Conformational Switches in Protein Design and Evolution

Hue Sun Chan

Departments of Biochemistry, of Molecular Genetics, and of Physics
University of Toronto, Ontario M5S 1A8 Canada
http://biochemistry.utoronto.ca/chan/bch.html
Folding Cooperativity: switch-like ‘all-or-none’
Folding and Unfolding

Effects of Photoswitchable Crosslinks on Folding/Unfolding (collaboration with G. Andrew Woolley, U Toronto)

Conformational Switches in Protein Evolution
(collaboration with Erich Bornberg-Bauer, U Münster)
Folding cooperativity means two-state-like folding/unfolding

Experimental criteria from:

- calorimetry: $\frac{\Delta H_{vH}}{\Delta H_{cal}} \approx 1$
- chevron plots
- more direct probes of two-state behaviors

$Q = \text{fractional number of native contacts}$
The “Levinthal” Paradox was posed in response to the discovery of two-state protein folding by calorimetry

R. L. Baldwin (1994)

“In 1968 I was listening to a seminar at Stanford by Cy Levinthal entitled ‘How to fold gracefully’. … Two years earlier, in 1966, Lumry, Biltonen and Brandts had argued persuasively that the reversible folding/unfolding reactions of small proteins follow a two-state model (U ↔ N, where U = unfolded, N = native) without observable intermediates. In his Stanford seminar, Cy Levinthal took their model and pushed it to a reductio ad absurdum. He made a simple calculation showing that it would take longer than the lifetime of the universe for a small protein to fold up by a random search of all possible conformations. Paul Flory was sitting next to me at his seminar, he nudged me and whispered ‘so there must be folding intermediates’.”
Folding Cooperativity and the Levinthal Paradox

In view of the relevant historical context, a wide-open funnel is not a solution to the Levinthal paradox.

As a simple consequence of polymer physics, intermediate conformations always exist but their populations need not accumulate during folding.

Kinetic manifestation of folding cooperativity: linear chevron plots

Linear folding and unfolding arms imply a linear relationship between log(folding/unfolding rate) and equilibrium stability.

Chevron rollover is indicative of less cooperative or noncooperative folding.

Data from Jackson et al., Biochemistry 32:11270 (1993)
Folding cooperativity is not a corollary of a sequence’s ability to fold to an essentially unique structure

**case in point:**
**A simplified atomic model of 3-helix protein**

- The general pairwise additive sidechain hydrophobic and directional H-bond interactions in this model are not sufficient for folding cooperativity

“Topological” Modeling: $C_{\alpha}$ $G\ddot{o}$ Models

\[ V_{\text{total}} = V_{\text{stretching}} + V_{\text{bending}} + V_{\text{torsion}} + V_{\text{non-bonded}} \]

\[ = \sum_{\text{bonds}}^{N-1} K_r (r - r_0)^2 + \sum_{\text{angles}}^{N-2} K_\theta (\theta - \theta_0)^2 \]

\[ + \sum_{\text{dihedrals}}^{N-3} \left\{ K_{\phi}^{(1)} \left[ 1 - \cos (\phi - \phi_0) \right] + K_{\phi}^{(3)} \left[ 1 - \cos 3(\phi - \phi_0) \right] \right\} \]

\[ + \sum_{i<j<3} \epsilon \left[ 5 \left( \frac{r_{ij}'}{r_{ij}} \right)^{12} - 6 \left( \frac{r_{ij}'}{r_{ij}} \right)^{10} \right] + \sum_{i<j-3} \epsilon \left( \frac{r_{\text{rep}}}{r_{ij}} \right)^{12} \]

Shea et al., PNAS (1999); Micheletti et al., Phys Rev Lett (1999); Clementi et al., JMB (2000); Koga & Takada, JMB (2001); Kaya & Chan, JMB (2003)

Native-centric local and non-bonded interactions
Langevin dynamics
Thermodynamically quite cooperative
Desolvation Barriers

(a) Potential of mean force; (b) Cheung, García & Onuchic, PNAS (2002); (c) Hillson, Onuchic & García PNAS (1999); (d) Karanicolas & Brooks, Protein Sci (2002)
Desolvation barriers enhance folding/unfolding cooperativity

Increasing height of pairwise desolvation barrier leads to ...

