can be identified by differential resistance to SDS denaturation

High throughput method for isolating all KS proteins in a cell lysate

(A) Cell extract SDS-PAGE separation on a narrow gel strip.
(B) The gel strip is immersed in boiling 1% SDS.
(C) Second dimension SDS-PAGE separation.
(D) The gel is stained. The non-kinetically stable proteins migrate equal distances in both dimension.





Meinhold, D., Boswell, S., Colon, W., *Biochemistry* 2005, *44*, 14715-14724

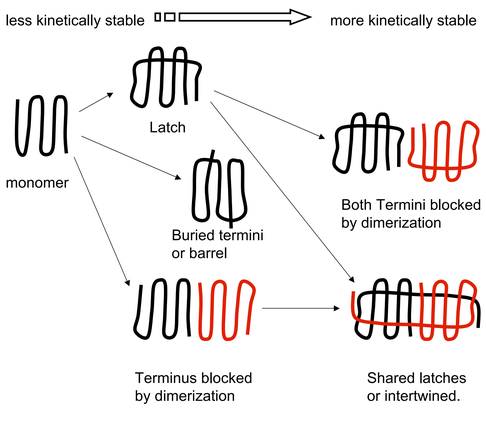
2. KS proteins in E.coli has recurrent structural themes

Selected seqs from E. coli lysate 2D gel

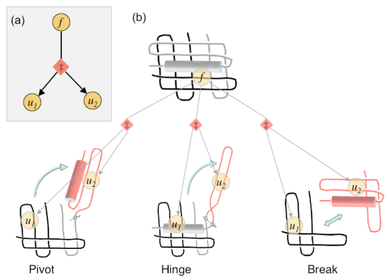
Kinetically stable proteins share structural features.

PDB code	MW	Length	4°	Class	OMP	Outer membrane protein M	Bound metal T	Termini buried L	Latch B	Termini blocked by oligomer
1dpc	26259	243	24	α/β				C	N	
1dwk	17263	156	10	α+α/β			N	N,C	N,C	
1hxx	37103	340	3	β	X					
1ino	19611	175	6	α/β		Mn	N,C	C	N,C	
1isc	21163	192	2	α+α/β		Fe	C			
1jsw	52427	478	4	α				N,C,loop		
1koj	63102	557	2	α/β				N,C,loop		
1mpf	37177	340	3	β	X					
1onr	35139	316	2	α/β			N,C	C		
1pd5	25700	219	3	α/β			C	N	N	
1pho	37633	330	3	β	X					
1s7c	35587	331	4	α/β			C		N	
1sbp	34542	310	1	α/β			N	C		
2ddm	30889	283	2	α/β			N,C		N	

Xia, K., Manning, M., Hesham, H., Lin, Q., *et al.*, *Proc Nat Acad Sci* 2007, *104*, 17329-17334



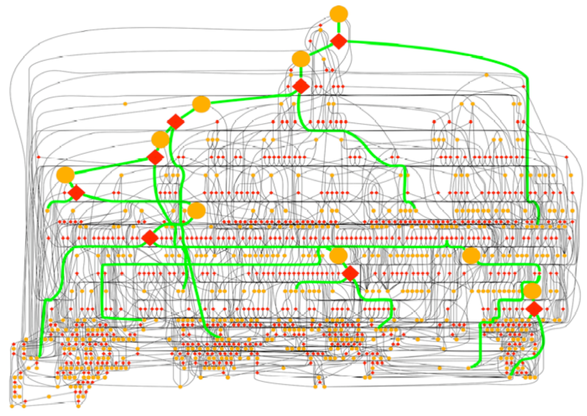
3. GeoFold models folding as a series of splittings of three types: break, pivot and hinge.



(a)Elemental subsystem for the kinetic model. *f* is a subset of the protein, and is partitioned into u_1 and u_2, passing over energy barrier ‡.
(b) Topological operators. A pivot motion is a rotation around a point. A hinge is a rotation around two points. A break is a translation. Rotations and translations must not cross chains.

