

Coarse-Graining Methods in Biology and Materials

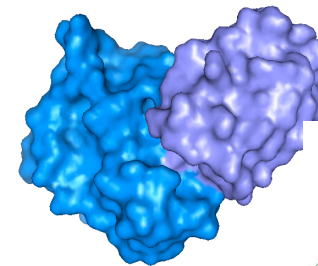
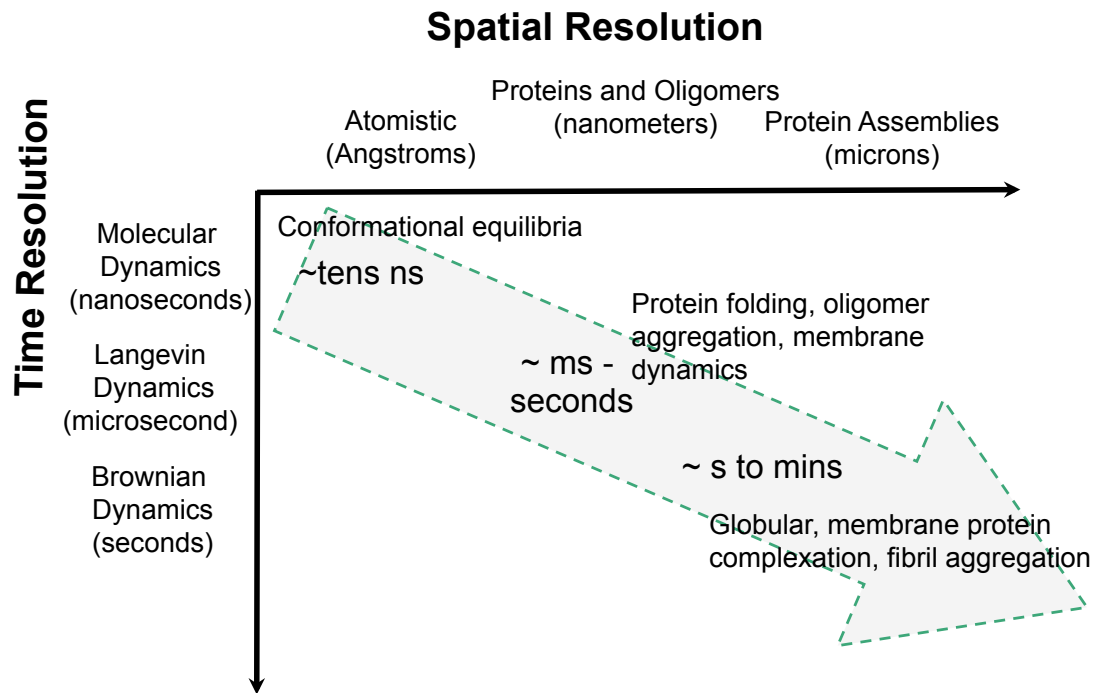
School on Multiscale Modeling and
Simulations of Hard and Soft Materials

Professor Teresa Head-Gordon
Department of Bioengineering
University of California, Berkeley



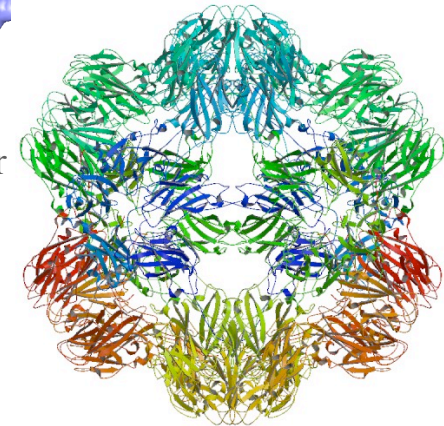
Lecture 3

Coarse-Graining and MultiScale



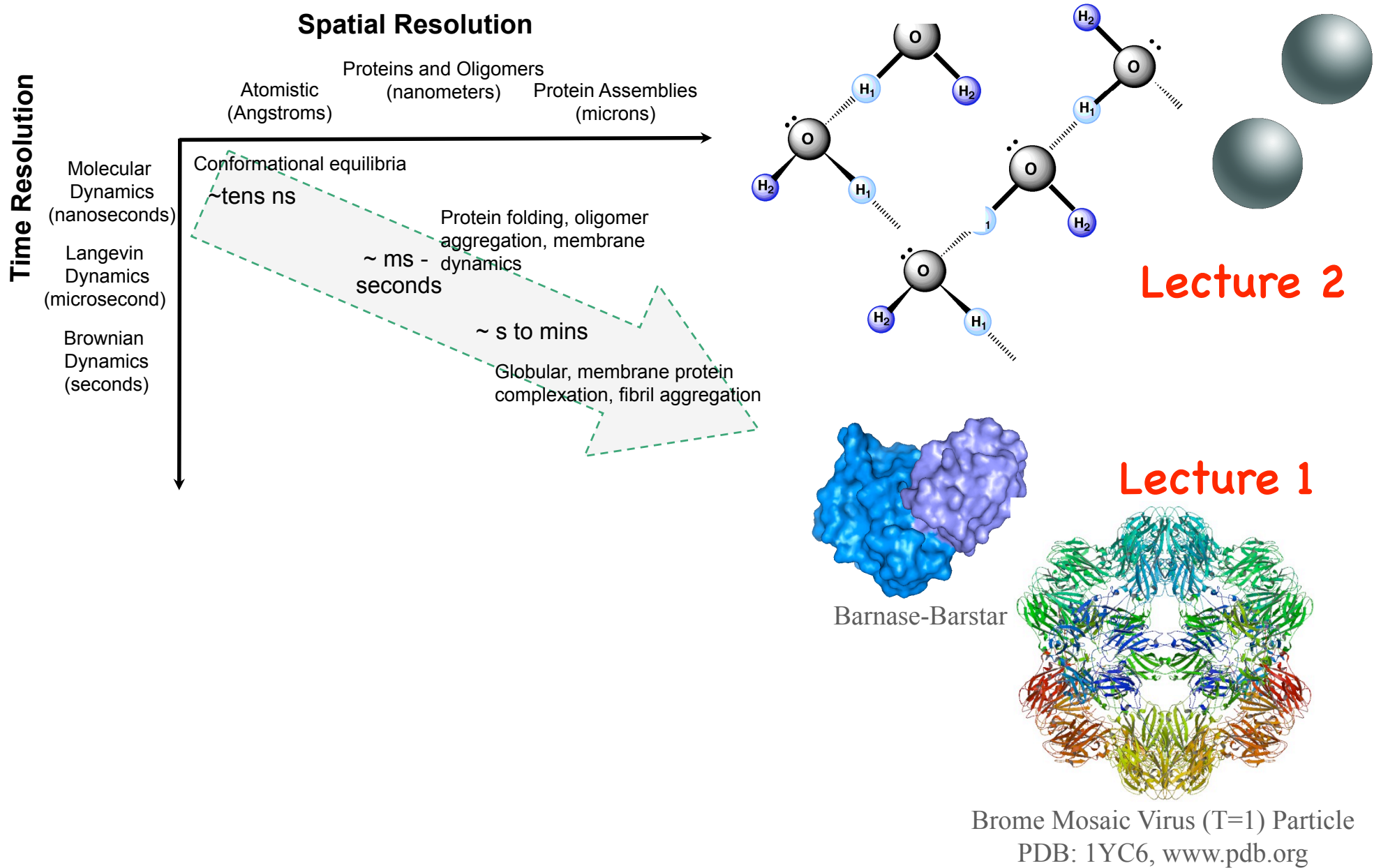
Barnase-Barstar

Lecture 1

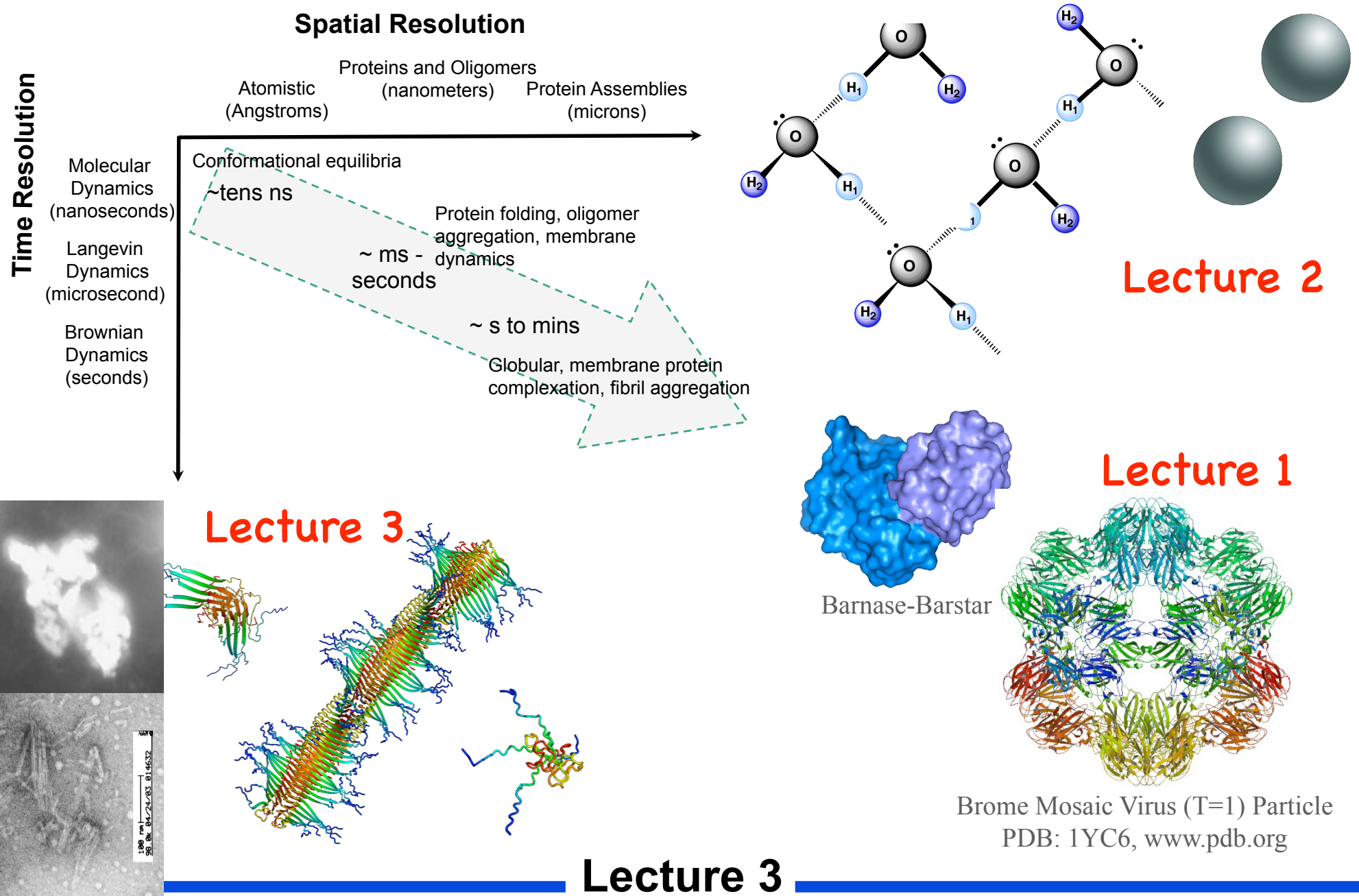


Brome Mosaic Virus (T=1) Particle
PDB: 1YC6, www.pdb.org

Coarse-Graining and MultiScale

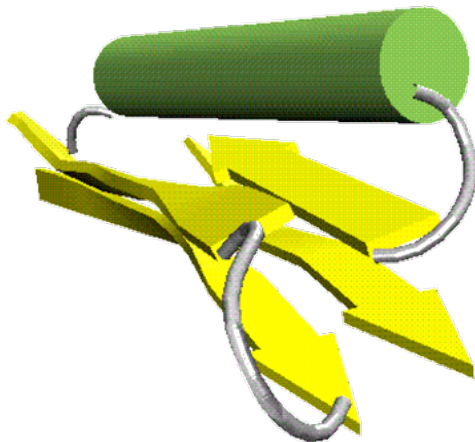
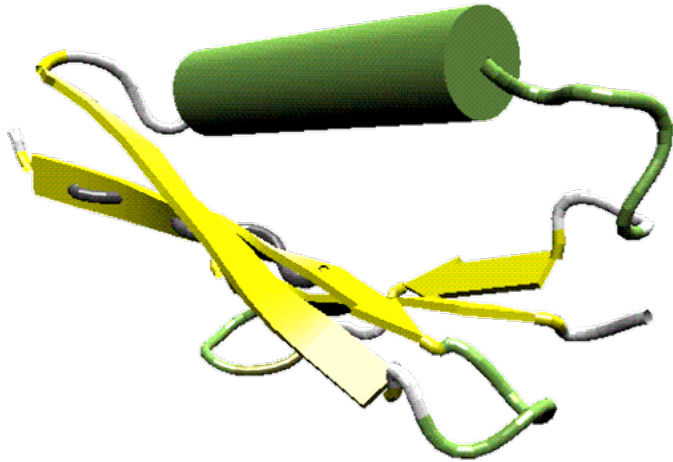


Coarse-Graining and MultiScale



Coarse-Grained Protein Models

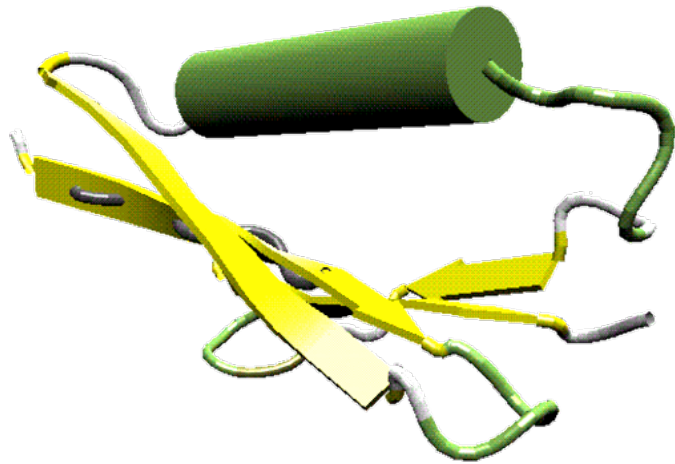
Why would these be expected to work? Missing amino acid chemistry, side chain packing, (hydrogen-bonding)



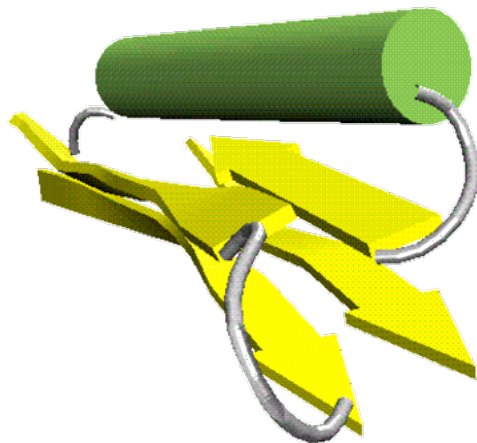
THG & Brown (2003). *Curr. Opin. Struct. Biol.* 13, 160-167.

Coarse-Grained Protein Models

Why would these be expected to work? Missing amino acid chemistry, side chain packing, (hydrogen-bonding)



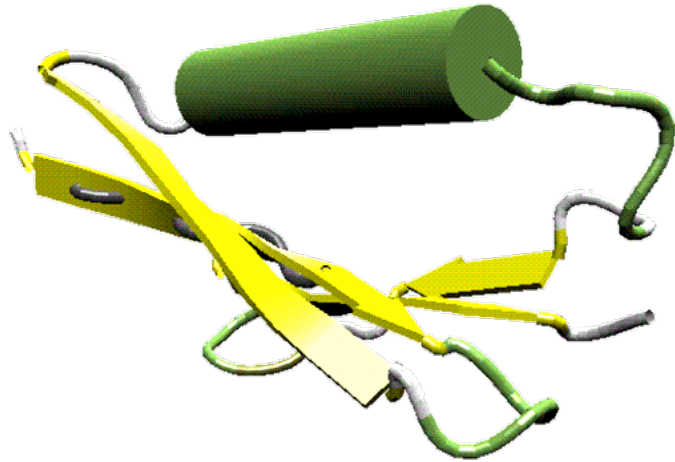
Topology is important: Folding rate for simple two-state folders is strongly correlated with average sequence separation between contacting residues in the native state (Plaxco et al., 2000).



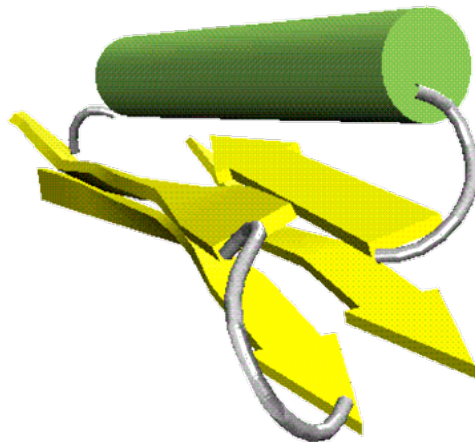
THG & Brown (2003). *Curr. Opin. Struct. Biol.* 13, 160-167.

Coarse-Grained Protein Models

Why would these be expected to work? Missing amino acid chemistry, side chain packing, (hydrogen-bonding)



Topology is important: Folding rate for simple two-state folders is strongly correlated with average sequence separation between contacting residues in the native state (Plaxco et al., 2000).

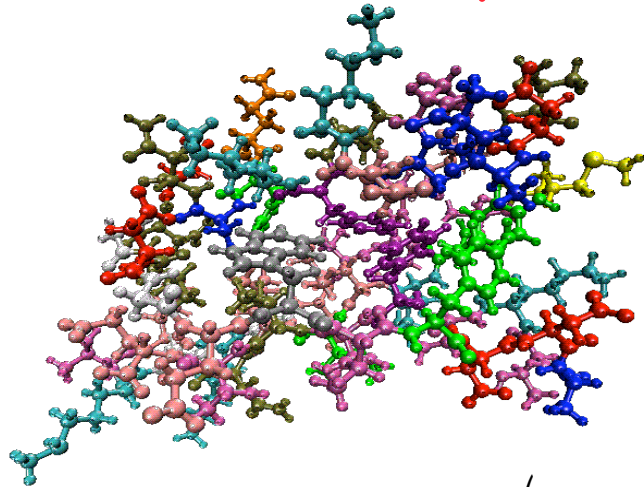


Coarse-grained sequences: needed for sequence dependence on folding. Should be highly appropriate for aggregation studies, since hydrophobic/hydrophilic sequence patterning (Broome, Hecht 2000; Schwartz and others 2001)

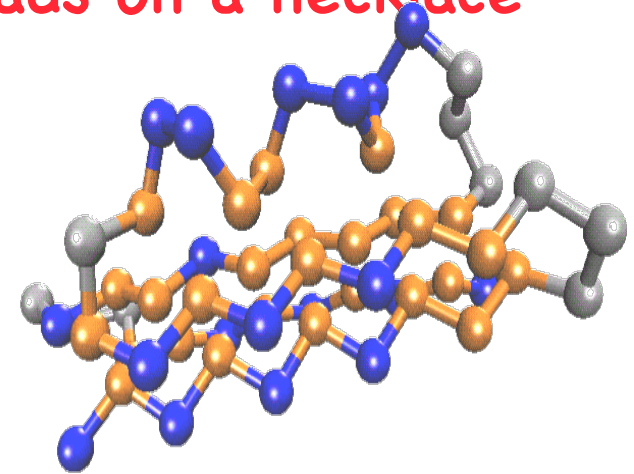
THG & Brown (2003). *Curr. Opin. Struct. Biol.* 13, 160-167.

Our Coarse-Grained Protein Model

Full atomic complexity



Beads on a necklace



$$\begin{aligned}
 H = & \sum_{\text{angles}} k_{\theta} (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} \{A[1 + \cos \phi] + B[1 - \cos \phi] + C[1 + \cos 3\phi] \\
 & + D[1 + \cos(\phi + \pi/4)]\} + \sum_{i,j \geq i+3} 4\epsilon_H S_1 \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - S_2 \left(\frac{\sigma}{r_{ij}} \right)^6 \right] + \\
 & \epsilon_{HB} G(r_{ij} - r_{HB}) G'(\mathbf{t}_{i,i-1} \cdot \mathbf{u}_{ij} - \cos \theta_{HB}) G'(\mathbf{t}_{j,j-1} \cdot \mathbf{u}_{ij} + \cos \theta_{HB})
 \end{aligned}$$

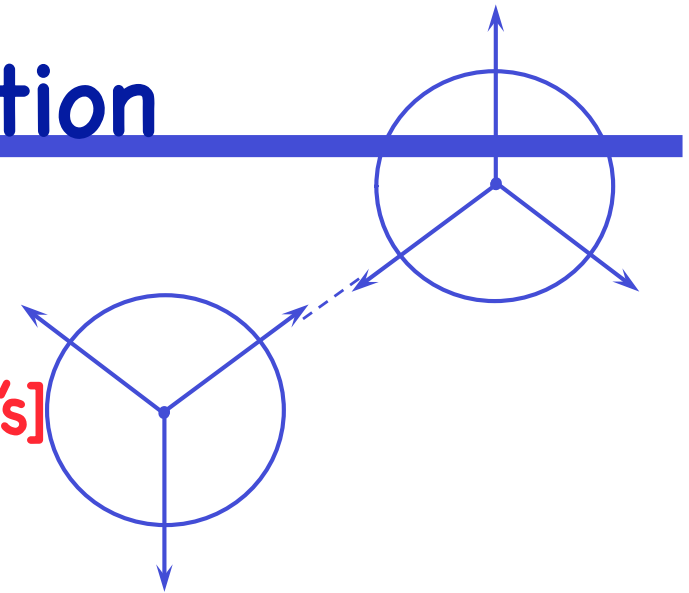
$$\epsilon_{HB} = 1.1\epsilon_H; \quad \epsilon_H \approx 1.5 \text{ kcal / mole}$$

Yap, Fawzi, THG (2008) Proteins

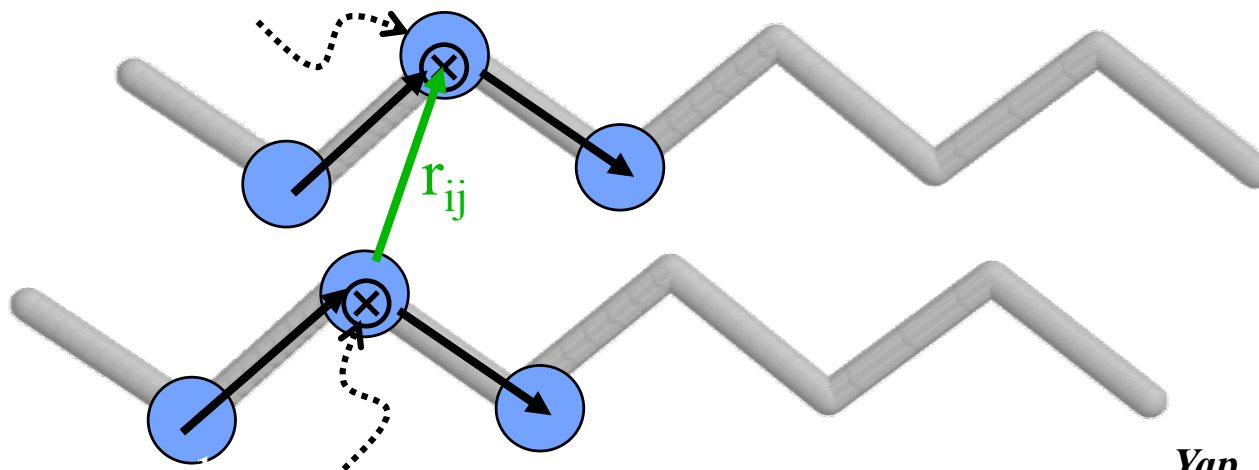
Hydrogen Bond Energy Function

H-bond energy function is inspired by
2-D Mercedes Benz Model of water

[Ben-Naim 1971; Dill and co-workers, 1990's]



$$U_{HB} = \underbrace{\varepsilon_{HB} F(r_{ij} - r_{HB})}_{\text{distance-dependent term}} \underbrace{G\left(\left|t_{HB,i} \cdot \hat{r}_{ij}\right| - 1\right) H\left(\left|t_{HB,j} \cdot \hat{r}_{ij}\right| - 1\right)}_{\text{angle-dependent terms}}$$