Higher overall free energy barrier

More linear chevron plots

Less native fluctuation

Cheung et al., *PNAS* (2002)

Desolvation barrier effects are a likely contributor to the remarkable diversity in the folding rates of small proteins.

Flores et al., J Mol Biol (2009);
Koga & Takada, JMB (2001);
Chavez, Onuchic & Clementi, JACS (2004)

Figure from: Chan, Zhang, Wallin & Liu, Annu Rev Phys Chem (2011)
Folding Barriers

Traditional free energy profile

Nature of barrier different

Energy landscape

Imageries of Protein Folding

ΔG

free energy of conformational state

progress variable

“reaction coordinate”

Each “coordinate” can represent collectively many different conformations.

= Potential of mean force (PMF), with solvent degrees of freedom pre-averaged.

entropic & enthalpic components
**Enthalpic barriers:**

Non-Arrhenius folding rates, positive unfolded-to-transition state enthalpy changes at some temperatures

Does this mean that the folding landscape is not funnel-like?


*Figure from: Chan, Zhang, Wallin & Liu, Annu Rev Phys Chem (2011)*
A common misunderstanding of the funnel picture of protein folding

"… A funnel-like landscape assumes mainly entropic barriers, which is in contradiction to the experimental work that is displayed in table 1. All of these experiments show major enthalpic contributions to the free energy barriers. …"
α–helix association in water as a model for rate-limiting events in protein folding

A pair of 20-residue poly-alanine or poly-leucine helices

~3,800 water molecules

Simulated constant-pressure free energy of association (potential of mean force, PMF) at five temperatures

MacCallum, Moghaddam, Chan & Tieleman, PNAS (2007)
Enthalpic desolvation barriers of ~ 50 kJ/mol comparable to that of protein folding

Dramatic enthalpy-entropy compensation at the desolvation step leading to low or non-existent free energy barriers

At 25 deg C,
Enthalpic folding barrier height for
CI2 is ~ 30kJ/mol
(Oliveberg et al., 1995)
CspB is ~ 32kJ/mol
(Schindler & Schmid, 1996)
Enthalpic Folding Barrier ≠ Non-Funnel Landscape

Figure from: Chan, Zhang, Wallin & Liu, Annu Rev Phys Chem (2011)
Enthalpic barriers caused by steric dewetting – “large” parts of the protein coming together at the rate-limiting step

Experimental correlation between activation volume and activation enthalpy?

MacCallum et al., Proc Natl Acad Sci USA (2007)
See also discussion in: Ferguson et al., J Mol Biol (2009)
Hummer et al. (1998):

From pressure dependence of pairwise and triplet hydrophobic interactions to physics of pressure denaturation


- explains why slightly expanded structures ~ ssm are favored at high *Ps*
- $\Delta V^\dagger(\text{ssm} \rightarrow \text{db}) \approx 3.8 \text{ ml/mol}$, and $\Delta V^\dagger(\text{cm} \rightarrow \text{db}) \approx 1.6 \text{ ml/mol}$ [from $\partial(\text{PMF})/\partial P$]

“..."
What can we learn from pairwise methane properties about pressure effects on protein folding?

- Void volume in the folded protein core is an essential factor in the thermodynamic equation, the \( \text{cm} \) configuration is not necessarily a good model for the folded state.

- Compressibility of the folded state is a key determinant of its stability under pressure; but this property may not be adequately reflected by the compressibility at \( \text{cm} \).

- Pairwise methane data provide valuable insights into \( P \)-dependent transition-state and denatured/unfolded-state properties; but do not by themselves address the relative stabilities of the folded vs unfolded states. [What if the two-methane volume is smaller in the protein core than at \( \text{ssm} \)?]

\[\text{cf. Neumaier et al. & Kiefhaber, PNAS (2013)}\]
**Kinetic manifestation of folding cooperativity: linear chevron plots**

- Desolvation-barrier (db) and Sidechain (SC) models are more cooperative than common Gō models.
- Transfer free energy models \([\sim O’Brien et al., PNAS 105:13403 (2008)]\) are only slightly more cooperative than SC-Gō models.