Pivots, hinges and **breaks** have different entropy values ΔS, because they add different numbers of degrees of freedom to the system.

4. GeoFold generates a kinetic model for a protein .



GeoFold generates a directed acyclic graph (DAG) linking all topologically possible transitions (red diamonds), each splitting one intermediate substructure (orange circles) into two. After running UnfoldSim (see 5), the pathways that account for most of the unfolding traffic are identified (think lines and large symbols).

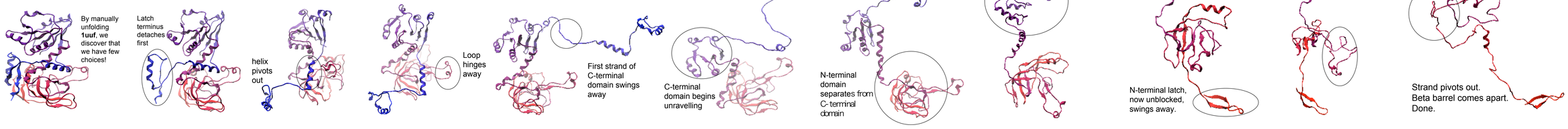
5. UnfoldSim runs a kinetic simulation .

The concentrations of all species except the native structure are set to zero. The change in concentration of *f* at each time step, for one transitions state (‡, red diamonds) is calculated as

$$k_{qu} = -\frac{d[f]}{dt} = \gamma[f]e^{\left(-\Delta G^{\ddagger}_{f-q}/RT\right)}$$

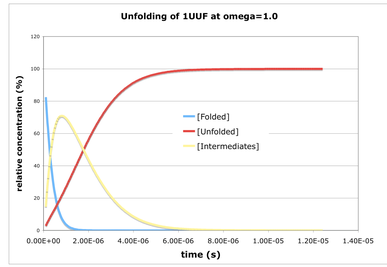
by the program UnfoldSim.

6. High traffic unfolding pathways may be dissected and visualized (here for 1UUF, alcohol dehydrogenase, a KS protein)



7. Unfolding rates can be measured as $k_u=\ln(2)/t_{1/2}$

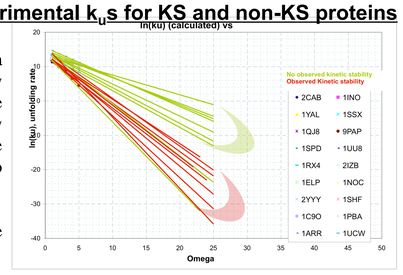
Following the changes in concentration with time, we sum over all the nodes that represent small segments of the chain, to get [Unfolded]. The time at which [Unfolded] reaches 50% is $t_{1/2}$, the half-life of unfolding. $t_{1/2}$ depends on temperature, urea concentration (1/omega in our simulations), and other parameters of the force field, including sidechain entropy, void volume, pivot entropy, hinge entropy, and break entropy.



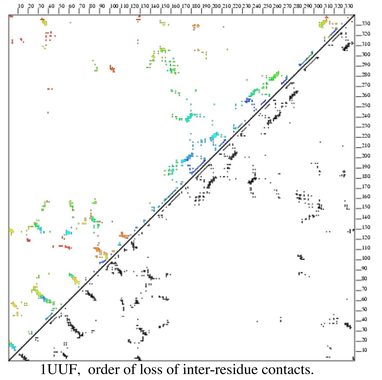
8. Simulated k_us agree with experimental k_us for KS and non-KS proteins

$\ln(k_u)$ can be plotted versus omega (solvation free energy, inversely related to [Urea]). We get the expected linear relationship. By extrapolating, we can infer the rates in pure water, which are too slow to simulate.

We find the GeoFold rates agree with experimental ones!



9. Unfolding pathways can be summarized using contact maps



1UUF, order of loss of inter-residue contacts.

When the simulation reaches 50% unfolded, we look at the concentration of each pairwise contact, rank the concentrations from small (red) to large (blue). Blue regions of the map show the folding core, while red regions show the early unfolding events (latches, dimer contacts, etc)