Yap, Fawzi, THG (2008) Proteins

Our Coarse-Grained Protein Model

Four flavors consisting of large attraction (B), small attraction (S), small repulsion (N), and large repulsion (L)

Trp	B	Met	B	Gly	N	Asn	L
Cys	B	Ala	S	Ser	N	His	L
Leu	B	Val	S	Thr	N	Gln	L
Ile	B	Tyr	S	Glu	L	Lys	L
Phe	B	Pro	N	Asp	L	Arg	L

Mapping 20-letter amino acid code to 4-Letter minimalist code

$$\begin{aligned}
 H = & \sum_{\text{angles}} k_{\theta} (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} \{A[1 + \cos \phi] + B[1 - \cos \phi] + C[1 + \cos 3\phi] \\
 & + D[1 + \cos(\phi + \pi/4)]\} + \sum_{i, j \geq i+3} 4\epsilon_H S_1 \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - S_2 \left(\frac{\sigma}{r_{ij}} \right)^6 \right] + Ehb
 \end{aligned}$$

S_1 and S_2 determine the different flavors

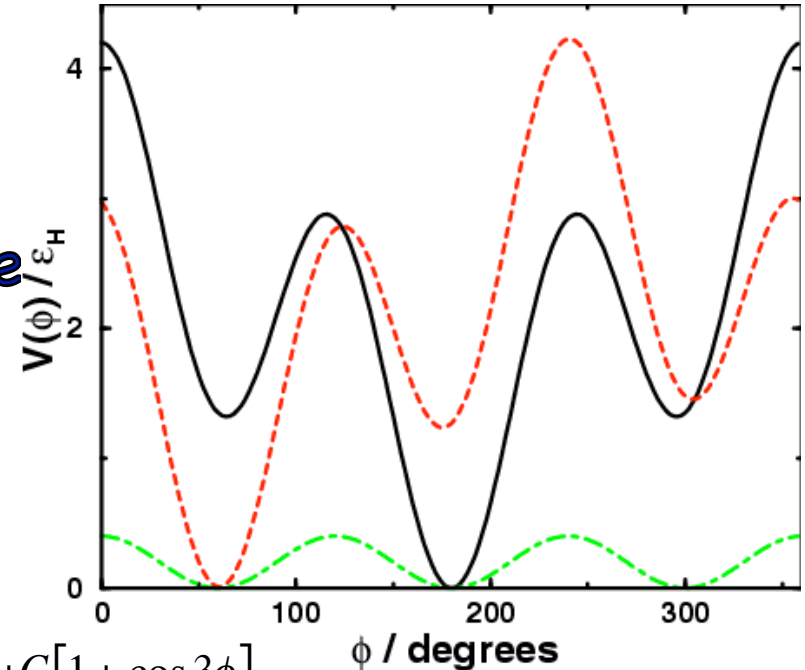
Yap, Fawzi, THG (2008) Proteins

Our Coarse-Grained Protein Model

Secondary structure specified: α , β , coil

Go-like (need to know native)

Not quite like Go model since can transition to other secondary structure (albeit through large barriers)



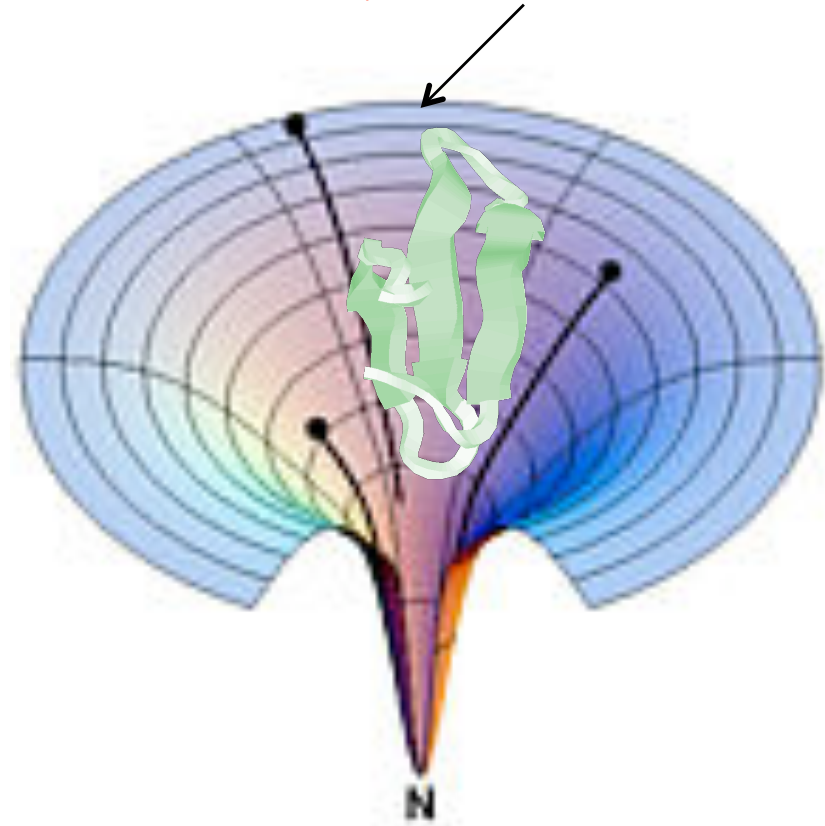
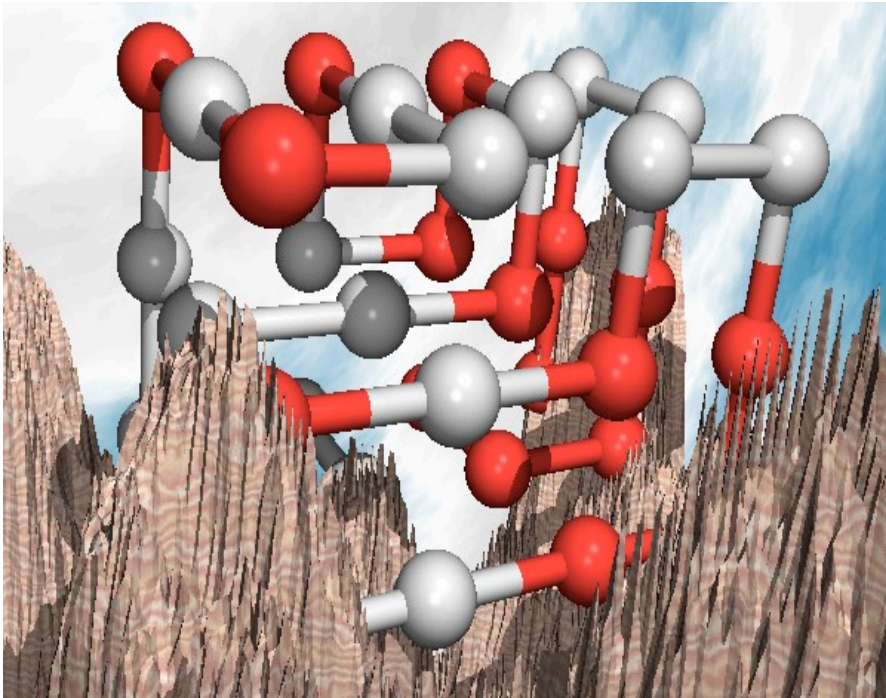
$$H = \sum_{\text{angles}} k_{\theta} (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} \{A[1 + \cos \phi] + B[1 - \cos \phi] + C[1 + \cos 3\phi]$$

$$+ D[1 + \cos(\phi + \pi/4)]\} + \sum_{i,j \geq i+3} 4\epsilon_H S_1 \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - S_2 \left(\frac{\sigma}{r_{ij}} \right)^6 \right] + Ehb$$

A,B,C,D determine the dihedral angle type

Principle of Minimal Frustration

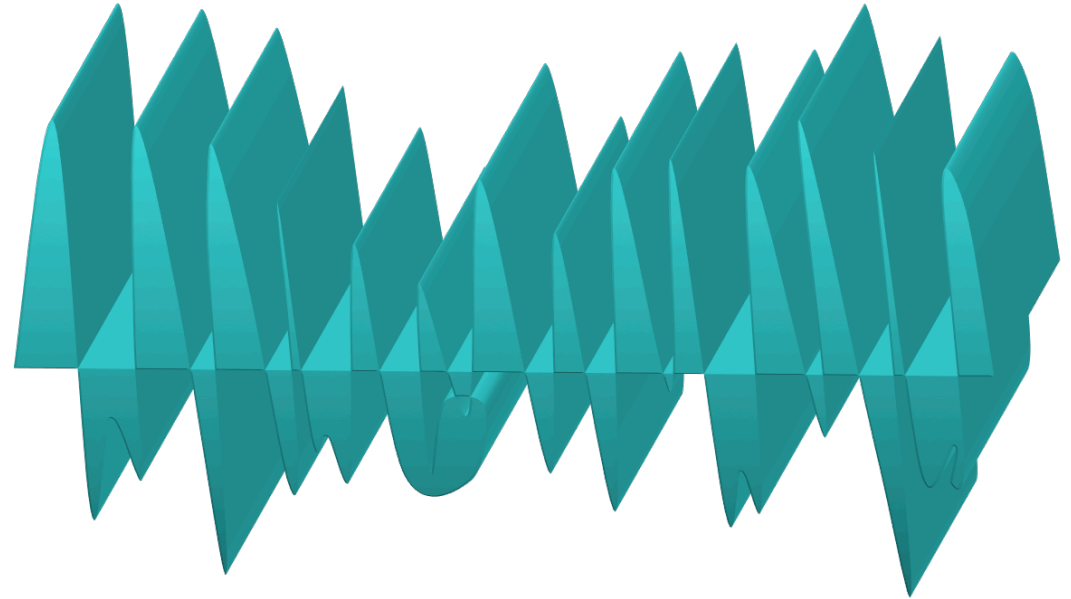
Random heteropolymers are unlike bipoymers:



“Evolution has designed native sequences to achieve efficient folding to a structurally organized ensemble with few traps arising from discordant energetic signals”

Principle of Minimal Frustration

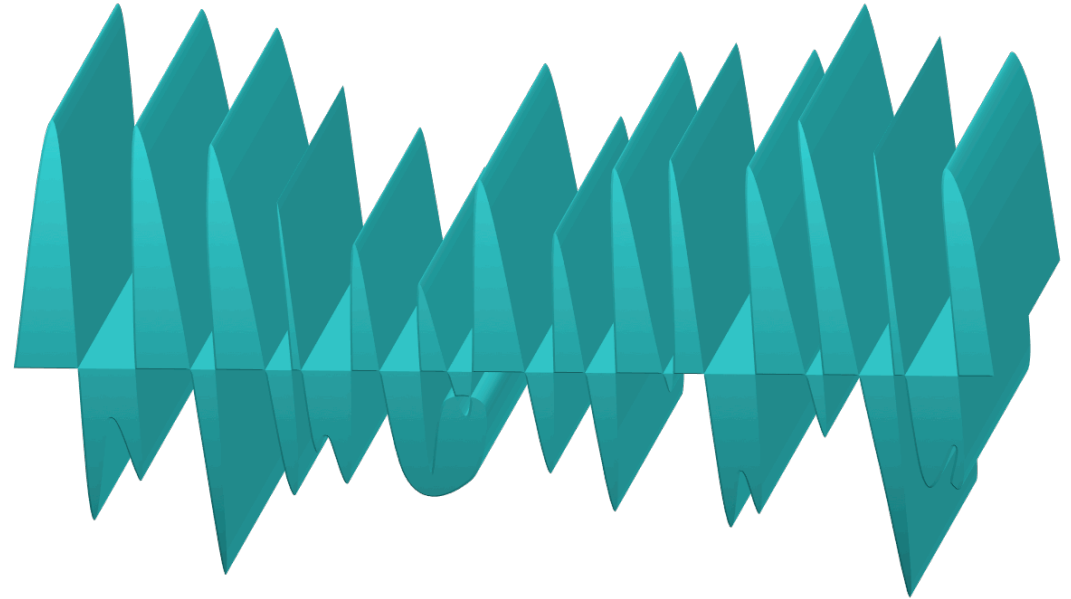
Choose a topology, a start sequence, and assign secondary structure



Long-lived traps of compact, non-native conformations with similar energetics

Principle of Minimal Frustration

Choose a topology, a start sequence, and assign secondary structure

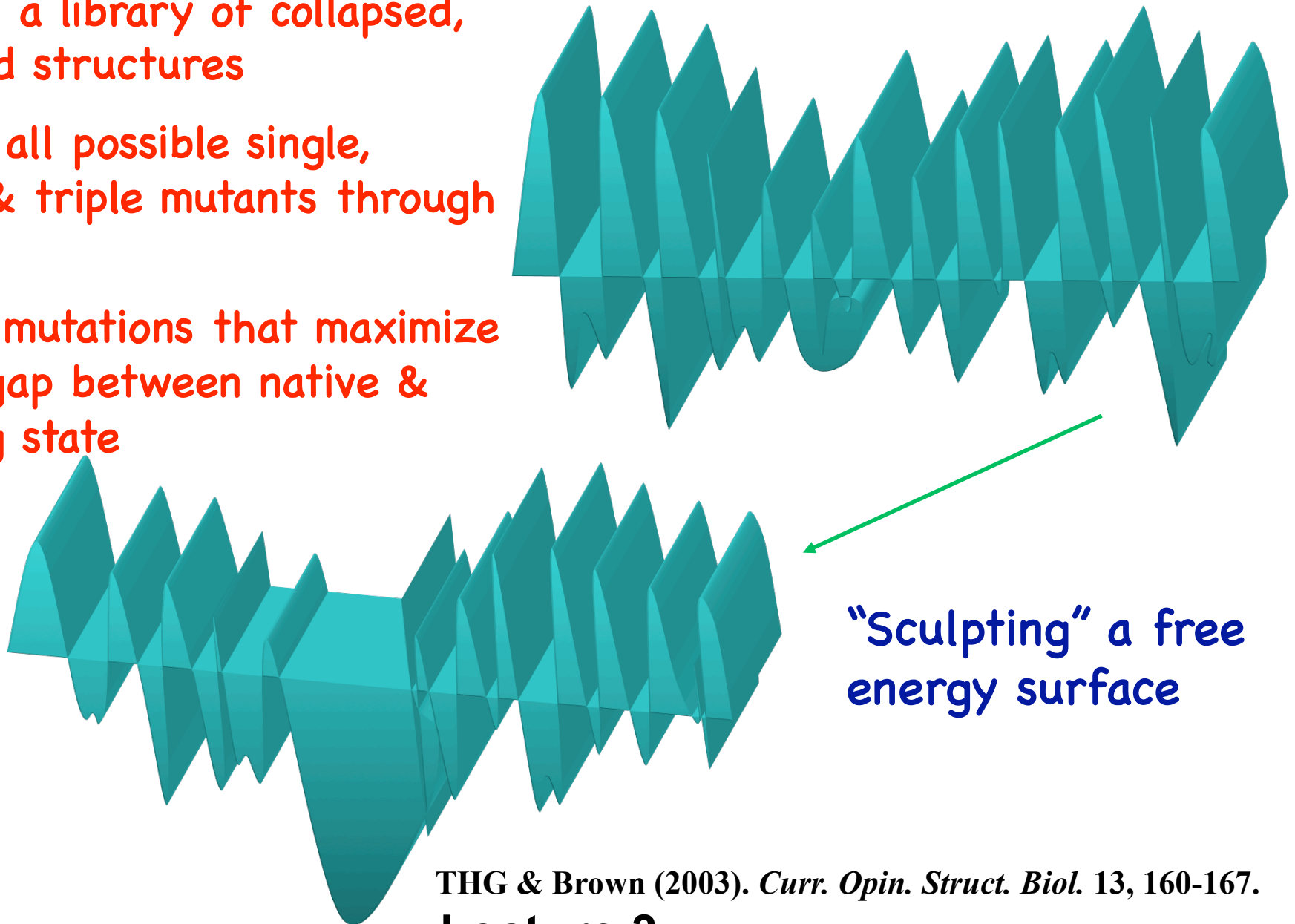


Long-lived traps of compact, non-native conformations with similar energetics

We use principle of minimal frustration to design a sequence close to the native sequence with a “folding funnel”

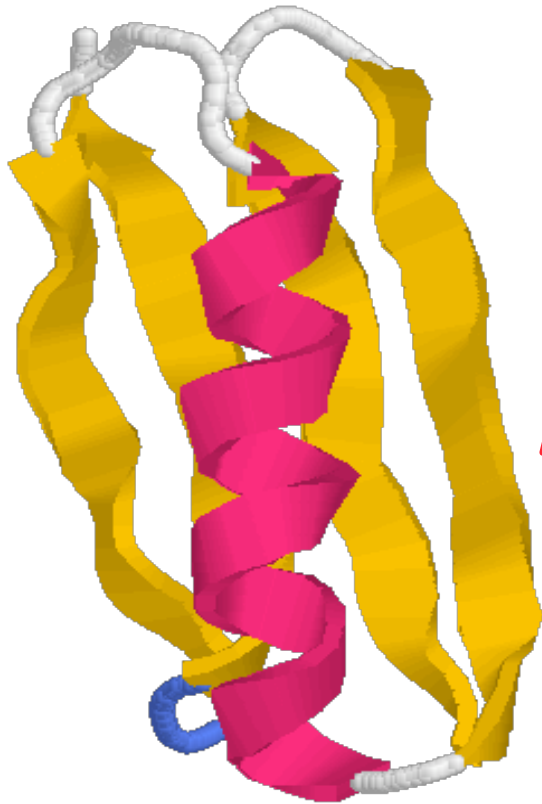
Principle of Minimal Frustration

- Develop a library of collapsed, misfolded structures
- Thread all possible single, double, & triple mutants through library
- Choose mutations that maximize energy gap between native & low lying state
- Repeat

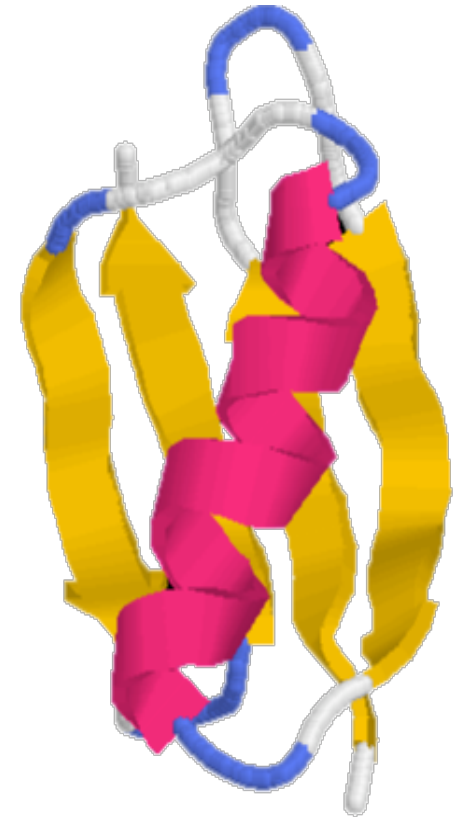


THG & Brown (2003). *Curr. Opin. Struct. Biol.* 13, 160-167.

IgG-Binding Protein's L & G



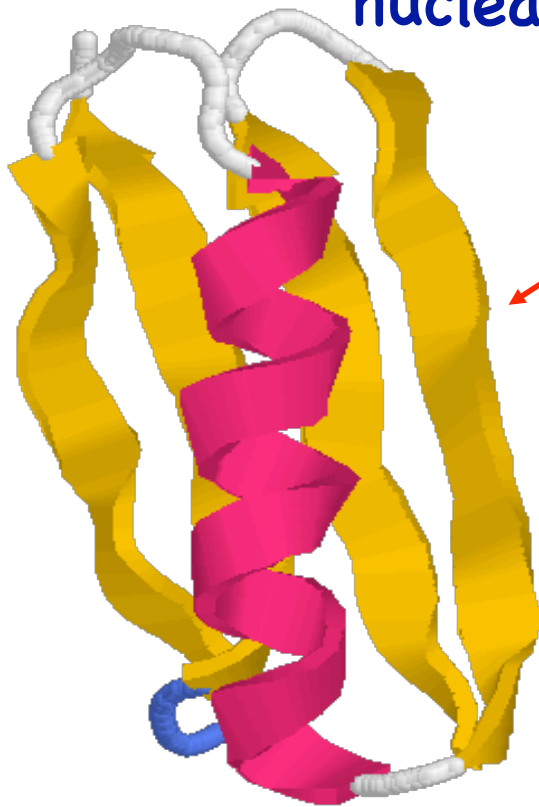
While Protein L and G have nearly identical tertiary fold:



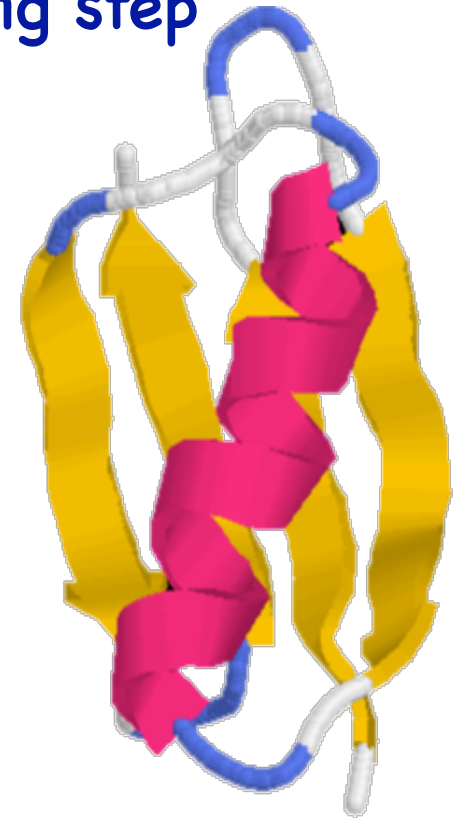
IgG-Binding Protein's L & G

Protein L:

two-state with a single barrier (Yi & Baker, 1996)
nucleating 1st β -hairpin is rate limiting step



While Protein L and G have nearly identical tertiary fold:



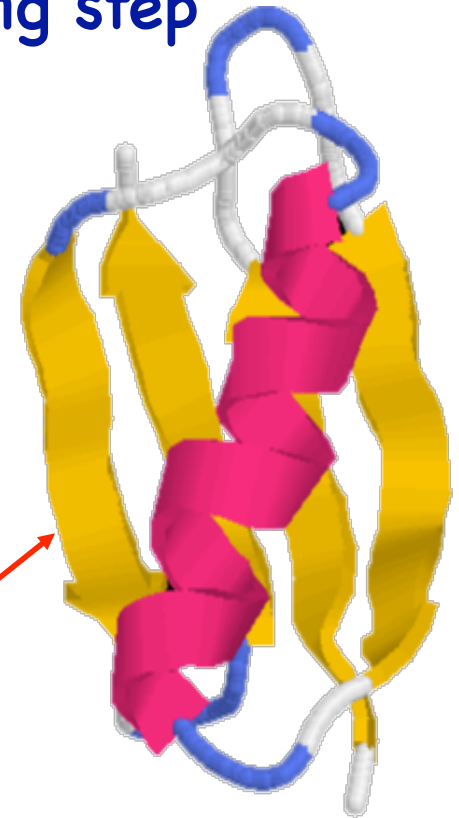
IgG-Binding Protein's L & G

Protein L:

two-state with a single barrier (Yi & Baker, 1996)
nucleating 1st β -hairpin is rate limiting step



While Protein L and G have nearly identical tertiary fold:



Protein G:

intermediate along folding pathway (Roder et al., 1997)
nucleating 2nd β -hairpin is rate limiting step

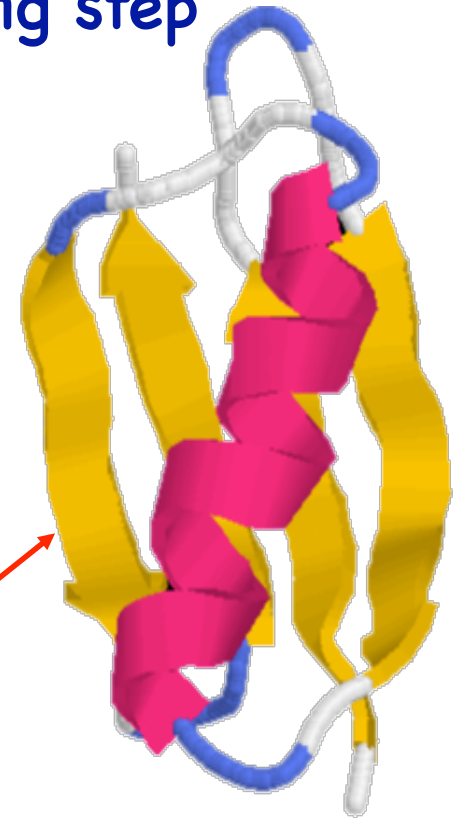
IgG-Binding Protein's L & G

Protein L:

two-state with a single barrier (Yi & Baker, 1996)
nucleating 1st β -hairpin is rate limiting step



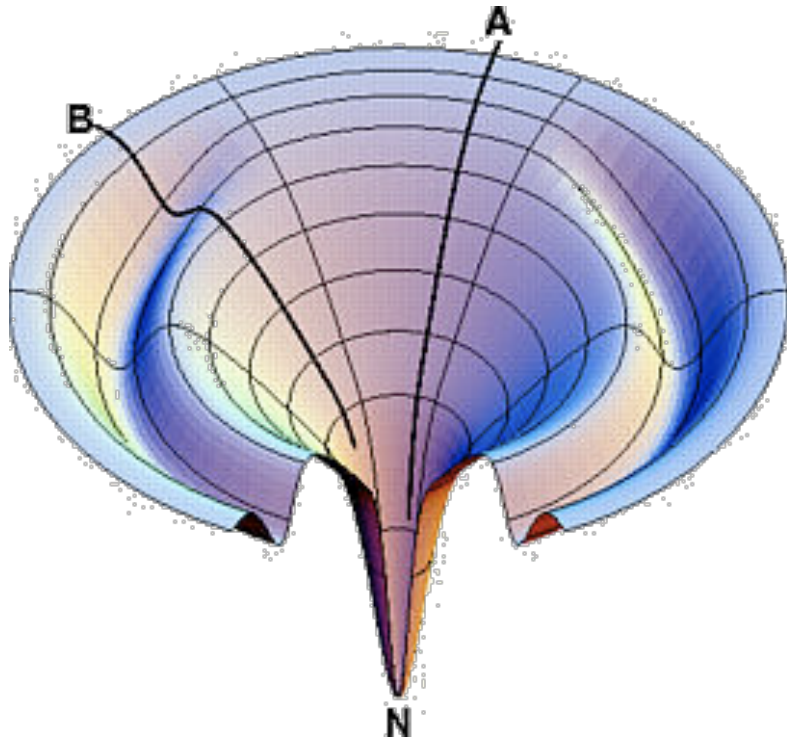
Many structurally homologous proteins have different folding mechanisms that may be informative about function/interaction partners in the cell



Protein G:

intermediate along folding pathway (Roder et al., 1997)
nucleating 2nd β -hairpin is rate limiting step

Early Folding Intermediates or Not?