*Scalley et al., Biochemistry 36:3373 (1997)*
Cooperativity effects of desolvation barriers in a more general perspective:

**Spatial Ranges of Driving Forces are a Key Determinant of Protein Folding Cooperativity and Rate Diversity**

- Restricting the range of native-centric attractive interactions to 1.2 times the native distance in common Cα Gō models.
- For a typical native Cα-Cα distance ~6.4Å, this range restriction is comparable to that imposed by the desolvation barrier (db) because (0.2)(6.4Å) = 1.3Å ≈ 1.4Å ≈ radius of a water molecule.

Restricting the spatial ranges of favorable interactions leads to a dramatic increase in folding rate diversity

Order of magnitude of folding-rate diversity:

- Experimental: $5.96$
- Long-range: $2.13$
- Short-range (1): $4.80$
- Short-range (2): $5.81$ (estimated)

Trade-off between interaction specificity and spatial range effects on folding cooperativity

Comparison with a recent model of cooperative homopolymer coil-globule transition:

- For native-centric protein models, the restrictions needed to produce cooperative folding are moderate and can readily arise physically from desolvation effects \([\approx 0.2(6.4\text{Å}) \approx 1.3\text{Å}]\).

- For a homopolymer of similar chain length and monomer size, the range of the nonspecific attraction needs to be several times narrower, viz., \(\approx 0.05(6.4\text{Å}) \approx 0.3\text{Å}\).

**Figures from:** Taylor, Paul & Binder, “All-or-none proteinlike folding transition of a flexible homopolymer chain,” *Phys Rev E* 79, 050801(R) (2009).

Desolvation barriers likely contribute significantly to folding cooperativity, the immense diversity in folding rates among different proteins, and the volumetric barriers to folding.

The sharpness of coil-globule transition and folding cooperativity generally increase with a decreasing spatial range of attractive intrachain interactions. The role of desolvation barriers in enhancing folding cooperativity may be understood in this general context.
Protein folding/unfolding controlled by photoswitchable linkers

Modeling Fyn SH3 with a BSBCA or an alkyne crosslinker

Molecular Dynamics (Amber99) simulations of free linkers:
The average end-to-end distance of trans2 > trans1

Destabilization of the native structure increases with the rigidity of the \textit{trans} linker.
What is critical for photocontrol is the availability of linker conformations that are consistent with the folded state, *not* the mean length of the free linker.

trans(2) longer than trans(1) on average, but trans(2) is less destabilizing than trans(1) because it is more flexible.

$\Delta G/k_B T$ of destabilization:

<table>
<thead>
<tr>
<th></th>
<th>Experiment</th>
<th>Theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-crosslinked</td>
<td>4.1</td>
<td>3.6</td>
</tr>
<tr>
<td>2-crosslinked</td>
<td>2.4</td>
<td>1.2</td>
</tr>
</tbody>
</table>

Destabilization is largely determined by the free trans linker fractional population with native-like end-to-end distances

Interplaying effects of trans linker stiffness and average end-to-end distance on folding stability

WT: $\Delta G/\kappa_B T \approx 0.4$

Contour plot obtained by varying $k_0$ and $\theta_0$

Folding cooperativity in the model Fyn SH3 with a 3-29 crosslink is enhanced by cis linkers but reduced by stiff trans linkers.

Native destabilization increases with increasing stiffness of the trans linker.

Free energy profiles at the models’ respective transition midpoints.

\[ \Delta G/k_B T \]

- Diagram showing a linear relationship between free energy change and stiffness for trans linkers.

- Diagram illustrating free energy profiles with different linkers.

Linker constraints on residue-residue distance

Stiff trans linkers can induce population of compact nonnative conformations

Coarse-grained model simulation shows that a *trans* linker between positions 3 & 29 in Fyn SH3 can lead not only to global unfolding but also partially unfolded (misfolded) conformations, consistent with experimental NMR data.
Coarse-grained models have the advantage of computational tractability and conceptual clarity. Inasmuch as their limitations are taken into account, coarse-grained models are useful tools for generating and evaluating hypotheses. They provide physical rationalizations for experiments and can be predictive. A good example is the present application to the study of destabilization effects of photoswitchable crosslinks.