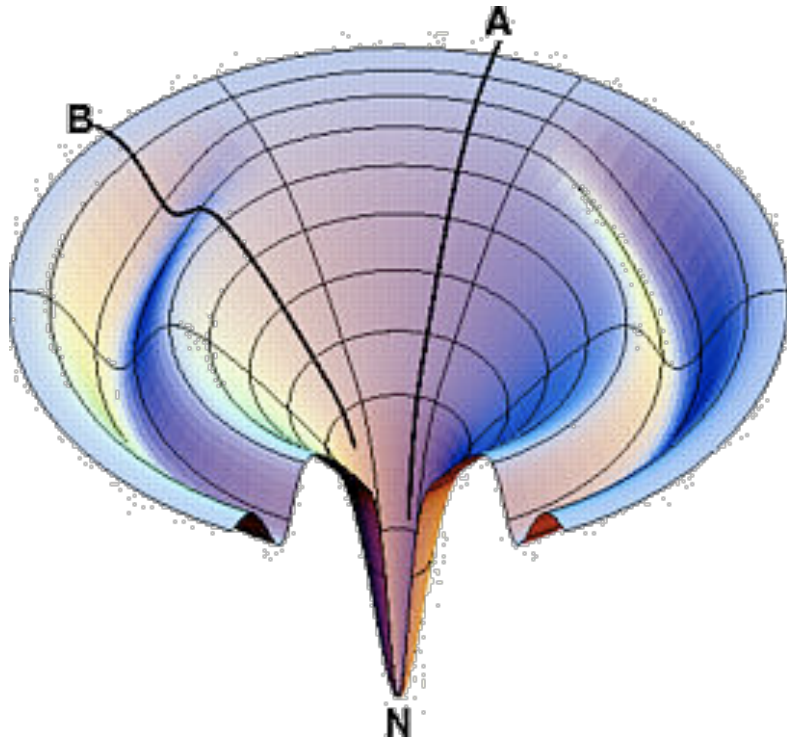


Pathway A: $U \rightarrow N$

Baker, *et al.*, *Nat. Struct. Biol.*, 2000

Englander, Sosnick & co-workers, 2002

Early Folding Intermediates or Not?



Pathway A: $U \rightarrow N$

Baker, et al., Nat. Struct. Biol., 2000

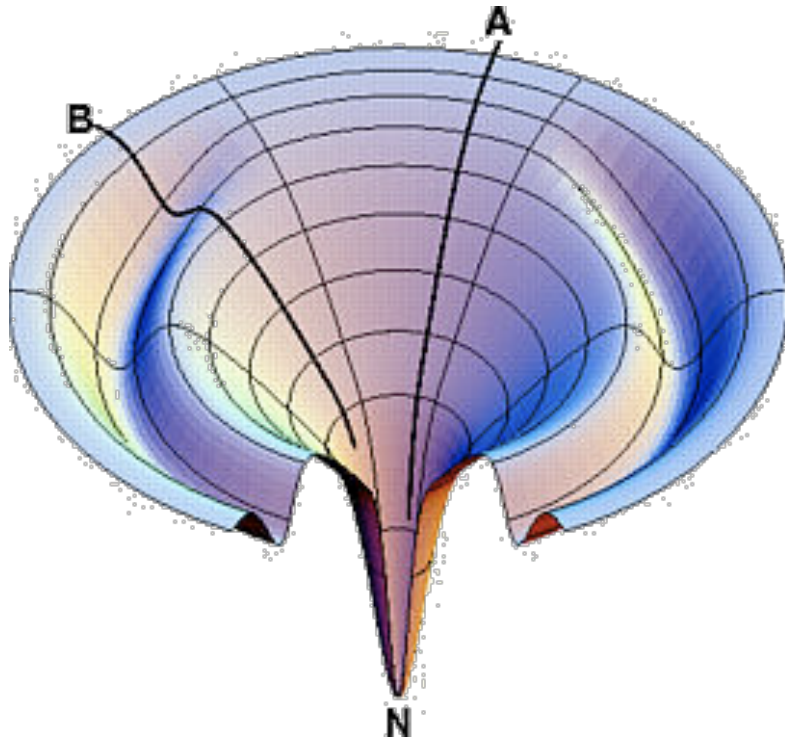
Englander, Sosnick & co-workers, 2002

Pathway B: $U \rightarrow I \rightarrow N$

Roder, et al., Nat. Struct. Biol., 1999

Gruebele and co-workers, 2002

Early Folding Intermediates or Not?



Pathway A: $U \rightarrow N$

Baker, et al., Nat. Struct. Biol., 2000

Englander, Sosnick & co-workers, 2002

Pathway B: $U \rightarrow I \rightarrow N$

Roder, et al., Nat. Struct. Biol., 1999

Gruebele and co-workers, 2002

In case of protein G, may involve nature of the “burst phase”

Is burst phase a downhill process?

Or a free energy barrier consistent with intermediate?

Simulation Protocol for Folding

Low-friction Langevin dynamics

friction coefficient: $0.05\tau^{-1}$ time step of 0.005τ

Thermodynamics:

multi-dimensional histograms at 15 different T (0.38–1.2)

E, Rg, contact order parameters $Q_{\alpha\beta 1}$, $Q_{\alpha\beta 2}$, etc

$$\chi = \frac{1}{M} \sum_{i, j \geq i+4}^K \theta(\varepsilon - |r_{ij} - r_{ij}^{Nat}|) \quad \chi_H, \chi_{\beta 1}, \chi_{\beta 2}, \chi_{\beta\alpha\beta}, \text{ etc}$$

3 independent trajectories at each temperature

10,000 data points per trajectory

Kinetics from Mean first passage times

~1000 folding trajectories

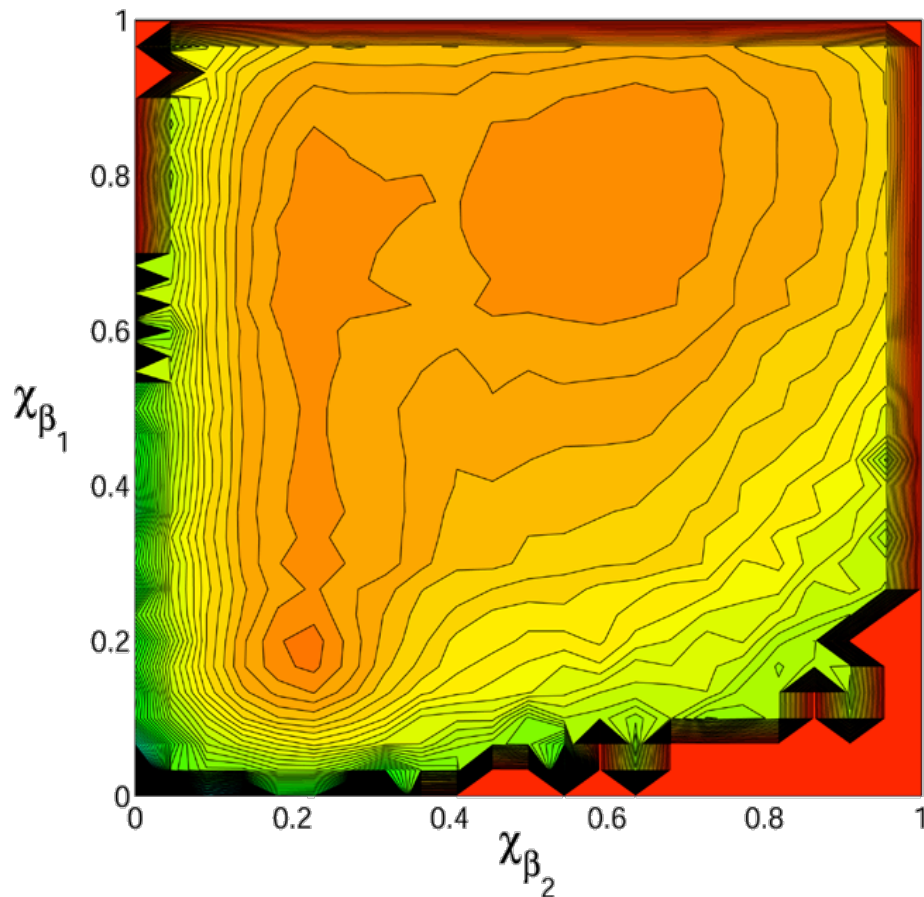
Pfold: determining transition state ensemble chain contacts

~10 trajectories per putative TSE member

Protein L: Thermodynamics and Kinetics

Protein L sequence consistent with experiment:

- path to native state through β_1 , with β_2 disrupted

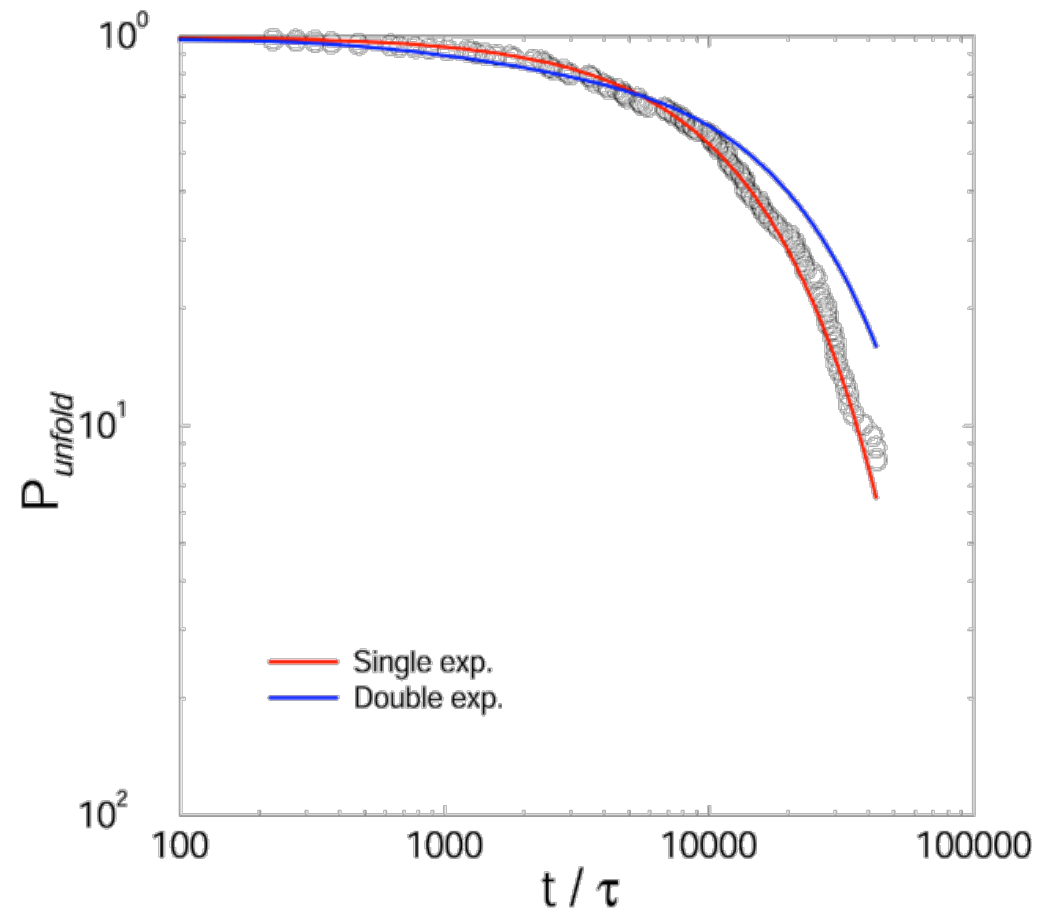
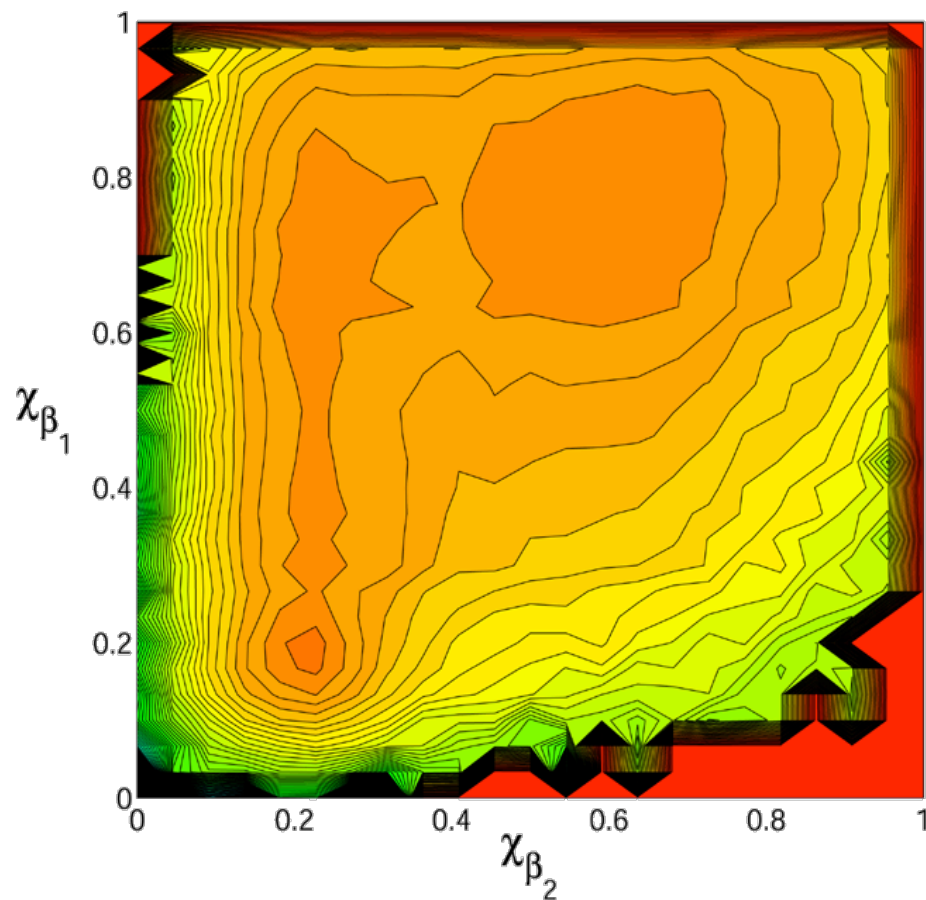


Brown & Head-Gordon (2004). *Protein Science* 13, 958-970

Protein L: Thermodynamics and Kinetics

Protein L sequence consistent with experiment:

- path to native state through β_1 , with β_2 disrupted
- single exponential kinetics

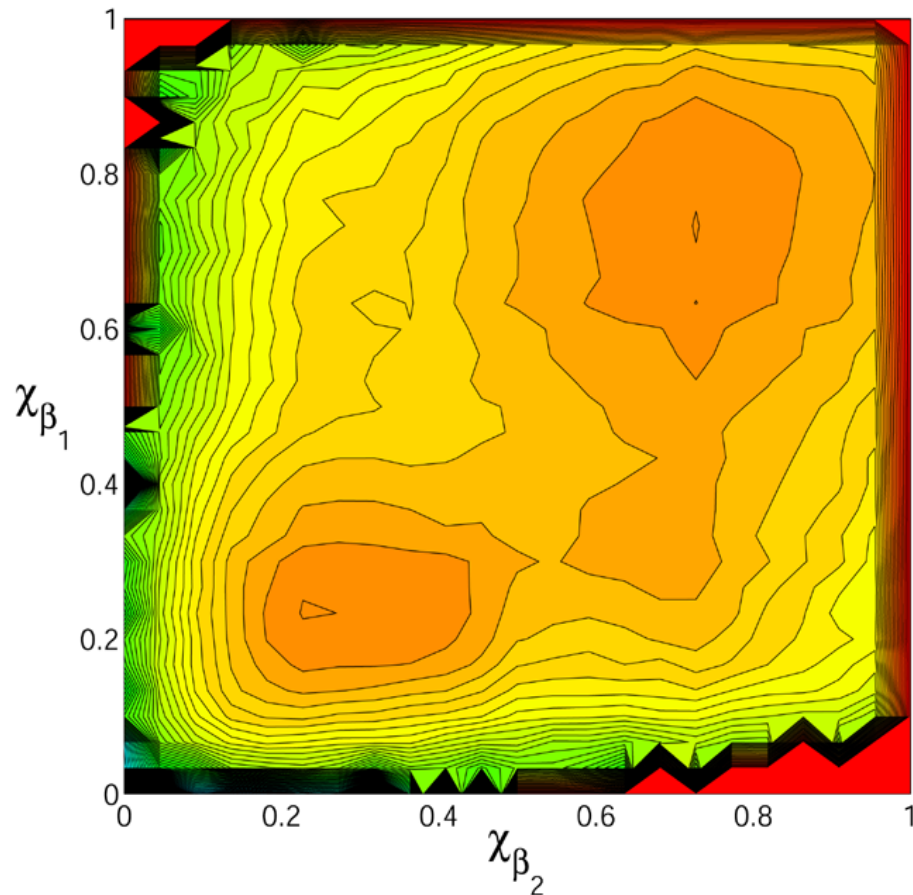


Brown & Head-Gordon (2004). *Protein Science* 13, 958-970

Protein G: Thermodynamics and Kinetics

Protein G sequence reflects:

➤ folding through β_2 region, with β_1 region disrupted

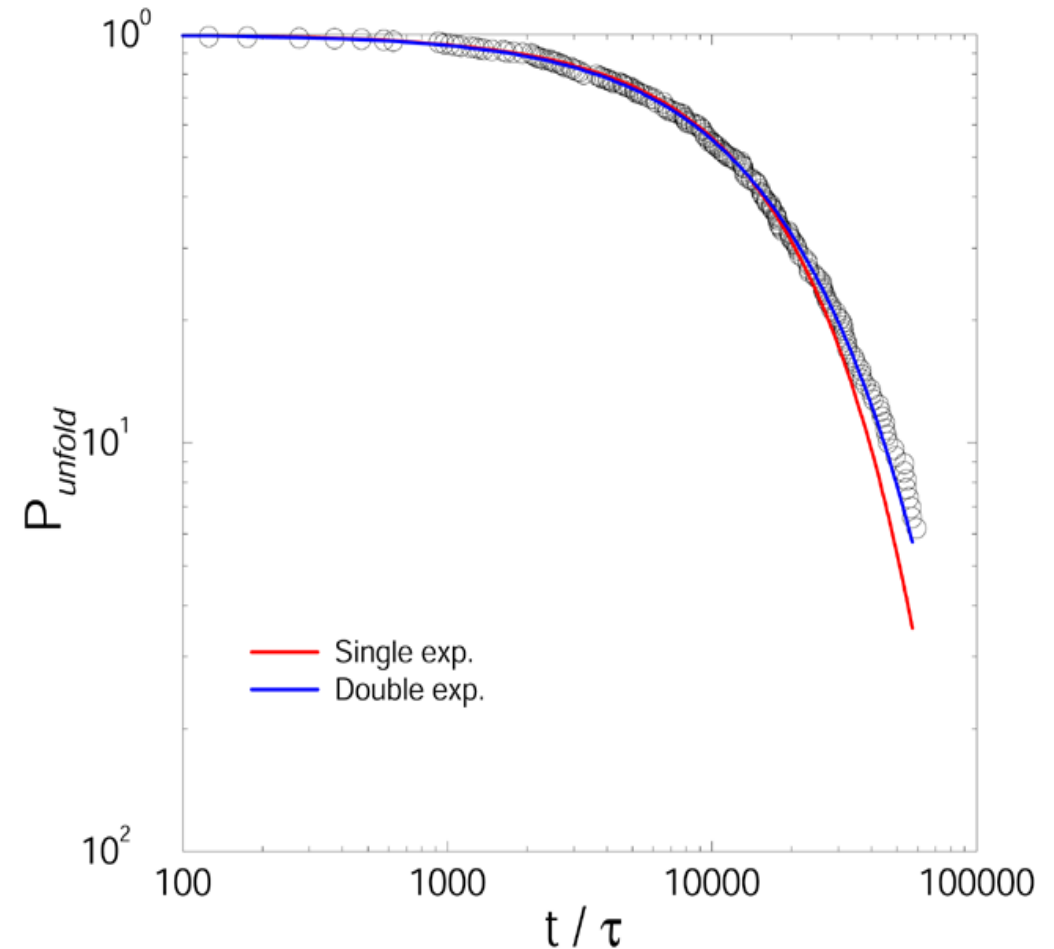
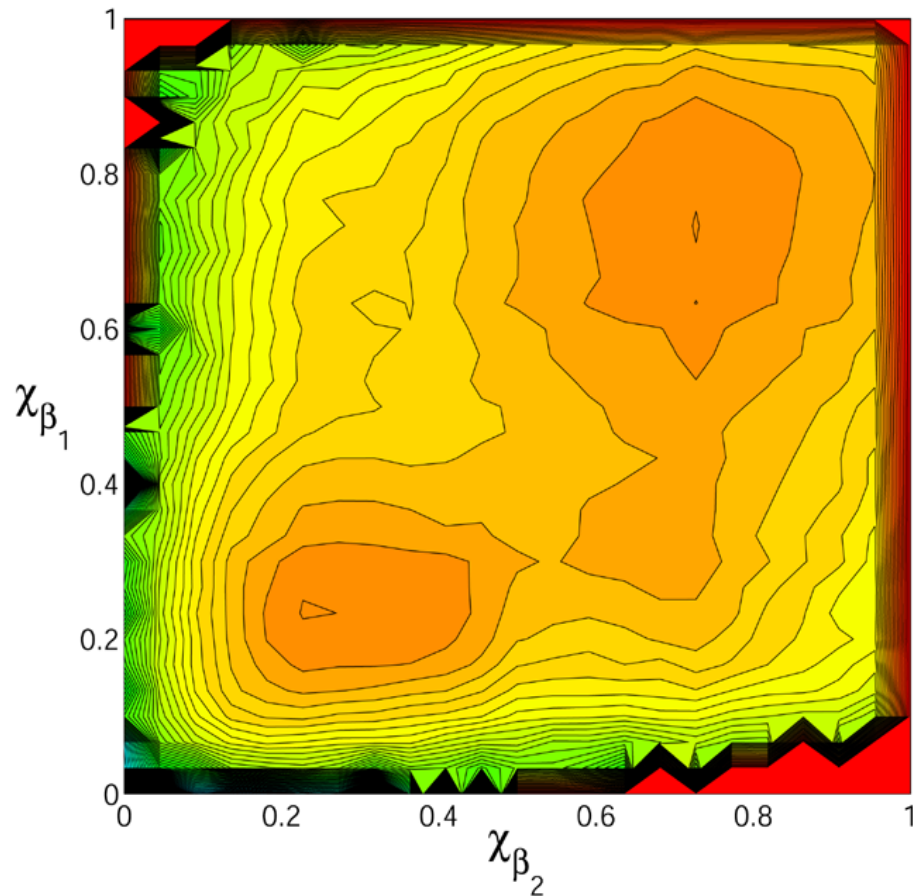


Brown & Head-Gordon (2004). *Protein Science* 13, 958-970

Protein G: Thermodynamics and Kinetics

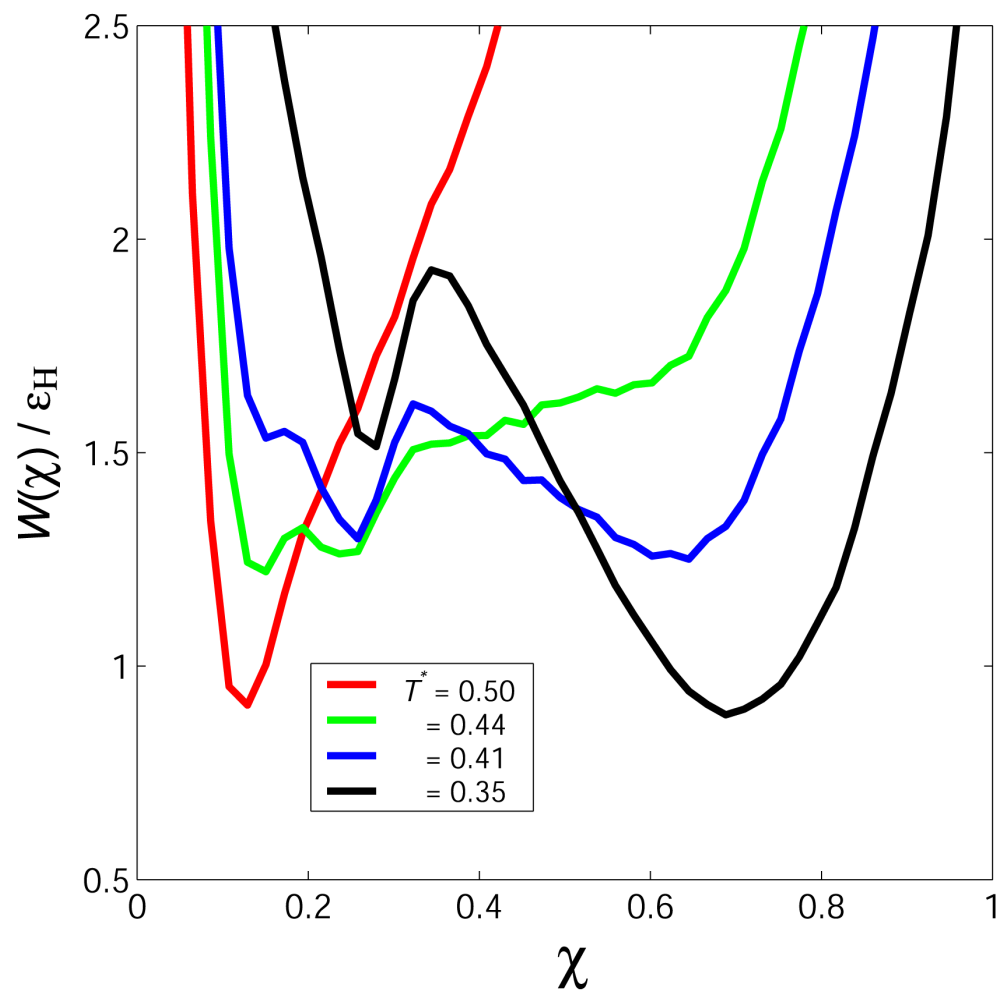
Protein G sequence reflects:

- folding through β_2 region, with β_1 region disrupted
- double exponential kinetics



Brown & Head-Gordon (2004). *Protein Science* 13, 958-970

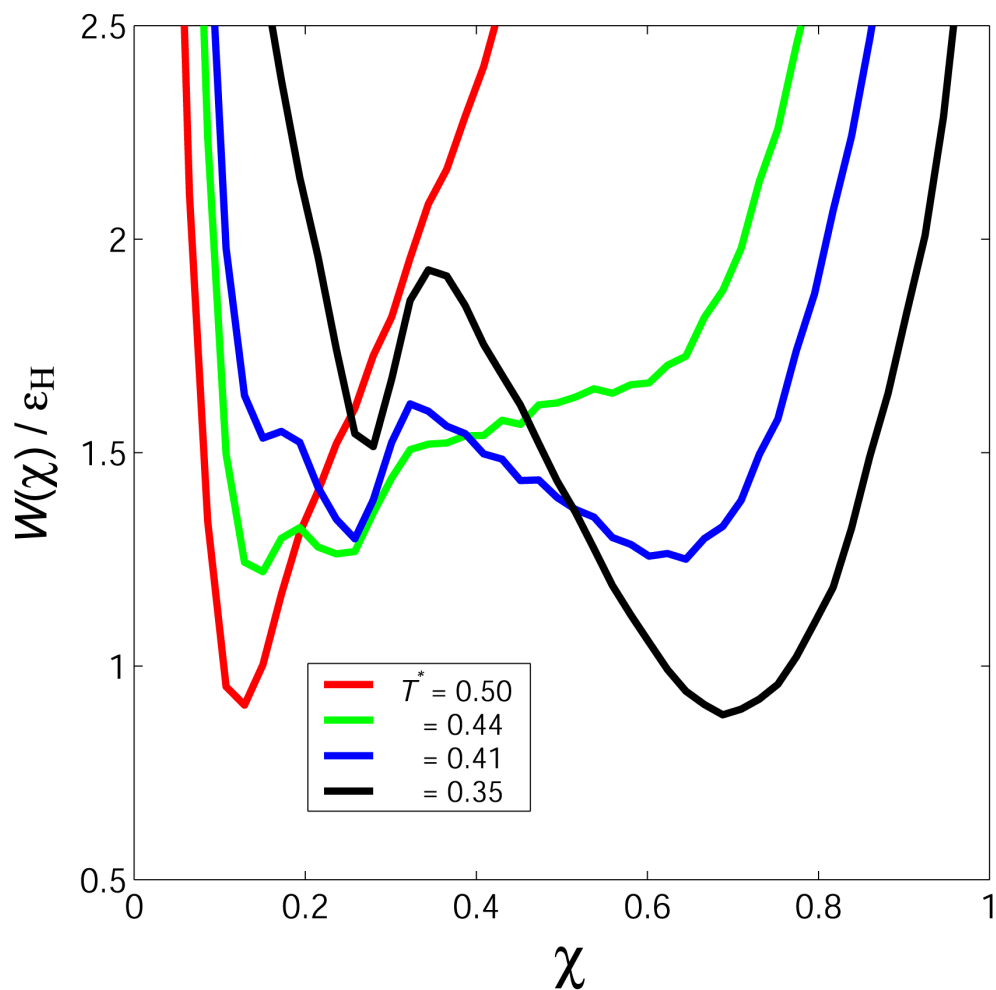
Folding Reaction Coordinates



Potential of mean force vs. native state similarity as a function of temperature for protein G. The folding temperature is $T_f = 0.41$.

Brown & Head-Gordon (2004). *Protein Science* 13, 958-970

Folding Reaction Coordinates



Potential of mean force vs. native state similarity as a function of temperature for protein G. The folding temperature is $T_f = 0.41$.

Based on this projection we might conclude that there is a shift in the unfolded population as we approach folding conditions

Brown & Head-Gordon (2004). *Protein Science* 13, 958-970

Transition State Ensembles for L and G

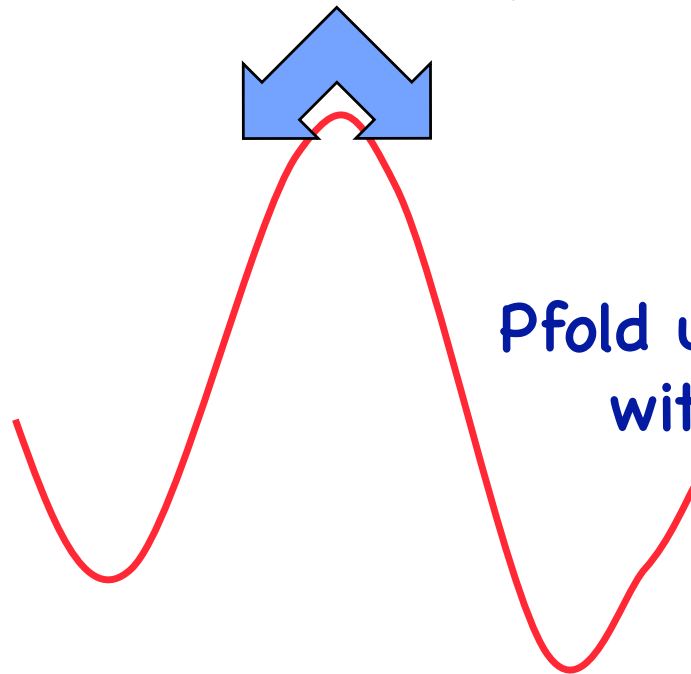
Considered contact order parameters α , β_1 , β_2 , $\beta_{2\alpha}$, $\beta_{3\alpha}$, $\beta_1\beta_2$, $\beta_1\beta_2\alpha$, $\beta_1\beta_4$, $\beta_2\beta_3$, $\beta_3\beta_4$, $\beta_3\beta_4\alpha$, $\beta_2\beta_3\alpha$, $\beta_1\beta_2\beta_3$, as well as a “diffuse” order parameter that was an expanded native state.

These were trapped as putative transition states from the kinetic trajectories at T_f , and

Transition State Ensembles for L and G

Considered contact order parameters α , β_1 , β_2 , $\beta_{2\alpha}$, $\beta_{3\alpha}$, $\beta_{1\beta_2}$, $\beta_{1\beta_2\alpha}$, $\beta_{1\beta_4}$, $\beta_{2\beta_3}$, $\beta_{3\beta_4}$, $\beta_{3\beta_4\alpha}$, $\beta_{2\beta_3\alpha}$, $\beta_{1\beta_2\beta_3}$, as well as a “diffuse” order parameter that was an expanded native state.

These were trapped as putative transition states from the kinetic trajectories at T_f , and



Pfold used to find whether actual TSE with good statistical confidence

Brown & Head-Gordon (2004). *Protein Science* 13, 958-970

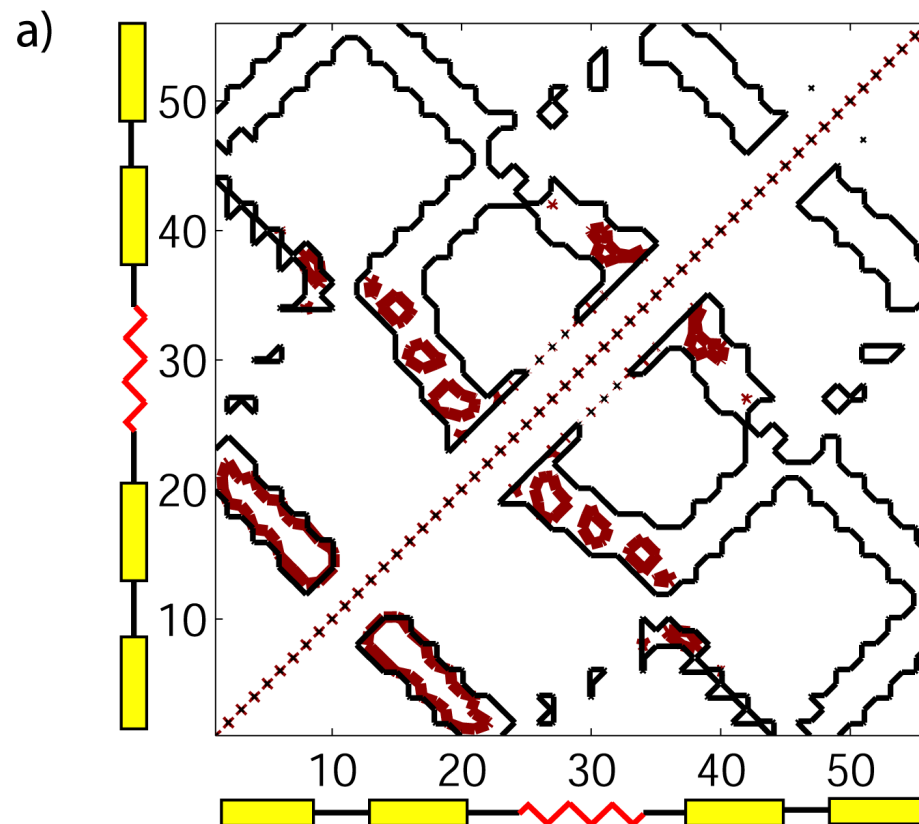
Transition State Ensemble for Protein L

Helix assisted β -hairpin1

$\beta_1\beta_2\alpha$

6:16 6:17 6:18 7:15 7:16 7:17 8:13 8:14 8:15 8:16 8:17 9:1
9:14 9:15 10:14 10:15 20:24 23:27 29:33 30:34

Contact map comparing native state (black) to contacts that are present for 90% or greater of TSE structures for folding of protein L (red)



Brown & Head-Gordon (2004). *Protein Science* 13, 958-970

Transition State Ensemble for Protein L

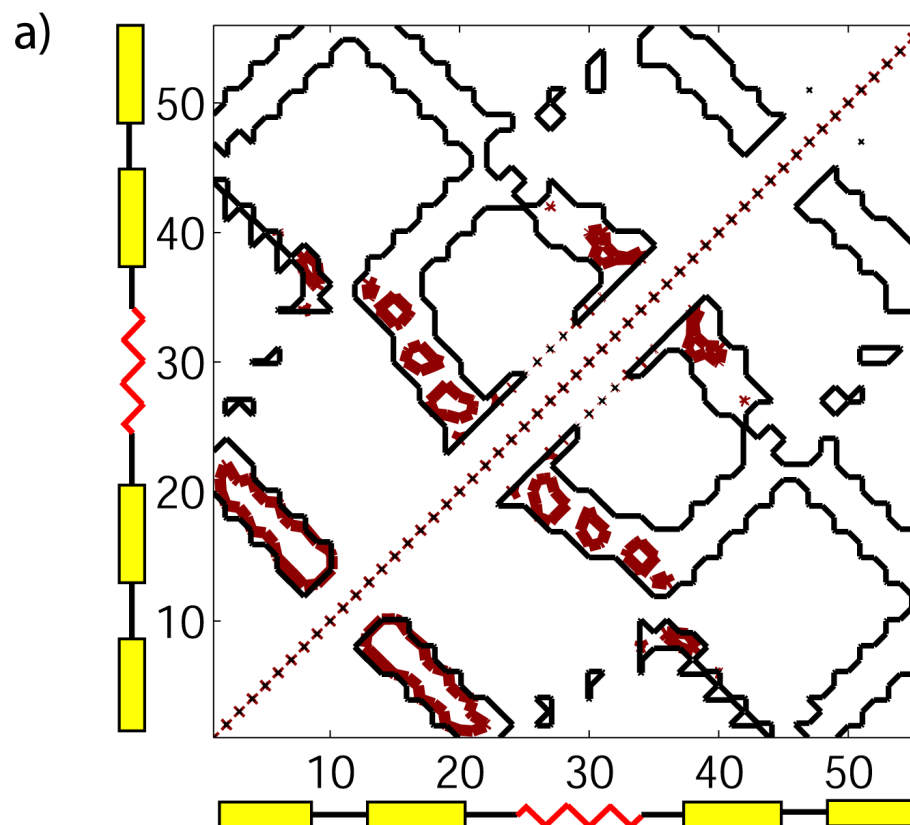
Helix assisted β -hairpin1

$\beta_1\beta_2\alpha$

6:16 6:17 6:18 7:15 7:16 7:17 8:13 8:14: 8:15 8:16 8:17 9:1
9:14 9:15 10:14 10:15 20:24 23:27 29:33 30:34

Contact map comparing native state (black) to contacts that are present for 90% or greater of TSE structures for folding of protein L (red).

TSE consistent with folding mechanism of formation of β -hairpin 1



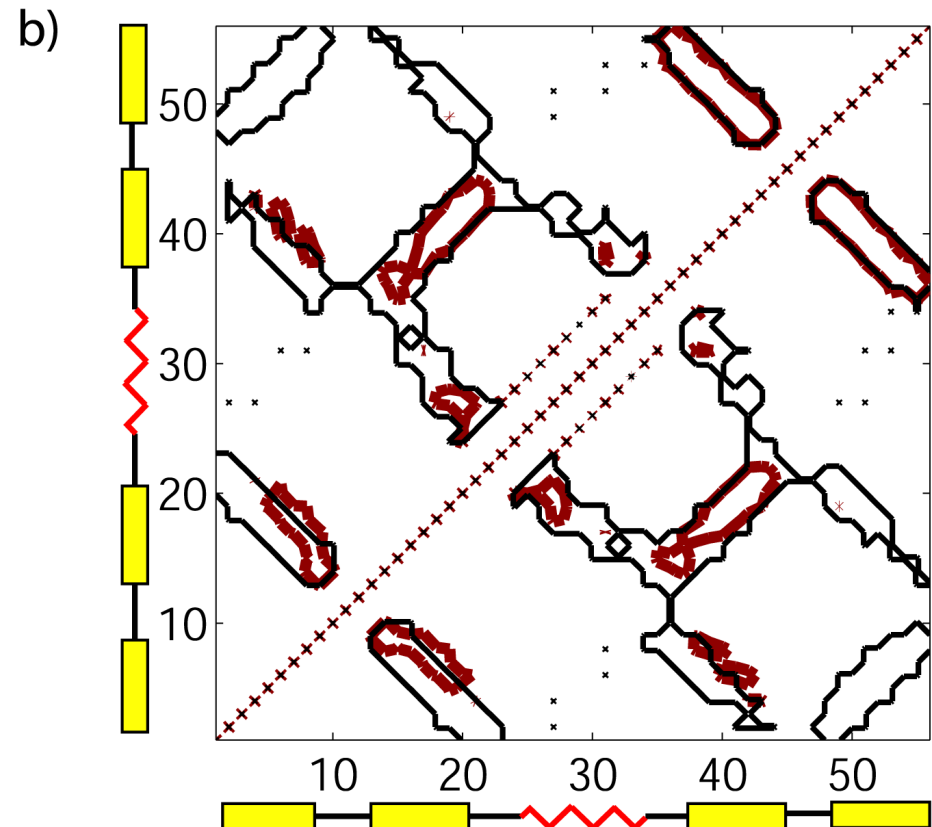
Brown & Head-Gordon (2004). *Protein Science* 13, 958-970

Transition State Ensemble for Fast Folding Pathway of Protein G

Helix assisted β -hairpin2

$\beta_3\beta_4\alpha$	10:14	20:24	20:27	23:27	24:28	27:31	36:53	36:54	36:55	37:53
	38:52	38:53	39:51	41:48	41:50	41:51	42:48	42:49	43:47	43:48
	43:49	44:48								

Contact map comparing native state (black) to contacts present for 90% or greater of TSE structures for fast folding pathway of protein G (red).



Transition State Ensemble for Fast Folding Pathway of Protein G

Helix assisted β -hairpin2

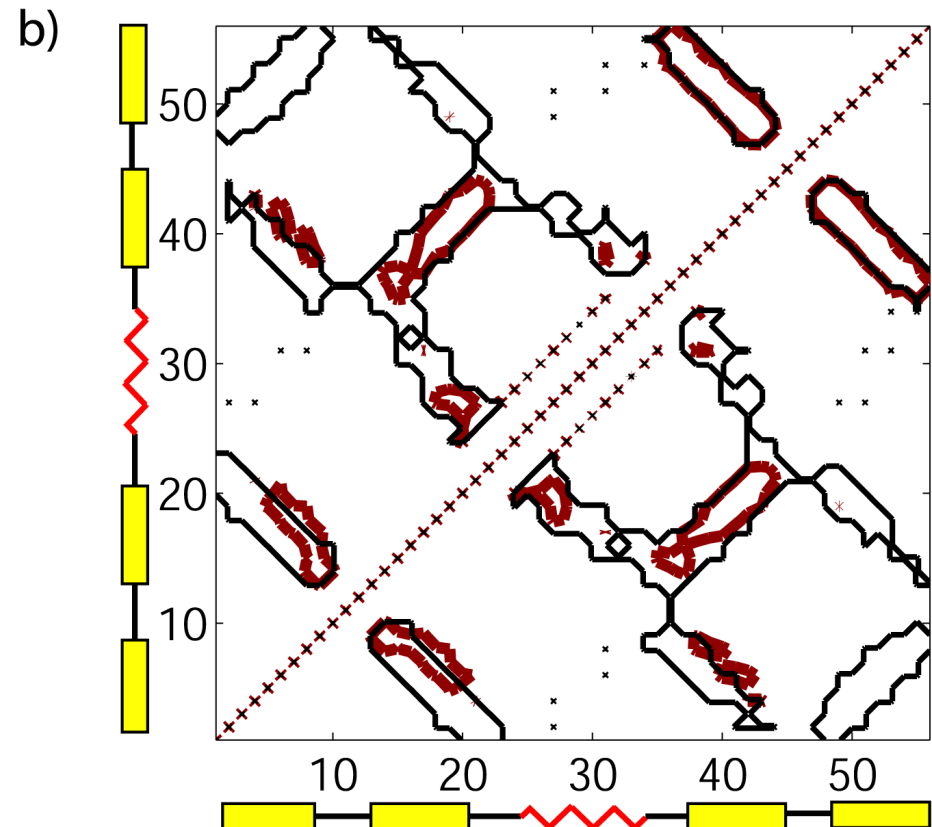
$\beta_3\beta_4\alpha$	10:14	20:24	20:27	23:27	24:28	27:31	36:53	36:54	36:55	37:53
	38:52	38:53	39:51	41:48	41:50	41:51	42:48	42:49	43:47	43:48
	43:49	44:48								

Contact map comparing native state (black) to contacts present for 90% or greater of TSE structures for fast folding pathway of protein G (red).