What is critical for photocontrol is the fractional population of linker conformations that are consistent with the folded state, *not* the mean length of the free linker.
Conformational Switches in Protein Evolution

evolving different protein folds: a conceptual study
A Simple Exact Biophysical Model: The Hydrophobic-Polar (HP) Model

The Hydrophobic-Polar (HP) Lattice Protein Model

- Hydrophobic-hydrophobic contacts are favorable
- Allows for exact (exhaustive) enumeration of sequences and conformations
- Provides an exact sequence-to-structure mapping for the study of evolution

Short-chain 2D HP model capture several essential features of the sequence-to-structure mapping of real proteins [Dill et al & Chan, *Protein Sci* (1995)].

- **hydrophobic-polar statistics:**

- **comparative (“homology”) modeling:**

- only ~ 2% of sequences are encoding uniquely

What properties of real proteins can simple HP model capture?
**HP** pattern is an important determinant of protein structure

Figure from: Cordes, Walsh, McKnight & Sauer, *Science* **284**, 325 (1999)

The **N11L** “bridge” single-mutant folds into two nearly equally populated native structures

Sequence Pattern Statistics of HP Model Proteins are similar to that of Real Proteins

\[ \sigma_i = \pm 1, \text{ where } +1 \text{ for } H; -1 \text{ for } P \]

for real proteins, \( H = L, I, V, F, M, \text{ or } W; P \) otherwise.

Consider mean-square block fluctuation

\[ \psi^{(S)} = \frac{s}{n} \sum_{k=1}^{n/s} \psi^{(s)}_k \]

where

\[ \psi^{(s)}_k \propto (\sigma^{(s)}_k - s \phi)^2 \text{ and } \phi = \frac{1}{n} \sum_{i=1}^{n} \sigma_i \]

- Similar nonrandomness is observed for uniquely encoding sequences in the HP model and for real proteins.

Superfunnels: Correlation between Thermodynamic and Mutational Stabilities

Neutral net of sequences encoding for a given structure tends to organized around a “prototype sequence” with maximum thermodynamic stability and mutational stability (robustness).

- Sequences with higher mutational robustness tend to have higher steady-state populations under evolutionary dynamics.

Bornberg-Bauer & Chan, *PNAS* (1999);
Interplay of network topology and fitness effects on evolutionary population

Mutational robustness alone may not be sufficient to account for the experimentally observed concentration of evolutionary populations at prototype-like sequences. Selection for native stability and kinetic stability is evident.
Neutral Nets and Supernets in Sequence Space

Sikosek & Chan, J R Soc Interface (2014), in press
Recombinatoric exploration of folded structures

- more efficient than single-point mutations

Cui, Wong, Bornberg-Bauer & Chan, PNAS (2002)
Sequences not reachable by single-point mutations are accessible by crossovers

= neutral net; 897 neutral nets ($n = 18$; 4,553 sequences total) in the “supernet”

Local Signal for Sequence Uniqueness

- Model proteins prefer certain local sequence patterns.

- Consistent with a subsequent experiment on β-lactamase showing that for a given number of amino acid substitutions, recombined variants are much more likely to retain function than variants generated by random point mutations [Drummond, Silberg, Meyer, Wilke & Arnold, *PNAS* (2005)].
Recombinatoric crossovers can help “tunnel” underneath rugged evolutionary landscape.

Autonomous folding units in the HP model


Cui, Wong, Bornberg-Bauer & Chan, PNAS (2002)
Latent Evolutionary Potentials: 
Selection of excited-state (promiscuous) functions can speed up evolution dramatically

Wroe, Chan & Bornberg-Bauer, HFSP J (2007)
Our model provides a biophysical rationalization for the intriguing, and otherwise puzzling experimental observation that adaptation to new requirements can proceed while the “old,” phenotypically dominant function is maintained along a series of seemingly neutral mutations.
Tawfik and coworkers characterized ~300 variants of the enzyme PON1 that are apparently neutral, or close to neutral, with respect to PON1’s levels of expression and native lactonase activity. Their activities with promiscuous substrates and ligands indicated significant changes in adaptive potentials.