TSE consistent with experimental folding mechanism of formation of β -hairpin 2

Brown & Head-Gordon (2004). *Protein Science* 13, 958-970

Lecture 3

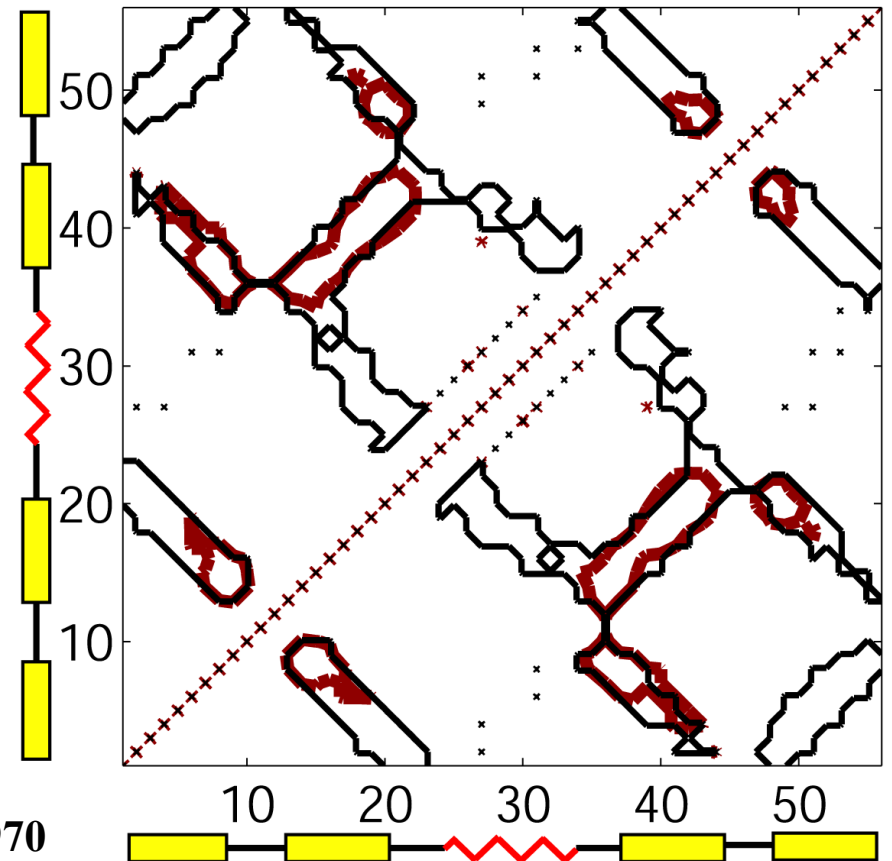


Transition State Ensemble for Slow Folding Pathway for Protein G

Protein G Transition State (slow pathway)

$\beta_1\beta_3\beta_2$	8:14 8:15 8:16 8:17 8:36 8:37 8:38 9:13 9:14 9:15 9:36 10:14
	14:36 15:36 17:38 19:40 19:41 19:42 20:41 20:42 20:49 21:41
	21:42 21:43 43:48

Contact map comparing native state (black) to contacts present for 90% or greater of TSE structures for slow folding pathway of protein G (red).



Brown & Head-Gordon (2004). *Protein Science* 13, 958-970

Determining the Intermediate State Ensemble for Slow Pathway of Protein G

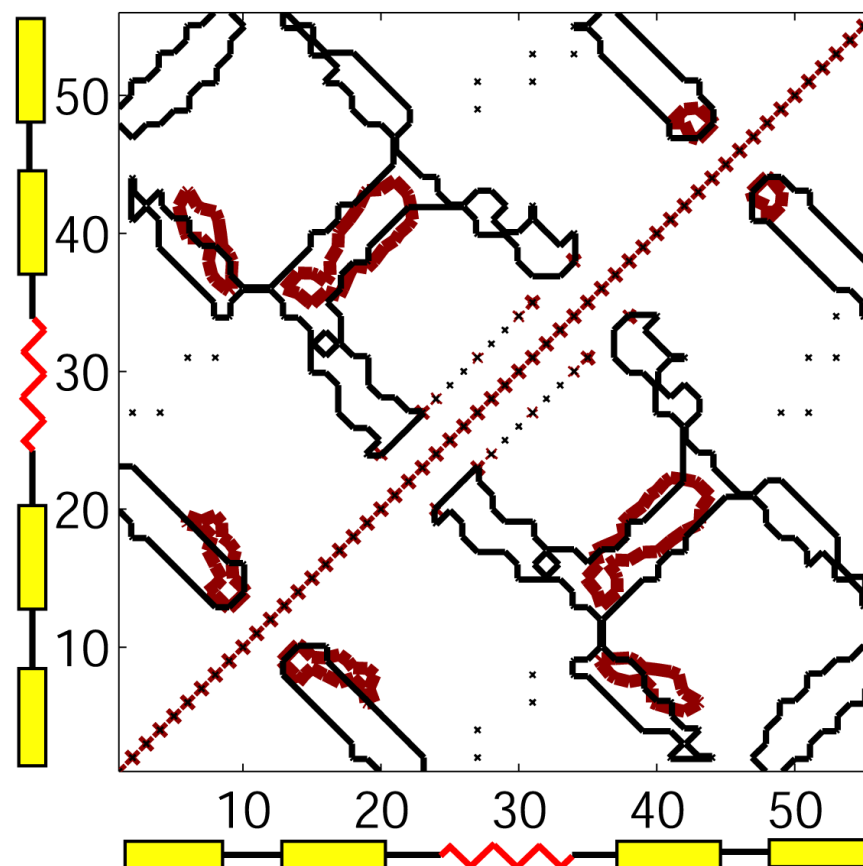
Protein G Intermediate: follow unfolding from TSE!

$\beta_2\beta_3\alpha$

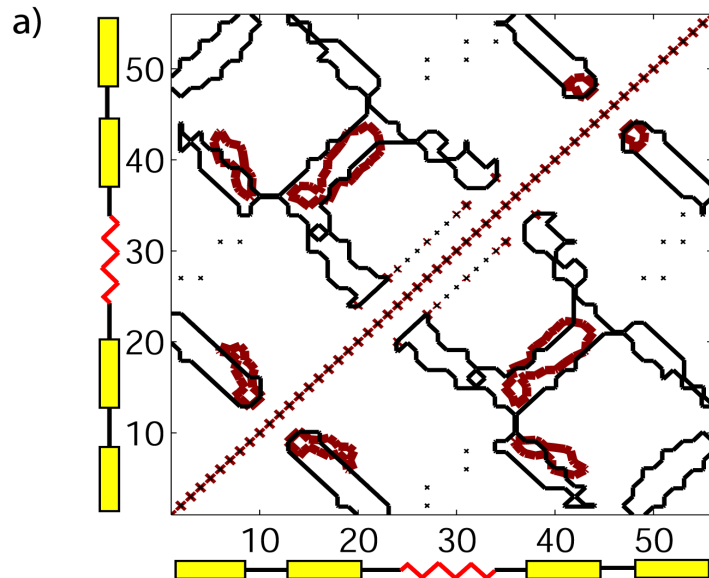
8:14 9:13 9:14 9:15 10:14 18:40 18:41 19:39 19:40 19:41
23:27 27:31 31:35 42:48 43:47 43:48 43:49 44:48

Contact map comparing native state (black) to contacts present for 90% or greater of ISE structures of slow folding protein G (red).

The intermediate is characterized by associated helix with β -strands 2 and 3 which are misaligned relative to the native state.



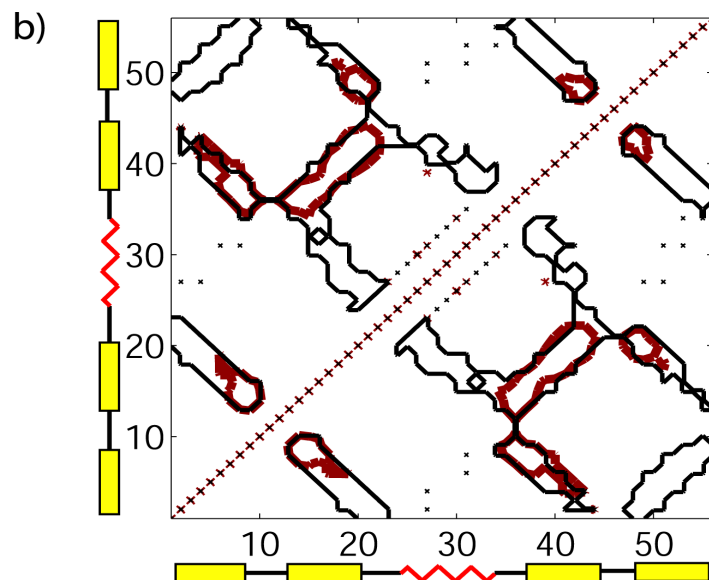
Nature of Intermediate for Protein G



Intermediate ensemble

Associated helix with β -strands 2 and 3 which are misaligned relative to the native state.

Brown & Head-Gordon (2004). *Protein Science* 13, 958-970



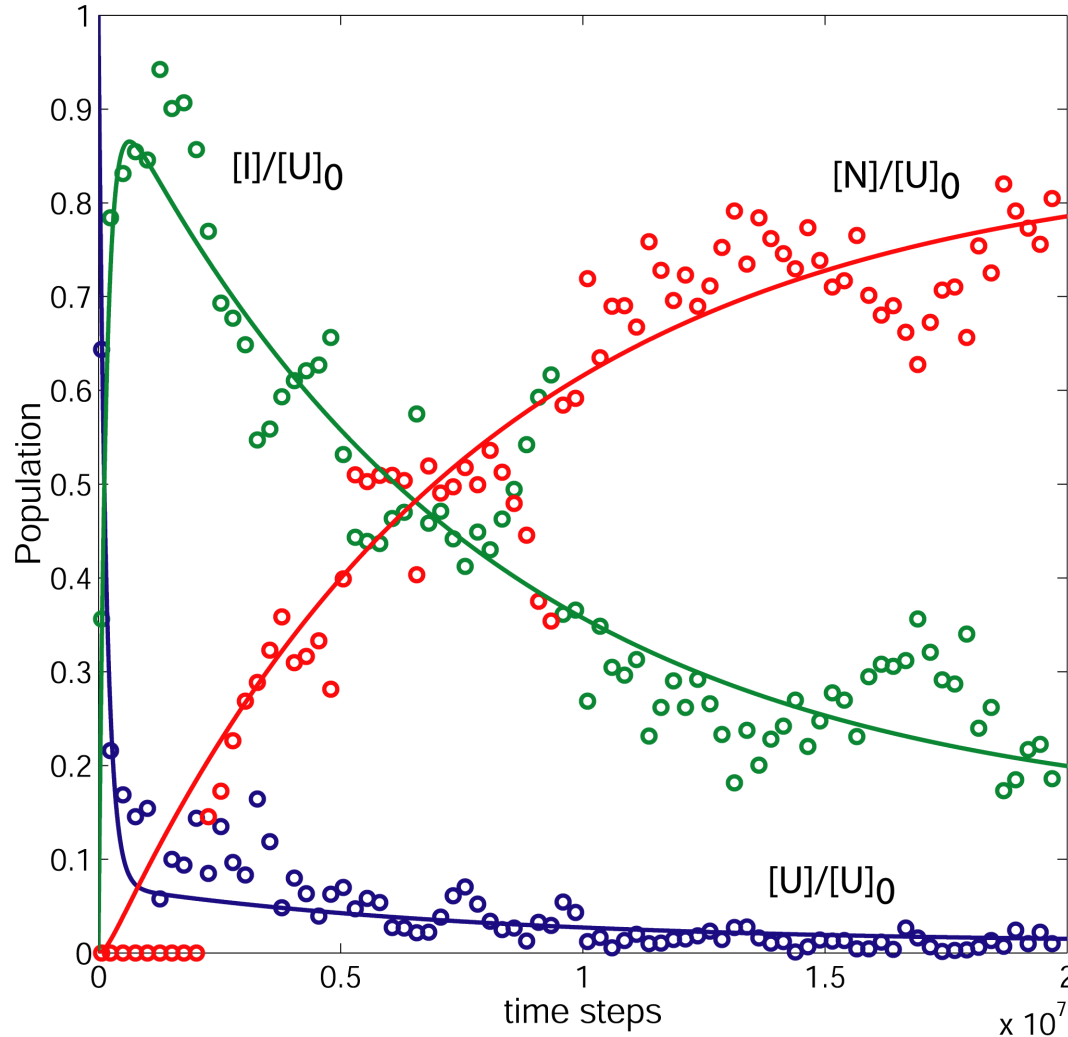
Transition-state ensemble

β -strands 1, 2, 3, with correct alignment for β -strands 2 and 3 relative to the native state.

Similar to that seen for Im7 and Im9

(Radford and co-workers, 2002)

Protein G Kinetics: [U] \longrightarrow [I] \longrightarrow [N]

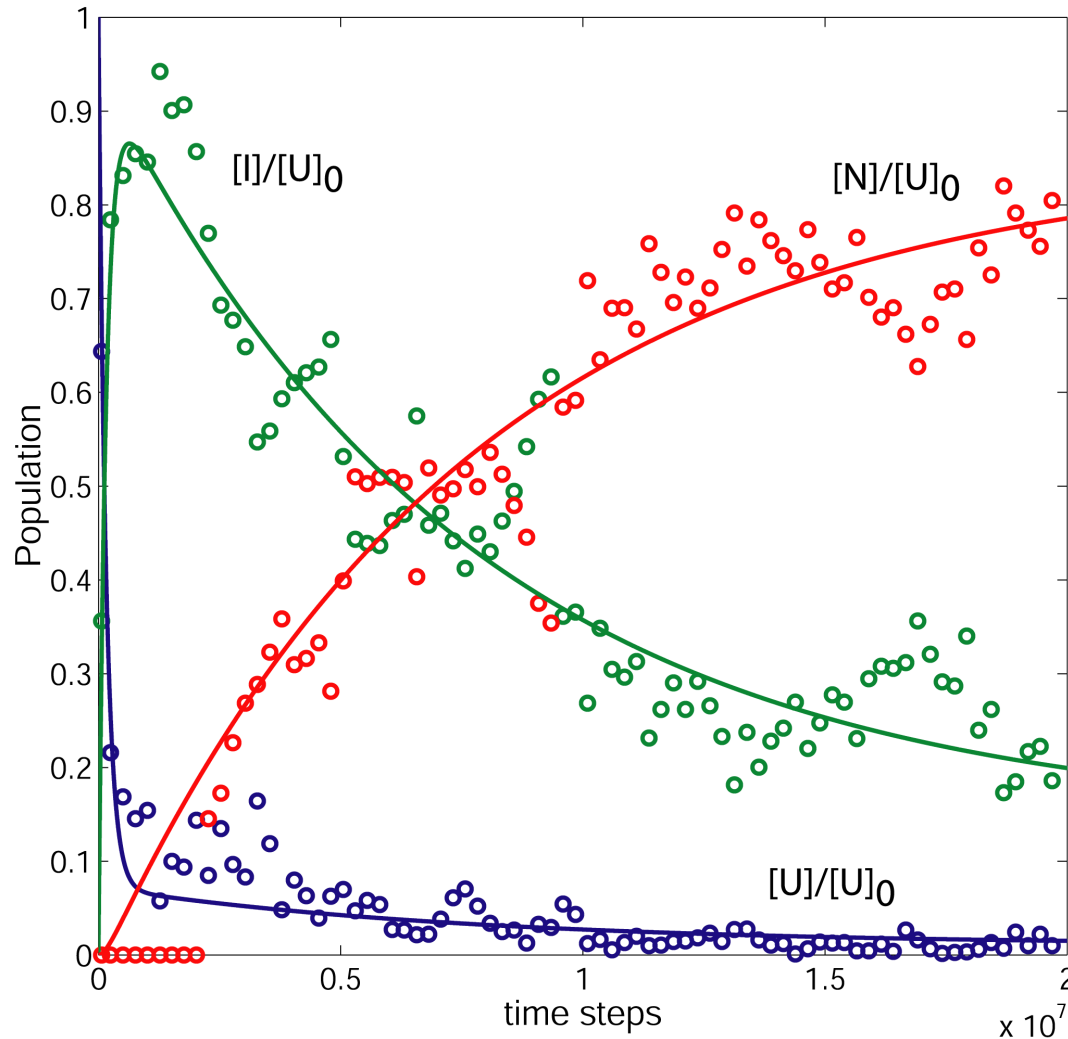


Collect populations of [U], [I], and [N] as a function of time, and fit data to two-step reversible mechanism:



Brown & Head-Gordon (2004). *Protein Science* 13, 958-970

Protein G Kinetics: [U] \longrightarrow [I] \longrightarrow [N]



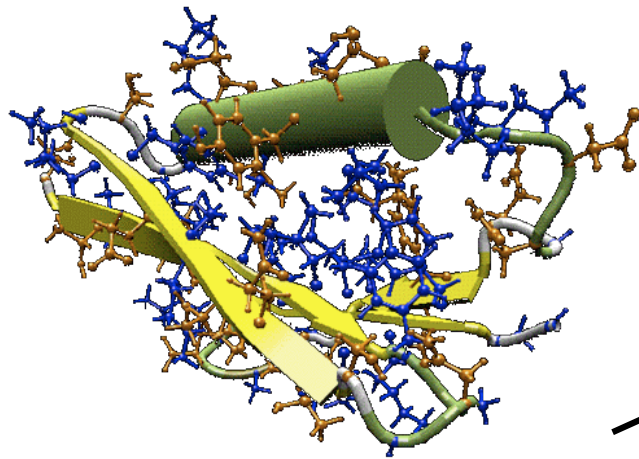
Collect populations of [U], [I], and [N] as a function of time, and fit data to two-step reversible mechanism:



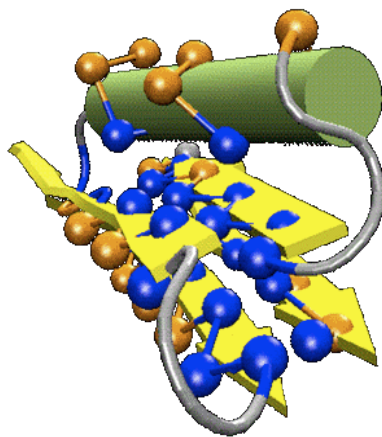
Protein G clearly folds through an intermediate and supports Roder et al

Brown & Head-Gordon (2004). *Protein Science* 13, 958-970

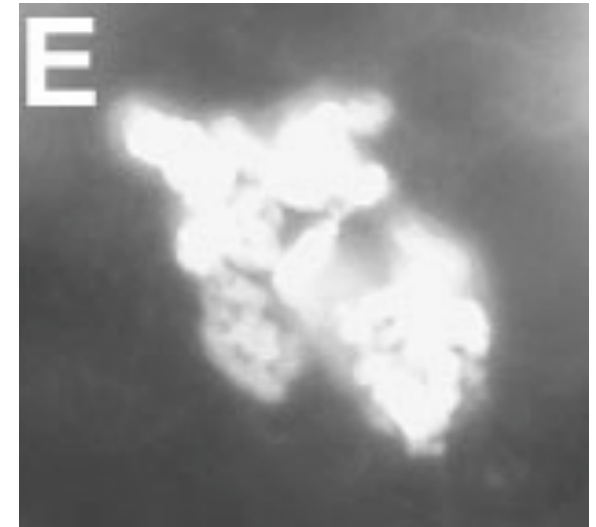
Protein Aggregation



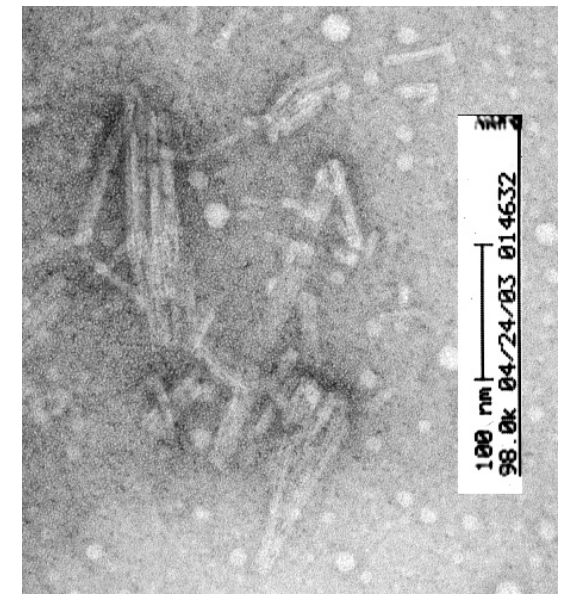
*Manufacturing of
Pharmaceutical
Proteins
(Biotech)*



*Alzheimers, Parkinson's,
Prion Diseases
(Health)*



<http://www.molvis.org/molvis/v8>



Alzheimer's Disease

~5% of total US population >65 years of age are diagnosed with Alzheimer's disease

~20% of population >80 years and ~30% >90 years are diagnosed with Alzheimer's disease.

Women have higher incidence than men (because women live longer on average)

A disease associated with an aging demographic in USA with grim statistics for ~2050 if no therapeutics arise

[http://www.nia.nih.gov/Alzheimers/
AlzheimersInformation/GeneralInfo/](http://www.nia.nih.gov/Alzheimers/AlzheimersInformation/GeneralInfo/)

Protein Aggregation and Disease



Fawzi, Okabe, Yap, THG (2007) J. Mol. Biol. 365, 535-550

Yap, Fawzi, THG (2008) Proteins 70, 626-638

Fawzi, Kohlstedt, Okabe, THG (2008) Biophys. J. 94 2007-16.

Fawzi, Yap, Okabe, Kohlstedt, Brown, THG (2008)

Acc. Chem. Res 41 1037-47

Fawzi, Ruscio, Phillips, Doucleff, Wemmer, THG (2008)

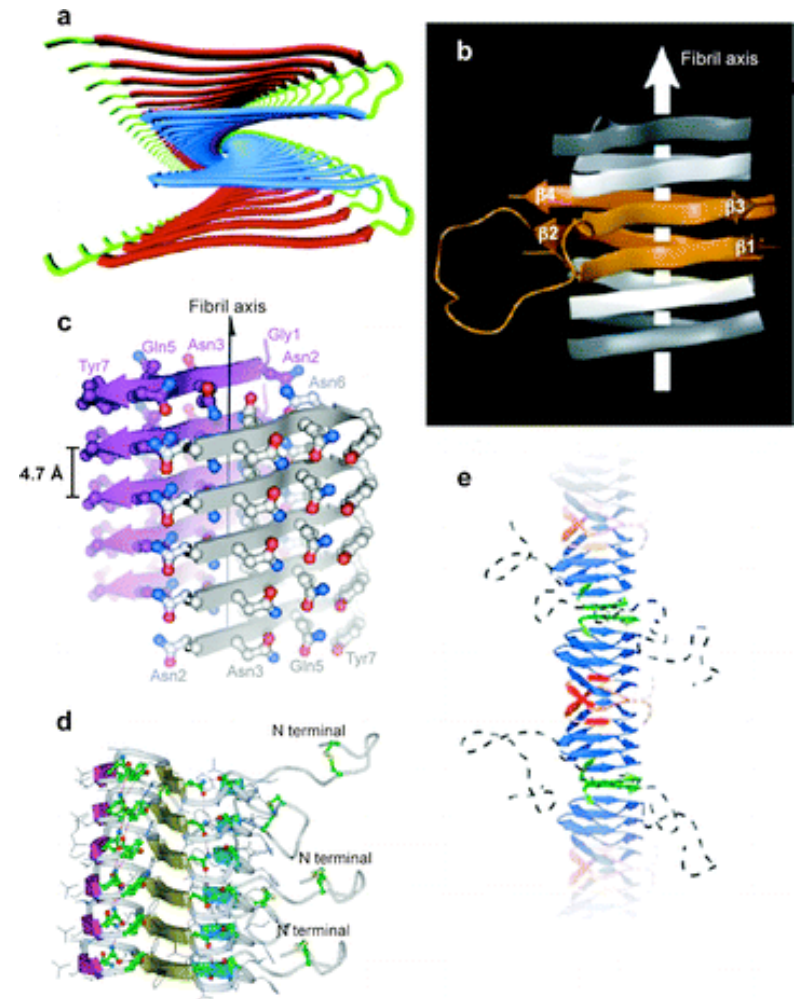
JACS 130, 6145-58


Sodt & THG (2009). Biophysical J. submitted



Amyloid Fibril Morphologies

Virchow in 1854 noted a macroscopic tissue abnormality that exhibited a positive iodine staining reaction, similar to starch or “amyloid”.

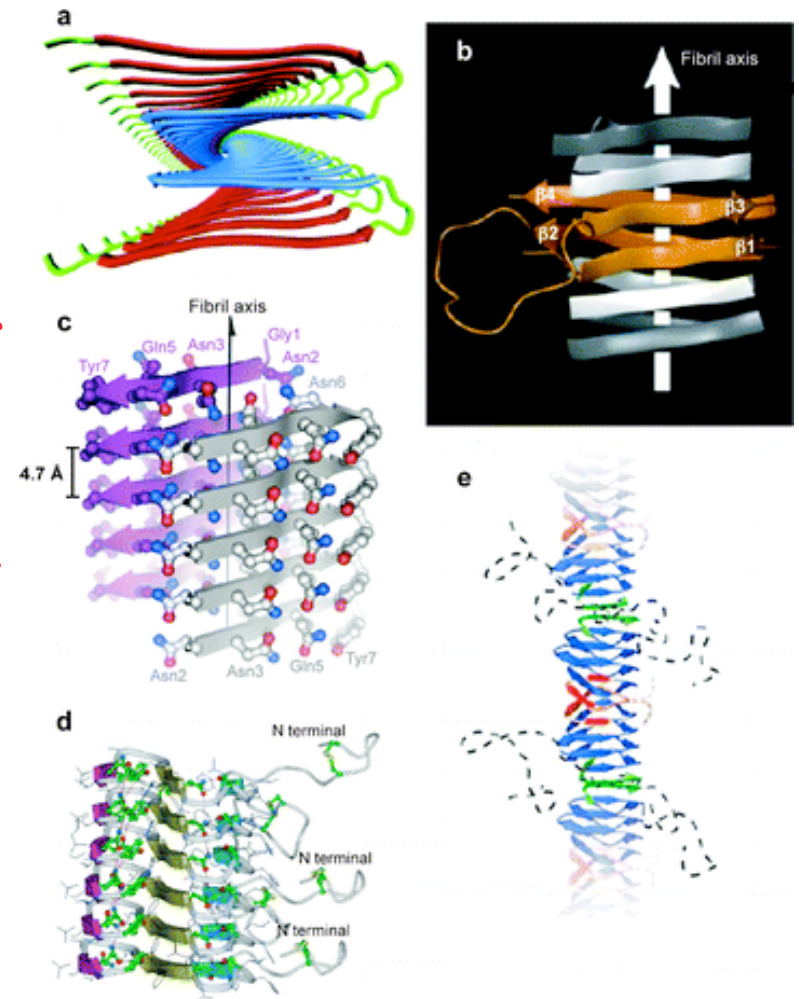


 Chiti F, Dobson CM. 2006.
Annu. Rev. Biochem. 75:333–66

Amyloid Fibril Morphologies

Virchow in 1854 noted a macroscopic tissue abnormality that exhibited a positive iodine staining reaction, similar to starch or “amyloid”.

Disease peptides or proteins that order into amyloid fibril morphologies bind specific dyes such as thioflavin T, allowing characterization with light microscopy and x-ray diffraction



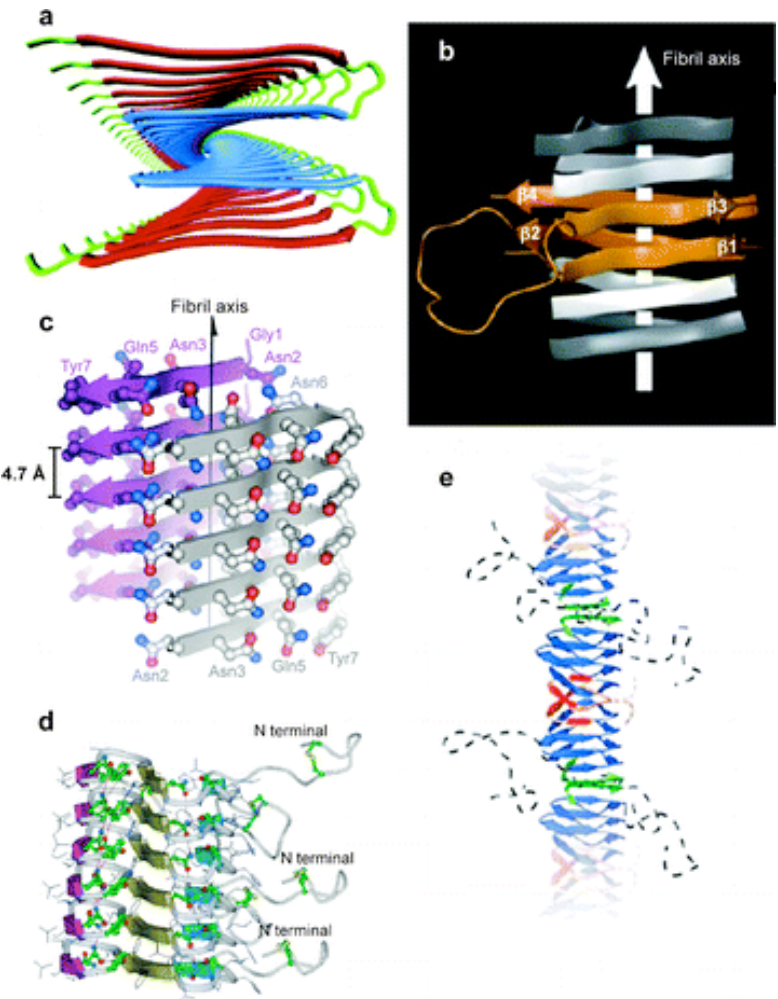
Chiti F, Dobson CM. 2006.
Annu. Rev. Biochem. 75:333–66

Amyloid Fibril Morphologies

Virchow in 1854 noted a macroscopic tissue abnormality that exhibited a positive iodine staining reaction, similar to starch or “amyloid”.

Disease peptides or proteins that order into amyloid fibril morphologies bind specific dyes such as thioflavin T, allowing characterization with light microscopy and x-ray diffraction

When characterized ex-vivo or in vitro, fibrils have characteristic “cross- β ” structure: intermolecular β -sheets (~10nm) run along fibril axis, stabilizing assemblies which can extend to microns in length.

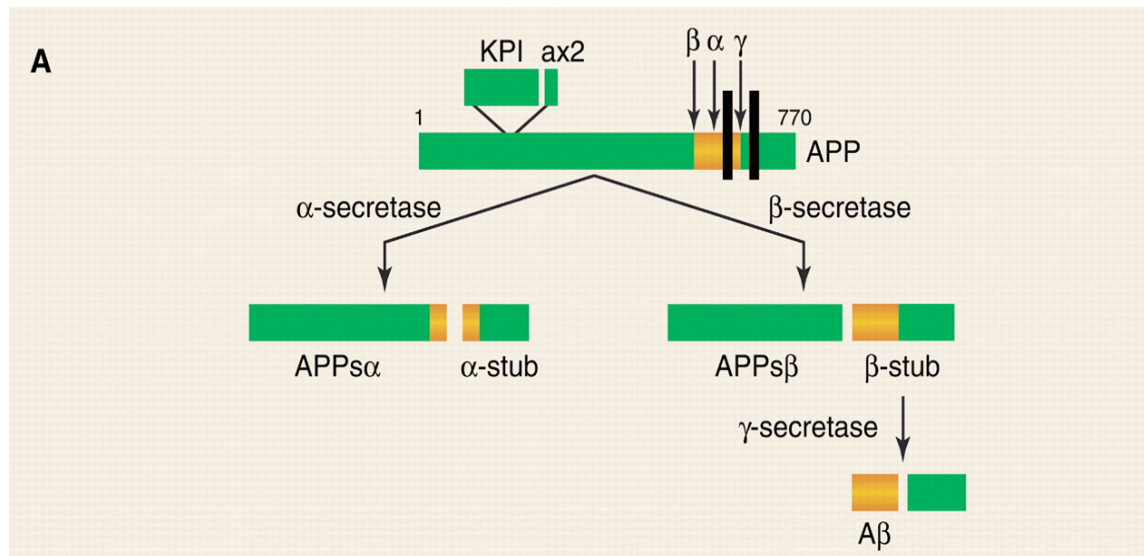


Chiti F, Dobson CM. 2006.
Annu. Rev. Biochem. 75:333–66

Alzheimer's Disease

Amyloid- β (function unclear) is generated by proteolytic cleavage of the transmembrane APP protein (function unknown)

β -secretase cleaves APP to delineate N-terminus



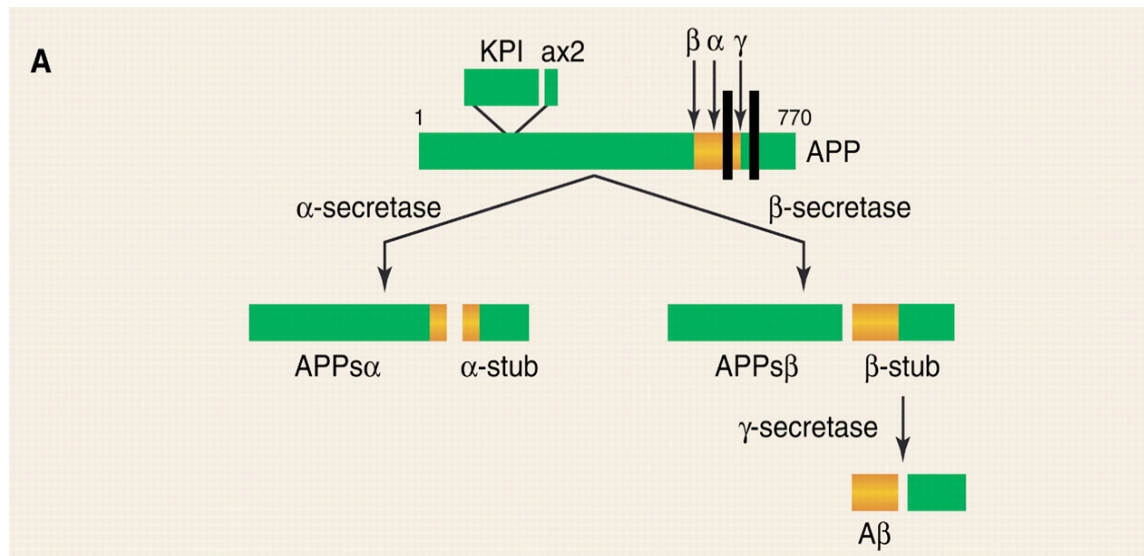
Goedert & Spillantini,
Science 2006

γ -secretase cleaves APP to delineate C-terminus

Alzheimer's Disease

Amyloid- β (function unclear) is generated by proteolytic cleavage of the transmembrane APP protein (function unknown)

β -secretase cleaves APP to delineate N-terminus



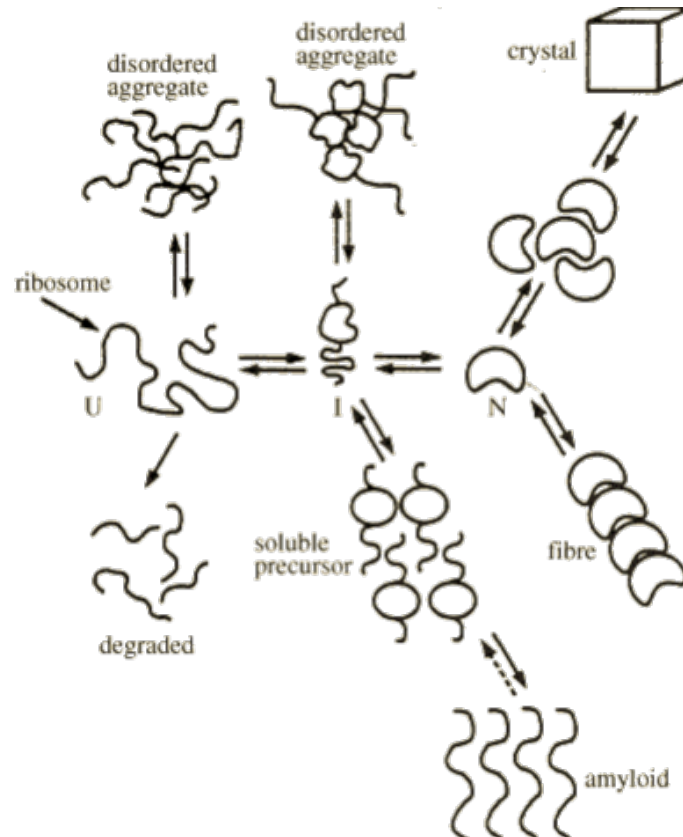
Goedert & Spillantini,
Science 2006

γ -secretase cleaves APP to delineate C-terminus

While early events involving A β are extremely unclear (ADDL vs. fibrils), tangle formation (tau protein), nerve cell degeneration, and dementia are downstream events.

Amyloid- β and Alzheimer's Disease

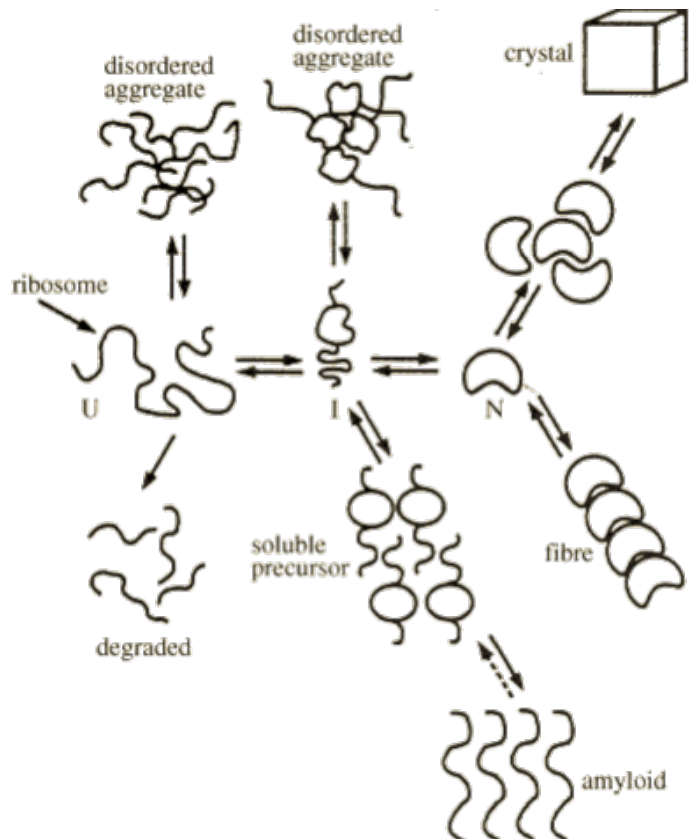
Accumulating evidence for soluble oligomers formed during early aggregation events as the major cytotoxic species



Protein Folding and Its Link with Human Disease. C. M. Dobson, 2002

Amyloid- β and Alzheimer's Disease

Accumulating evidence for soluble oligomers formed during early aggregation events as the major cytotoxic species

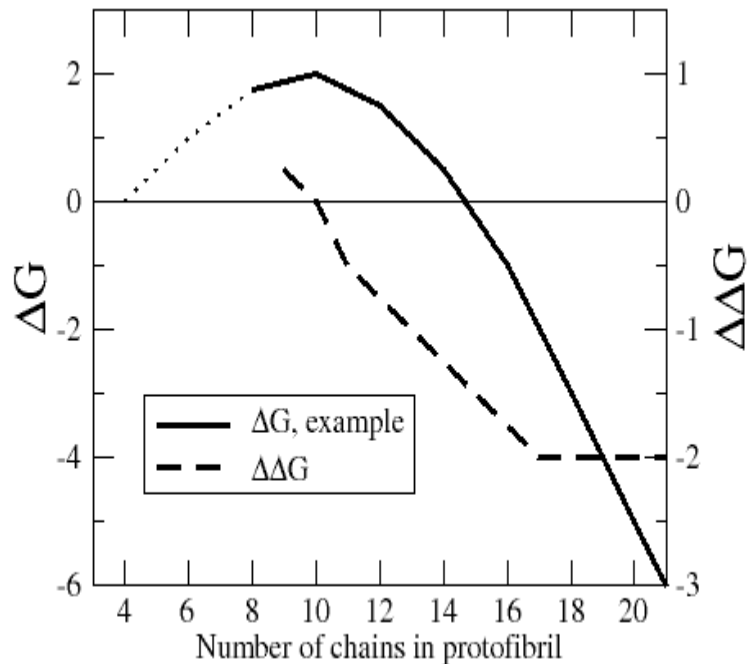


Protein Folding and Its Link with Human Disease. C. M. Dobson, 2002

Alternatively, fibrils have different morphologies depending on preparation. Strains of prion disease are known to arise from different morphologies.

Amyloid- β and Alzheimer's Disease

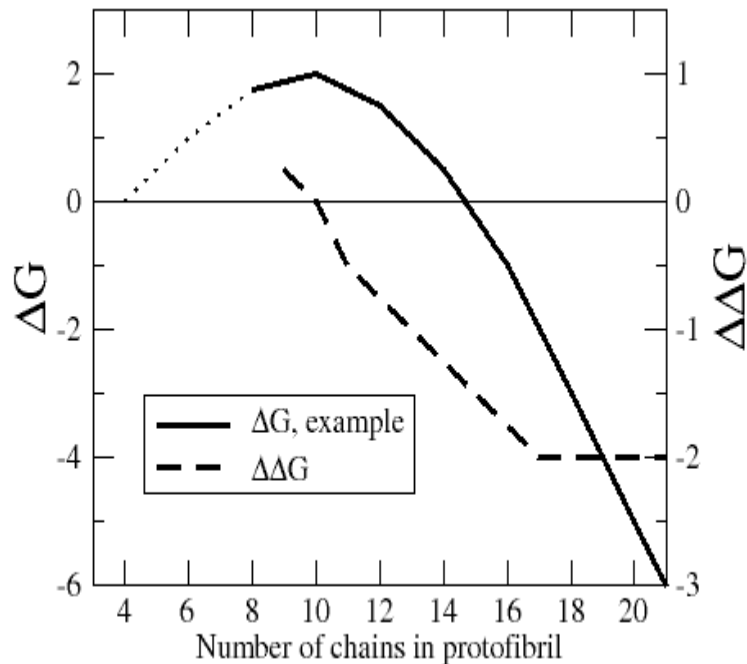
Amyloid fibrils vs. soluble oligomers as the cause of disease, toxic species to the cell.



Kinetics of fibrillization is well described by nucleation and polymerization mechanism (Ferrone)

Amyloid- β and Alzheimer's Disease

Amyloid fibrils vs. soluble oligomers as the cause of disease, toxic species to the cell.



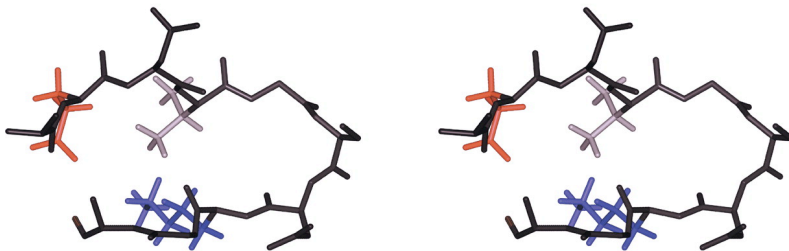
- What is critical nucleus size and structure?
- What fibril lengths define the growth and plateauing kinetic profiles?
- How do these species change under point mutations: Familial Alzheimer's Disease.
- Monomers/Oligomers in solution and membrane

Kinetics of fibrillization is well described by nucleation and polymerization mechanism (Ferrone)

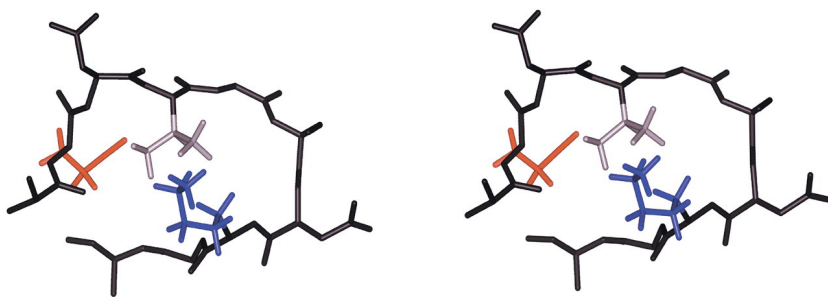
Amyloid- β Monomer

At present there are no reliable structural studies of the monomeric forms of $A\beta_{1-40,42}$

Family I



Family II



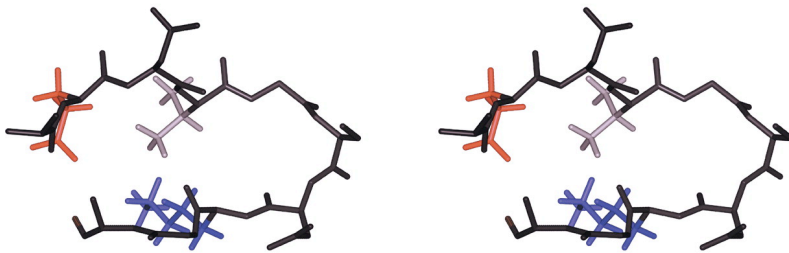
(Teplow & co-workers, Proteins 2005)

Amyloid- β Monomer

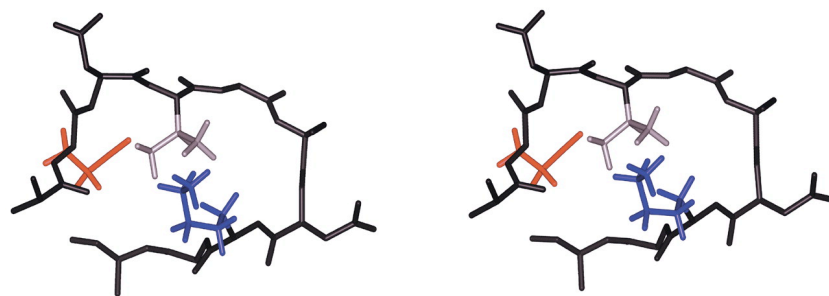
At present there are no reliable structural studies of the monomeric forms of $A\beta_{1-40,42}$

$A\beta_{1-40,42}$ monomer in context of *protofilament* from SS-NMR constraints

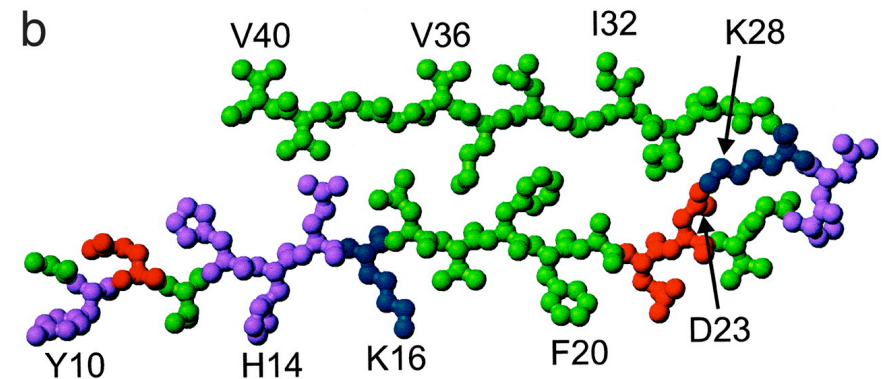
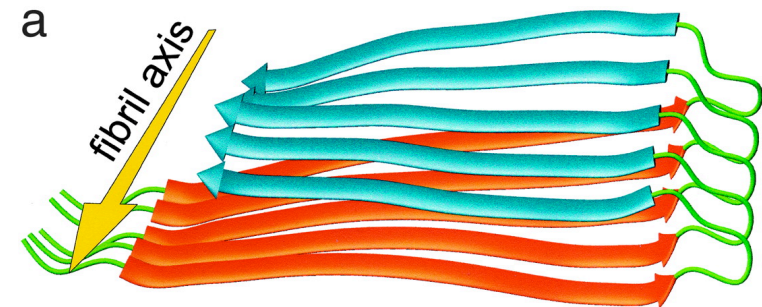
Family I



Family II



(Teplow & co-workers, Proteins 2005)

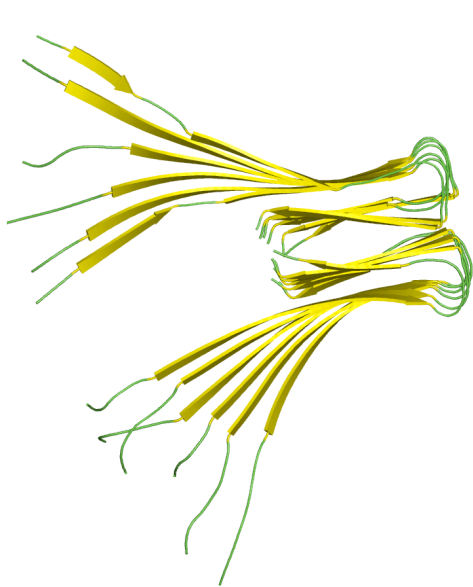


(Petkova et al. PNAS 2002)