“Inasmuch as the selection of nonnative functions and structures is operative, the attraction of any superfunnel towards its prototype may extend to proteins that are not yet part of its neutral network.” — Sikosek et al, *PNAS* (2012)
Conformational switches in real proteins


- One-mutation switch between the human serum albumin-binding domain (G_A) and the IgG-binding domain (G_B) of *Streptococcus* protein G [Alexander, He, Chen, Orban & Bryan, *PNAS* **106**, 21149 (2009)]. NMR structures were determined for G_A95 & G_B95 (the 56aa sequences are 95% identical). Three substitutions in G_A95 or G_B95 shift the equilibrium from >99% in one structure to >99% in the other.

Model calculations we performed on the experimental G_A/G_B series of sequences indicate that the stability of the excited (latent) state gradually increases (i.e., with decreasing free energy) as the switch point is approached.

Experimentally designed bi-stable proteins and mutation-induced conformational switches

Bouvignies et al., Nature (2011)


Meier et al., Curr Biol (2007)

Alexander et al., PNAS (2009)

Anderson et al., Protein Eng Des Sel (2011)
Role of excited-state selection in Escape from Adaptive Conflict

Escape from Adaptive Conflict Follows from Weak Functional Trade-Offs and Mutational Robustness.

- Biophysics-based network connections.
- Evolutionary dynamics under mutations and gene duplications computed using both an analytical master equation and stochastic Monte Carlo simulations.
- Fitness is proportional to the stability (concentration) of the functional structures up to a certain optimum concentration above which fitness does not increase further with concentration.
- The optimum concentration corresponds to a measure of selection pressure.

Evolving a new folded structure:

traversing two superfunnels
Neofunctionalization follows from strong selection pressures (strong trade-offs), whereas Subfunctionalization results from intermediate selection pressures (weak trade-offs).
Escape from Adaptive Conflict (EAC) in the real world

Experiments showing that a reconstructed common ancestor of the fluorescent proteins in corals that emit either red or green light can emit light of both colors are indicative of EAC.

**Subfunctionalization (SUBF) as a Consequence of Neutral Network Topology**

- A strong tendency toward mutationally robust genotypes (i.e., those with many sequence–space neighbors): a sequence–space “entropy” effect.
- Because of this “entropic” driving force, SUBF can be nonadaptive.

---

![Graph](image)

The graph shows the logarithmic steady-state populations of gene pairs against the number of neutral genotypes within distance 2. The inset highlights the relationship between biophysics-based model seq-space connections and random seq-space connections. The data is from Sikosek, Chan & Bornberg-Bauer, *PNAS* (2012).
Subfunctionalization can be driven solely by Mutational Robustness

generalists less robust than specialists
Summary

- Biophysical sequence-to-structure mappings based upon simple explicit-chain protein models are versatile conceptual tools for addressing general principles of evolution.

- The superfunnel paradigm underscores a fundamental positive correlation between thermodynamic stability of the native structure and the mutational stability of a protein.

- Excited-state selection of promiscuous function can speed up evolution considerably.

- Subfunctionalization after duplication of a bi-stable gene with dual functions can be driven by sequence-space topology (i.e., mutational robustness) in an essentially nonadaptive manner.
Coworkers

**University of Toronto**
*Current group members:*
Tao Chen • Loan Huynh
Tobias Sikosek • Jianhui Song

*Former group members:*
Artem Badasyan
Mikael Borg
Cristiano Dias
Allison Ferguson
Hüseyin Kaya
Michael Knott
Zhirong Liu
Maria Sabaye Moghaddam
Seishi Shimizu
Stefan Wallin
Zhuqing Zhang

---

**University of Münster**
Prof. Erich Bornberg-Bauer
Richard Wroe
David Vernazobres
Tobias Sikosek

**Baylor College of Medicine**
Prof. E. Lynn Zechiedrich
Jennifer K. Mann

---

*Supported by the Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada & the Canada Research Chairs Program*