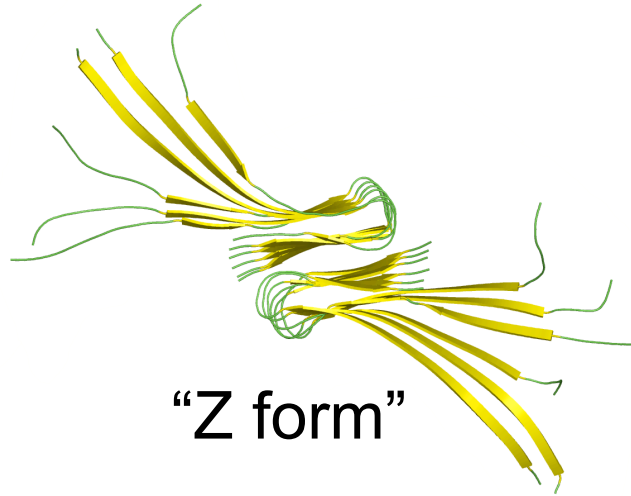
Solid State NMR Data Summary

Two quaternary structure (C2x and C2z) were proposed for the *protofibril* in the 2002 model and later C2z form in 2006

“X form”



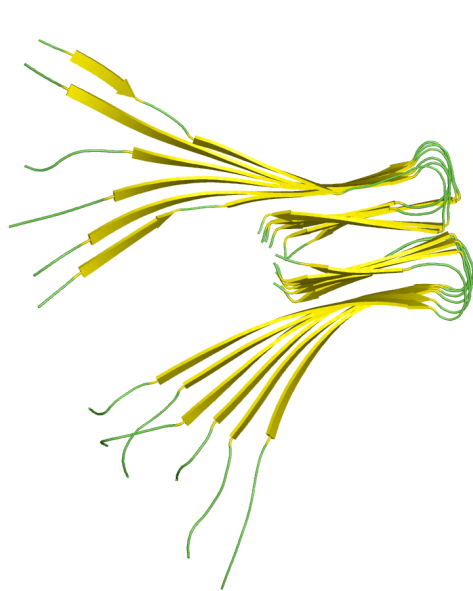
“Z form”



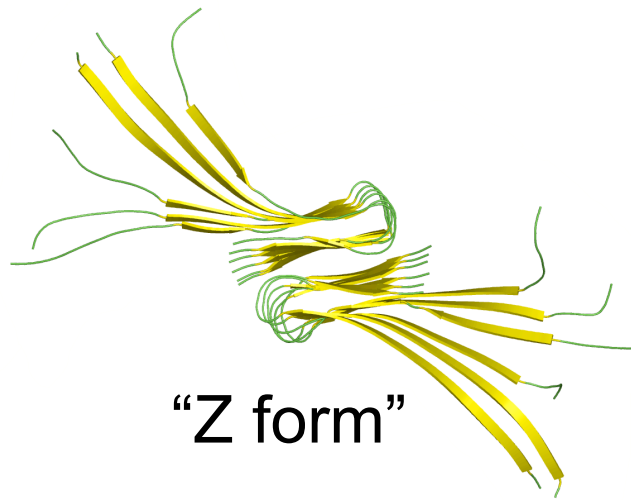
Solid State NMR Data Summary

Two quaternary structure (C2x and C2z) were proposed for the *protofibril* in the 2002 model and later C2z form in 2006

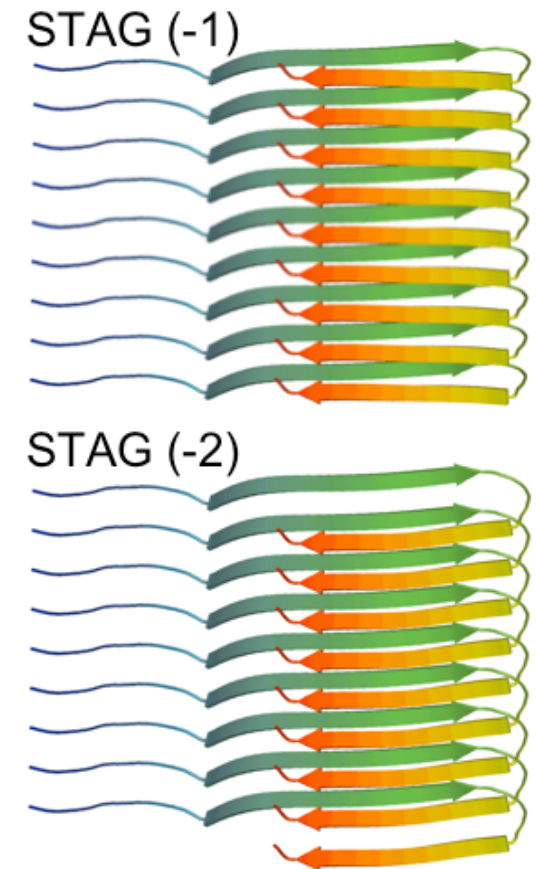
“X form”



“Z form”

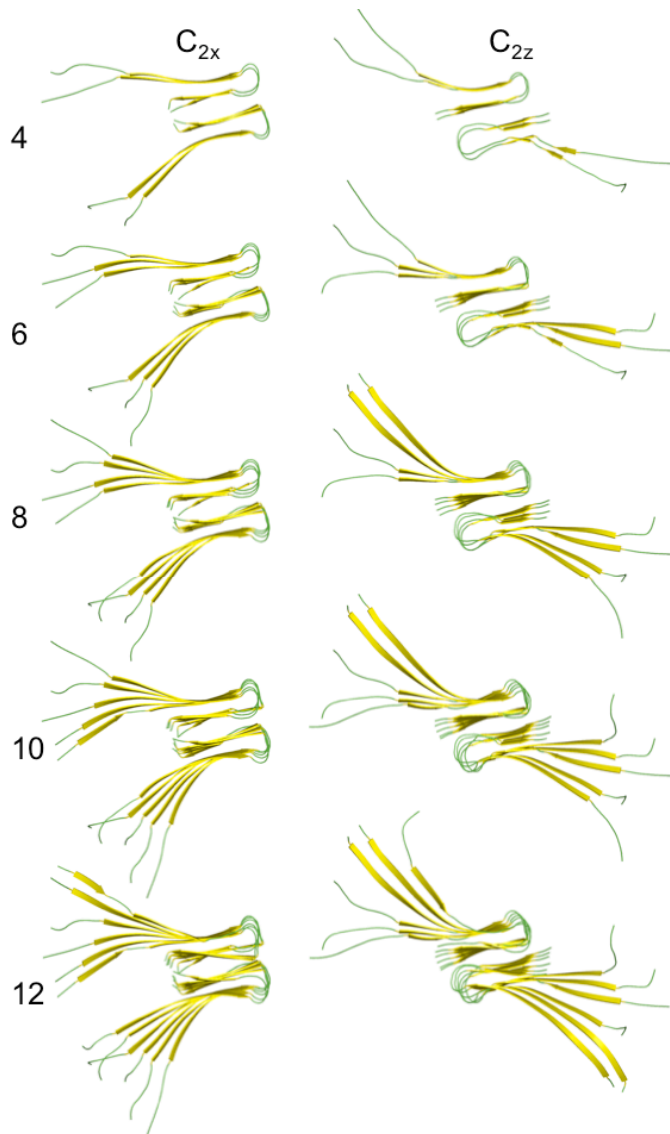


Petkova et al. PNAS 2002



Optimal interactions result in interdigitation of sidechains between strands to adopt a stagger on each *protofilament*

Model for Amyloid- β Fibrillization

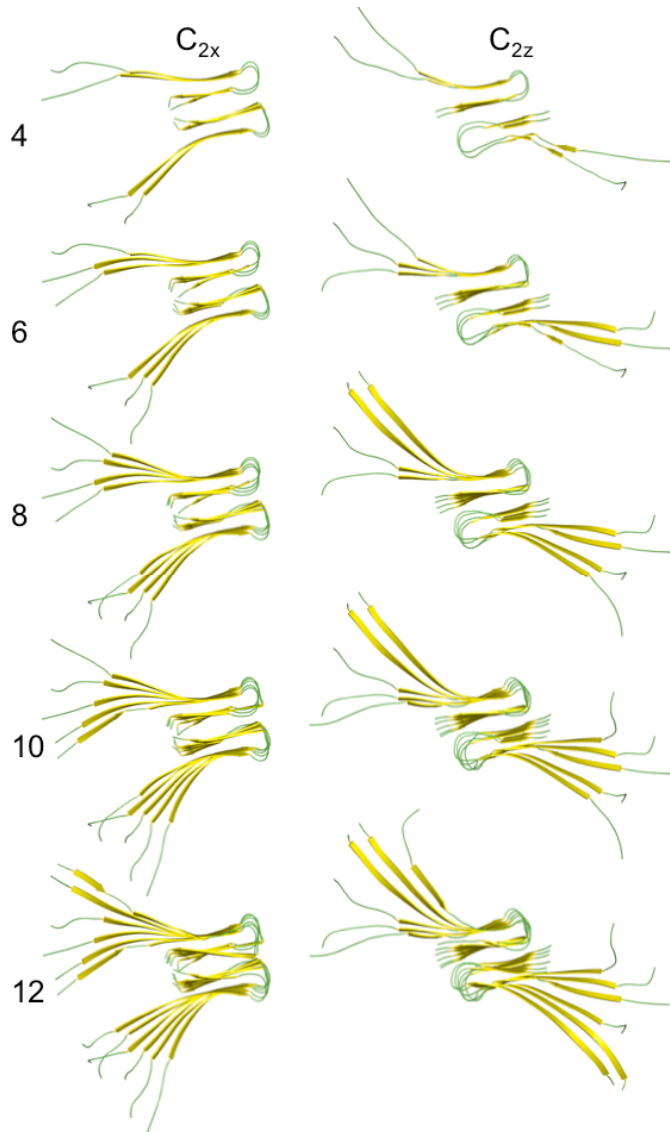


Large 40-chain *protofibril* was model built under the two symmetry forms

Satisfied most of SS-NMR restraints as of 2004. We found restraints in turn region were unphysical

Fawzi, Okabe, Yap, THG (2007) *J. Mol. Biol.* 365, 535-550

Model for Amyloid- β Fibrillization



Large 40-chain *protofibril* was model built under the two symmetry forms

Satisfied most of SS-NMR restraints as of 2004. We found restraints in turn region were unphysical

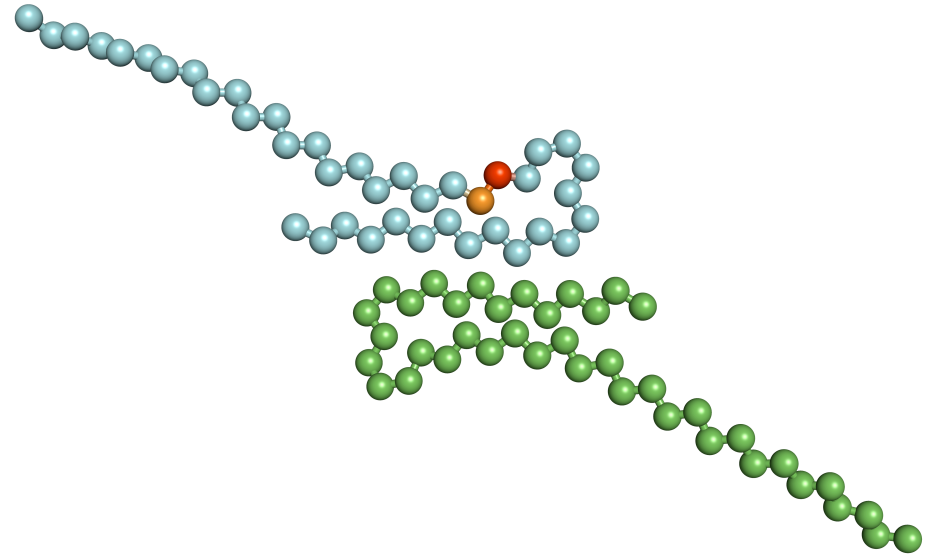
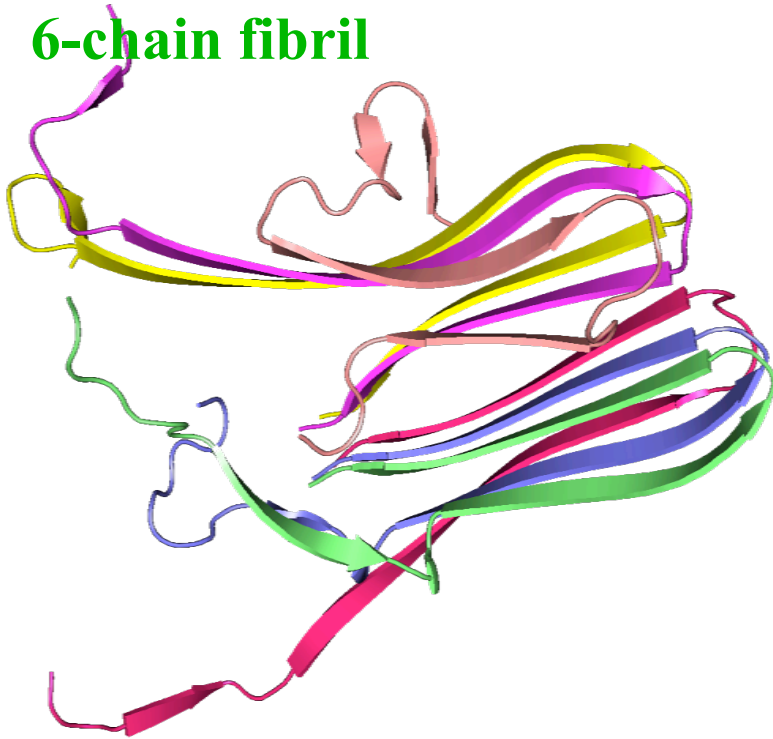
Large protofibril was equilibrated at $T=0.45$ ($\sim 335\text{K}$); -1 stagger develops

Smaller protofibrils were extracted from innermost region of large protofibril, ~ 100 trajectories were run at temperature $T=0.45$ ($\sim 335\text{K}$) for 1,000,000 timesteps (5000τ , $\sim 100\text{ns}$)

Fawzi, Okabe, Yap, THG (2007) J. Mol. Biol. 365, 535-550

Protofibril Stability as a Function of Size

6-chain fibril

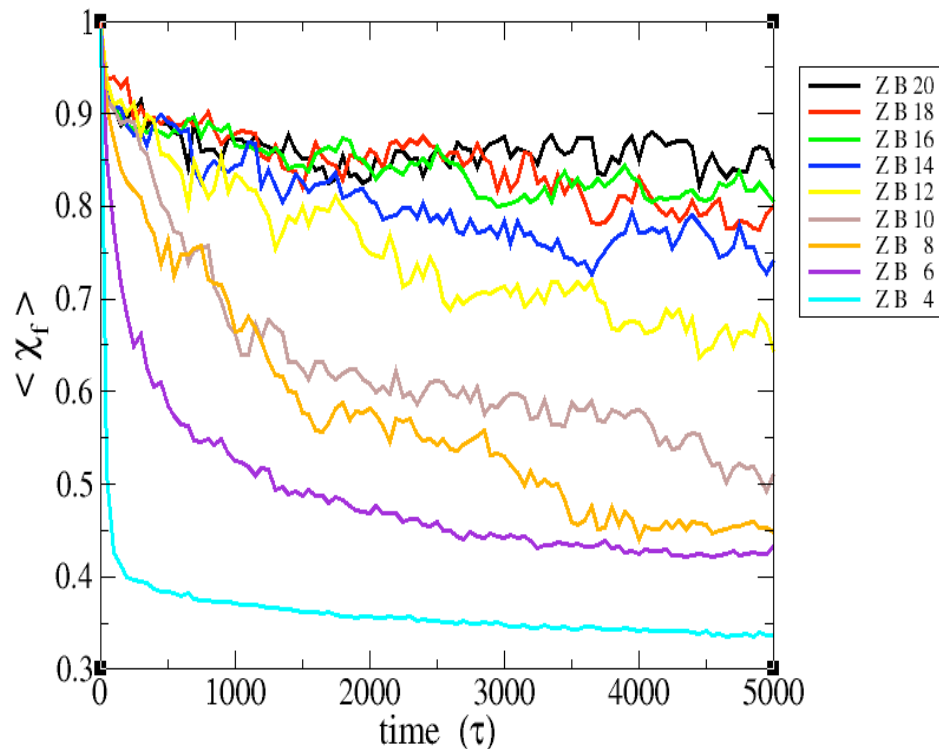


Protofibril morphology vs time was quantified using χ similarity parameter to characterize protofibril (structural) stability

$$\chi = \frac{1}{M} \sum_{\substack{c,d,c \neq d \\ c \text{ and } d \\ \text{edge chain}}}^N \sum_{\substack{i,j \\ \text{in } \beta \text{ region}}}^K \theta(\varepsilon - |r_{ij} - r_{ij}^{Nat}|)$$

Defibrilization (Structural Stability)

Measure $\langle \chi_f \rangle$ vs. time at T=335K for systems with seeds of varying number of chains: 4, 6, 8, 10, 12, 14, 16, 18, 20

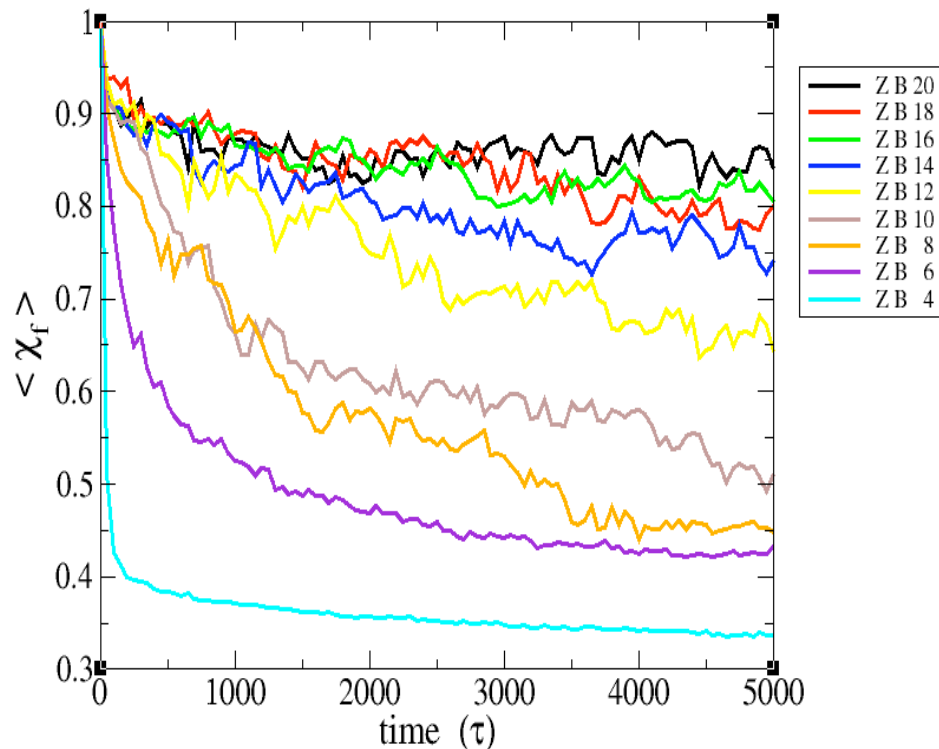


$$\chi = \frac{1}{M} \sum_{\substack{c,d,c \neq d \\ c \text{ and } d \\ \text{edge chain}}}^N \sum_{\substack{i,j \\ \text{in } \beta \text{ region}}}^K \theta(\varepsilon - |r_{ij} - r_{ij}^{Nat}|)$$

Fawzi, Okabe, Yap, THG (2007) J. Mol. Biol. 365, 535-550

Defibrilization (Structural Stability)

Measure $\langle \chi_f \rangle$ vs. time at T=335K for systems with seeds of varying number of chains: 4, 6, 8, 10, 12, 14, 16, 18, 20



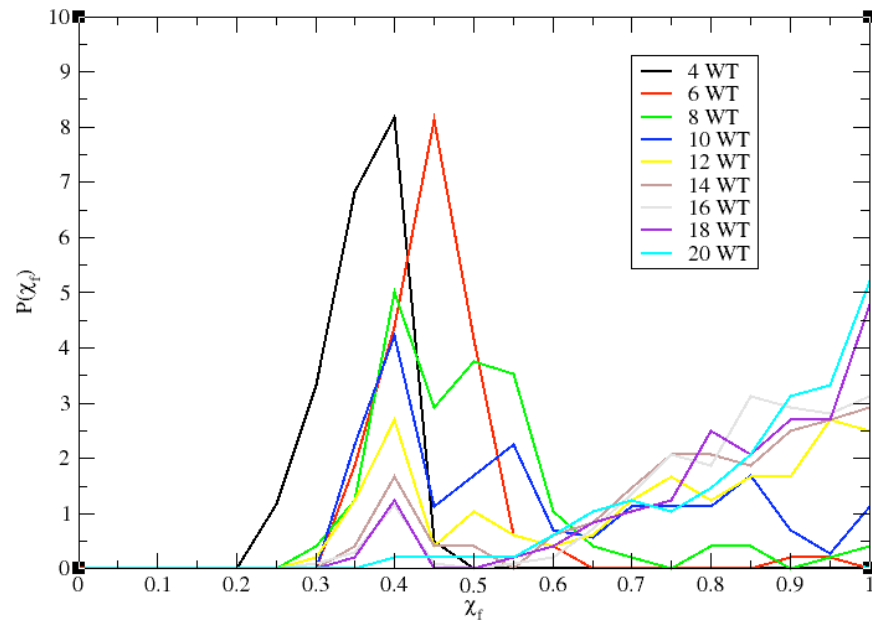
$$\chi = \frac{1}{M} \sum_{\substack{c,d,c \neq d \\ c \text{ and } d \\ \text{edge chain}}}^N \sum_{\substack{i,j \\ \text{in } \beta \text{ region}}}^K \theta(\varepsilon - |r_{ij} - r_{ij}^{Nat}|)$$

Defibrilization is complete for smallest oligomers but at ~16 chains, the fibril order is preserved. No difference between C2X and C2Z

Fawzi, Okabe, Yap, THG (2007) J. Mol. Biol. 365, 535-550

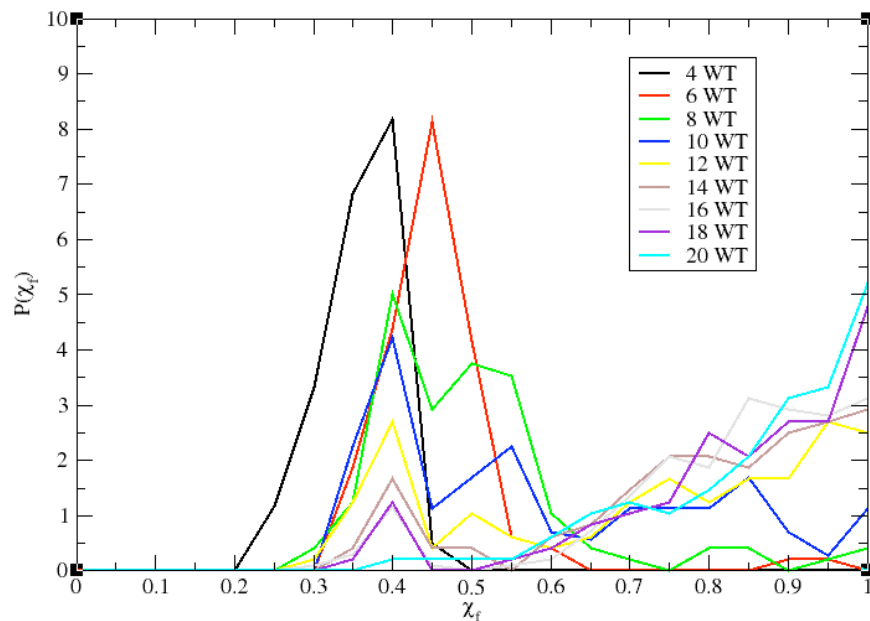
What is the Critical Nucleus?

Fraction of trajectories with $\chi_f > 0.65$ measures population, P_n , of n-ordered monomers in protofibril, in equilibrium with ensemble that lost structural order of one monomer end, P_{n-1} .



What is the Critical Nucleus?

Fraction of trajectories with $\chi_f > 0.65$ measures population, P_n , of n-ordered monomers in protofibril, in equilibrium with ensemble that lost structural order of one monomer end, P_{n-1} .

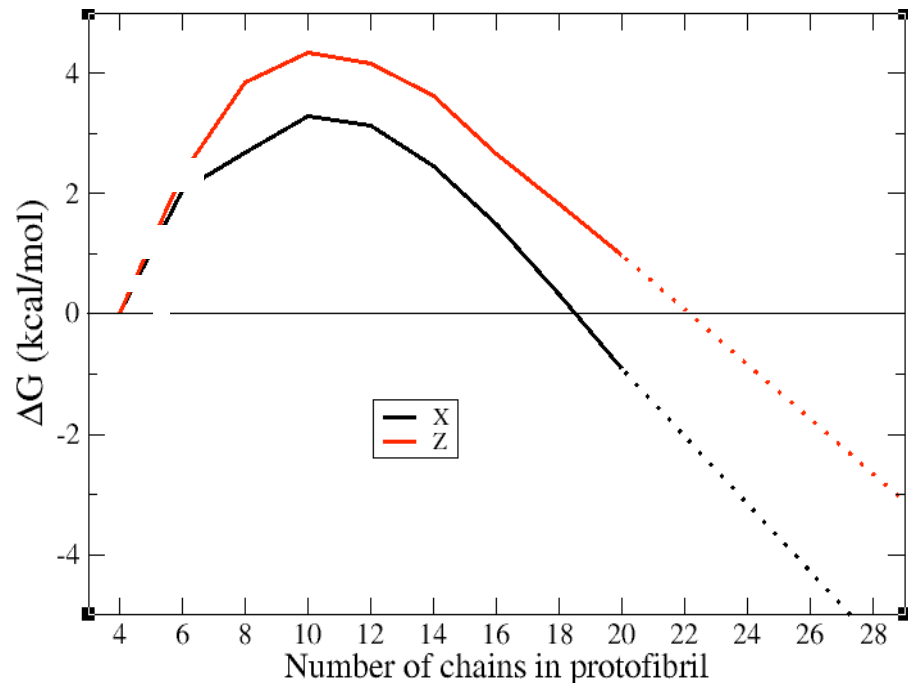


Integration over n of

$$\frac{d\Delta G}{dn} = -kT \ln \left(\frac{[P_{n-1}]}{[P_n]} \right)$$

What is the Critical Nucleus?

Fraction of trajectories with $\chi_p > 0.65$ measures population, P_n , of n-ordered monomers in protofibril, in equilibrium with ensemble that lost structural order of one monomer end, P_{n-1} .



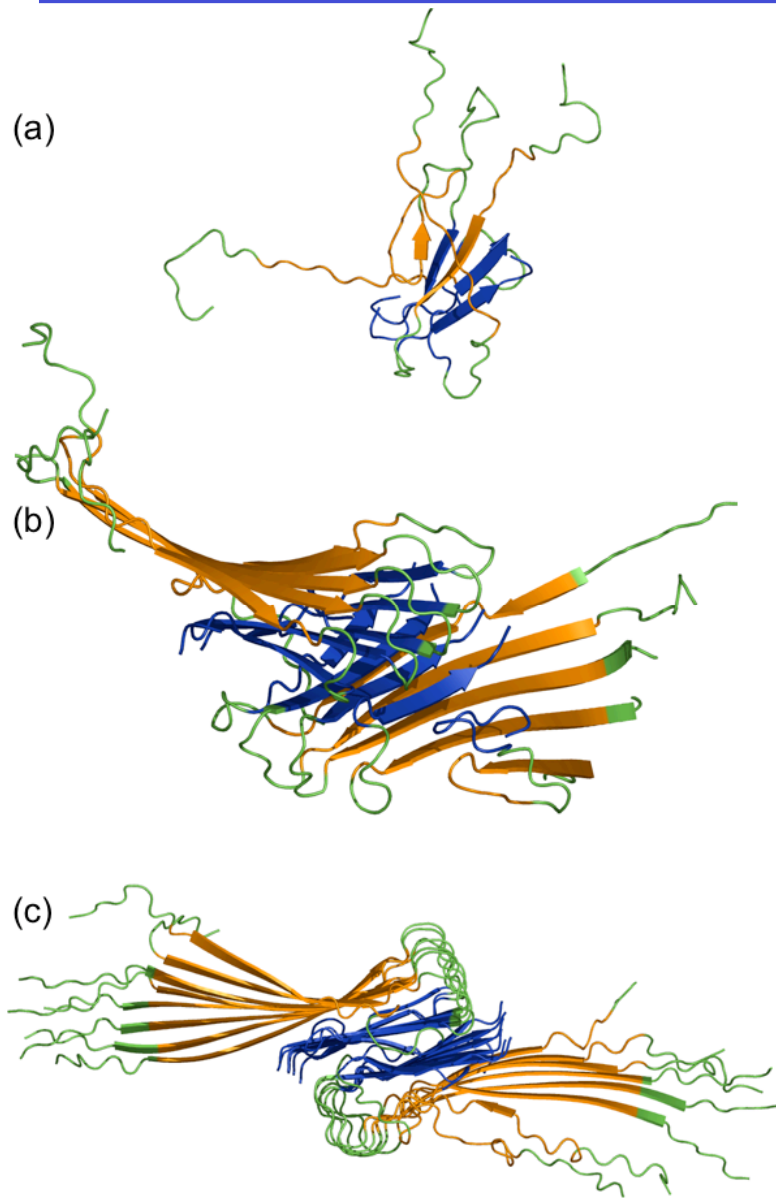
Integration over n of

$$\frac{d\Delta G}{dn} = -kT \ln \left(\frac{[P_{n-1}]}{[P_n]} \right)$$

Allows us to determine critical nucleus as maximum in free energy

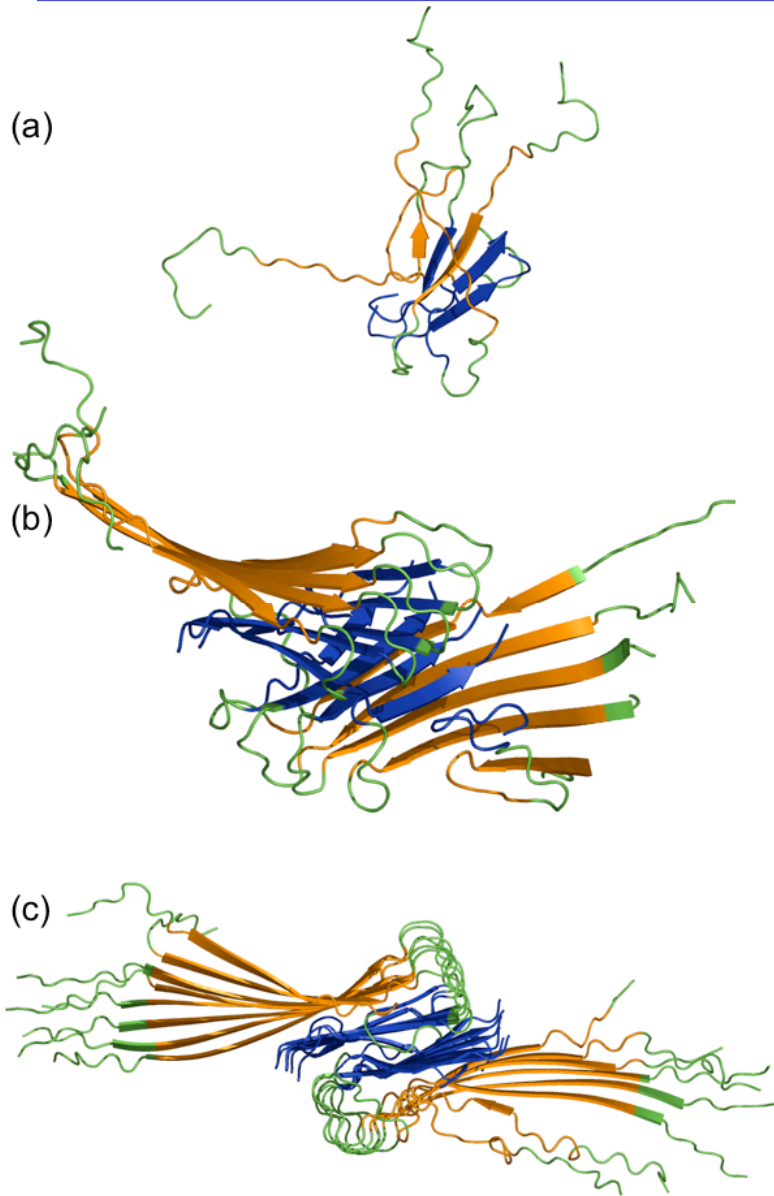
< 8 chains: mechanism not reversible; reaction coordinate for small oligomers needs further study

What is the Critical Nucleus?



Below the critical nucleus (~8 chains) there is no protofilament or protofibril order but some β -strand structure. Entropy wins out

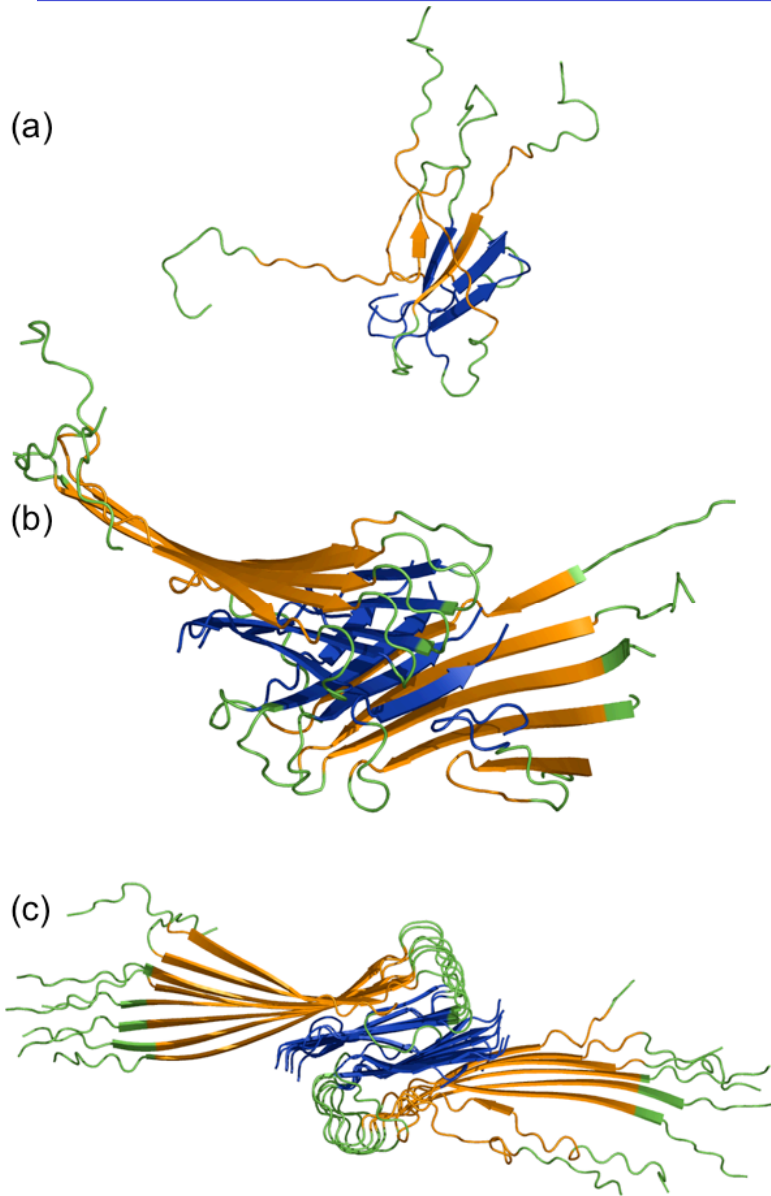
What is the Critical Nucleus?



Below the critical nucleus (~8 chains) there is no protofilament or protofibril order but some β -strand structure. Entropy wins out

At the critical nucleus of 10–12 chains protofilament order but no protofibril order. Entropy of quaternary disorder wins out

What is the Critical Nucleus?



Below the critical nucleus (~8 chains) there is no protofilament or protofibril order but some β -strand structure. Entropy wins out

At the critical nucleus of 10–12 chains protofilament order but no protofibril order. Entropy of quaternary disorder wins out

Post critical nucleus (>10–12 chains) quaternary structure is stabilized by sufficient enthalpic gains to overcome all entropic factors

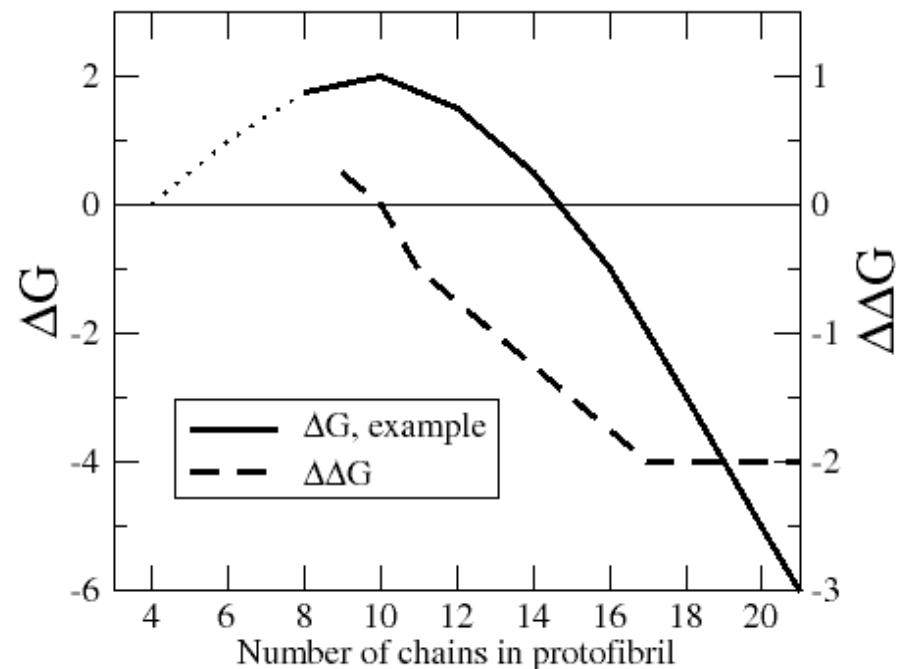
Oligomers vs. Fibrils

Fibril extension studies and Familial Alzheimer's Disease

Fawzi, Okabe, Yap, THG (2007) *J. Mol. Biol.* 365, 535-550

Fawzi, Kohlstedt, Okabe, THG (2008) *Biophys. J.* 94 2007-16.

Fawzi, Yap, Okabe, Kohlstedt, Brown, THG (2008) *Acc. Chem. Res* 41 1037-47



From Monomer to Oligomers

Fawzi, Phillips, Ruscio, Doucleff, Wemmer, THG (2008). *JACS* 130, 6145-58

Sodt & THG (2009). *Biophysical J.* submitted

Protofibril Extension Studies

Peptides with *arbitrary* configurations for extending protofibril

Start with one protofibril seed and many randomly oriented chains at approximately 50mg/mL

Determine efficiency and mechanism of addition

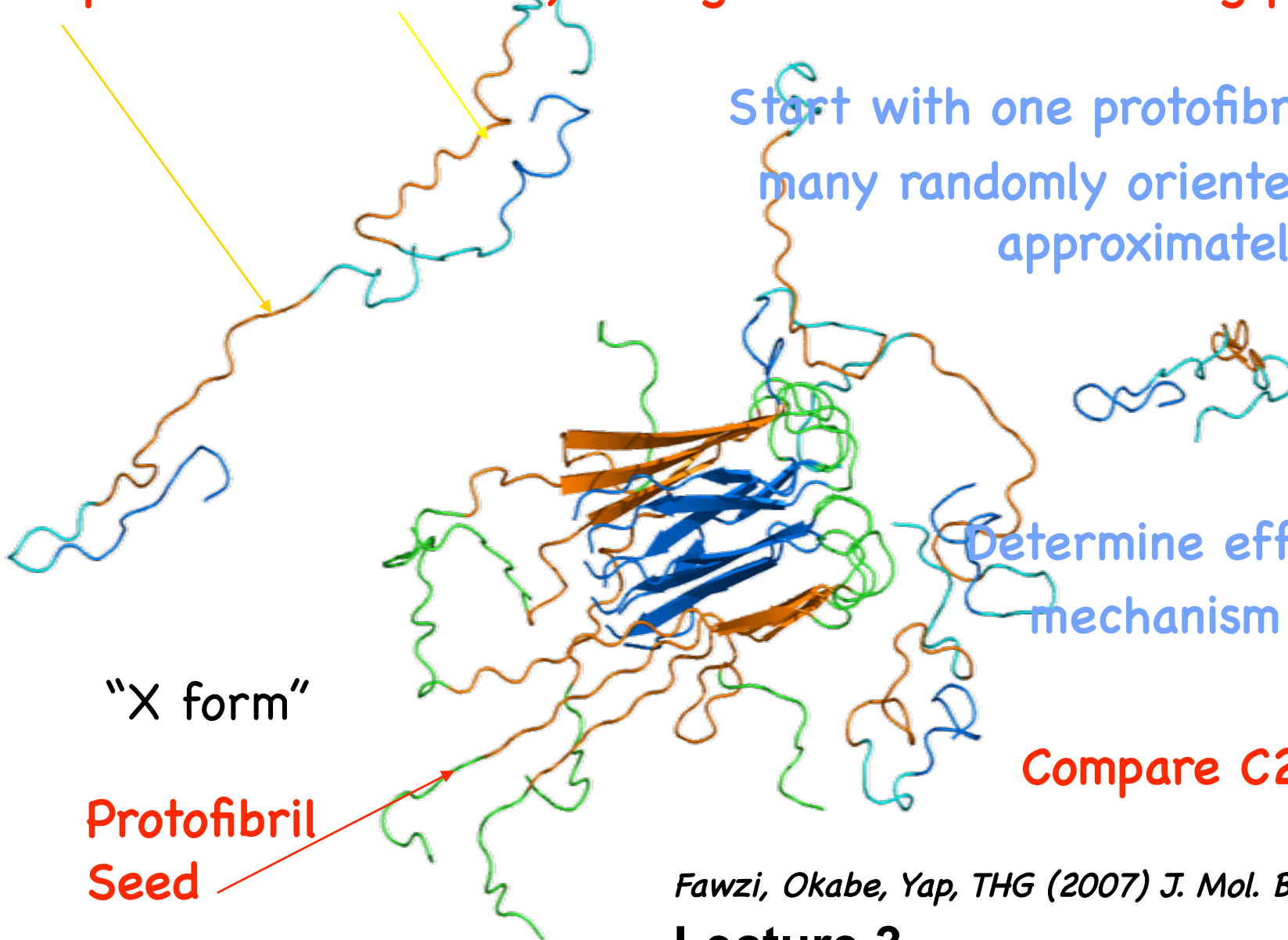
Compare C2X and C2Z

"X form"

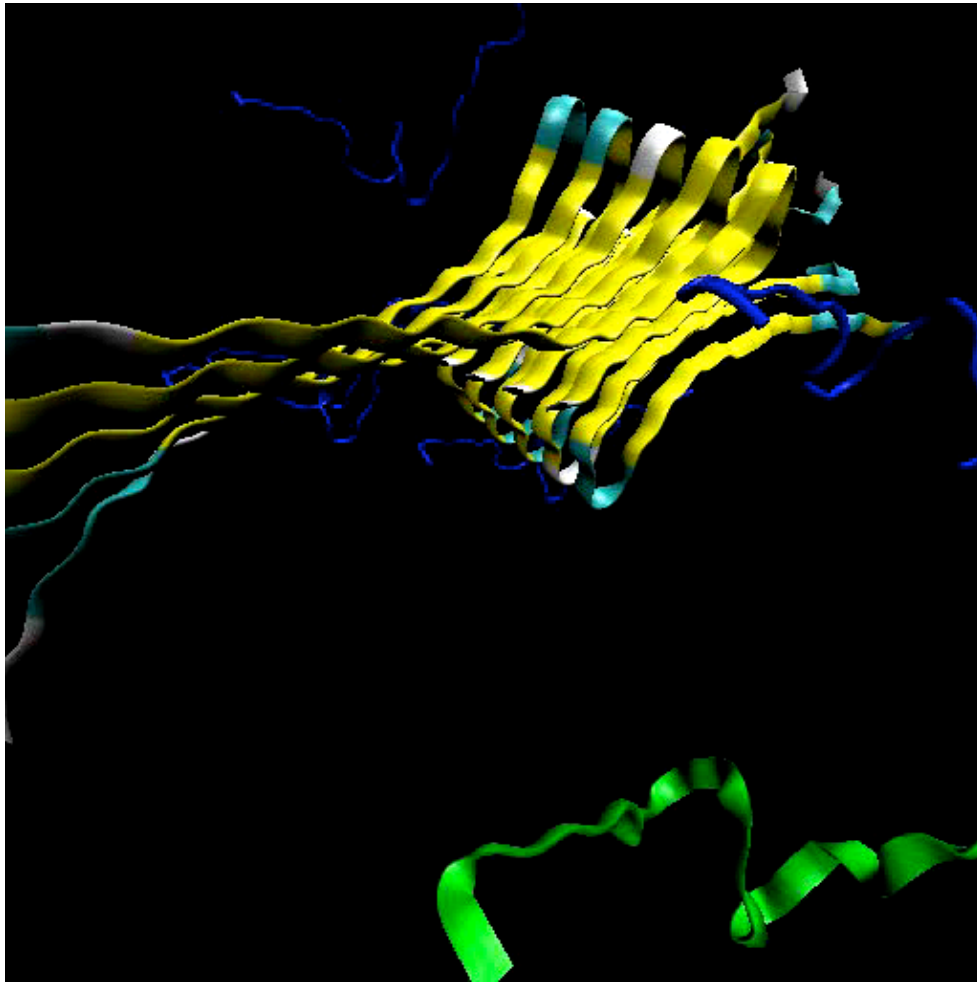
Protofibril
Seed

Fawzi, Okabe, Yap, THG (2007) *J. Mol. Biol.* 365, 535-550

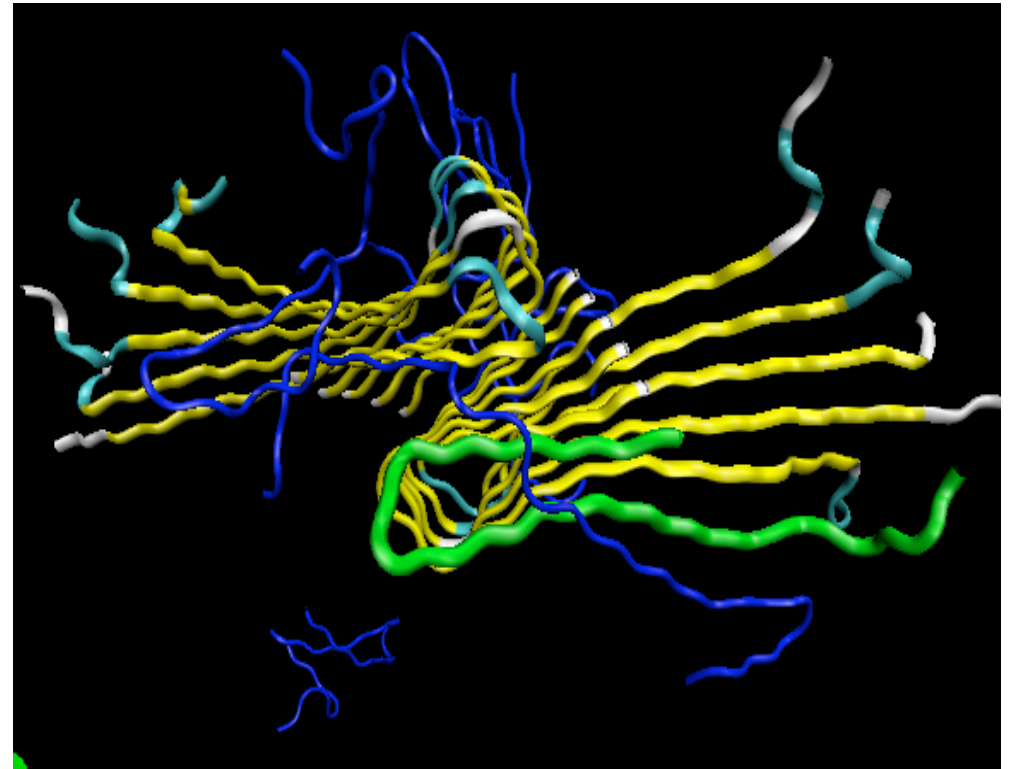
Lecture 3



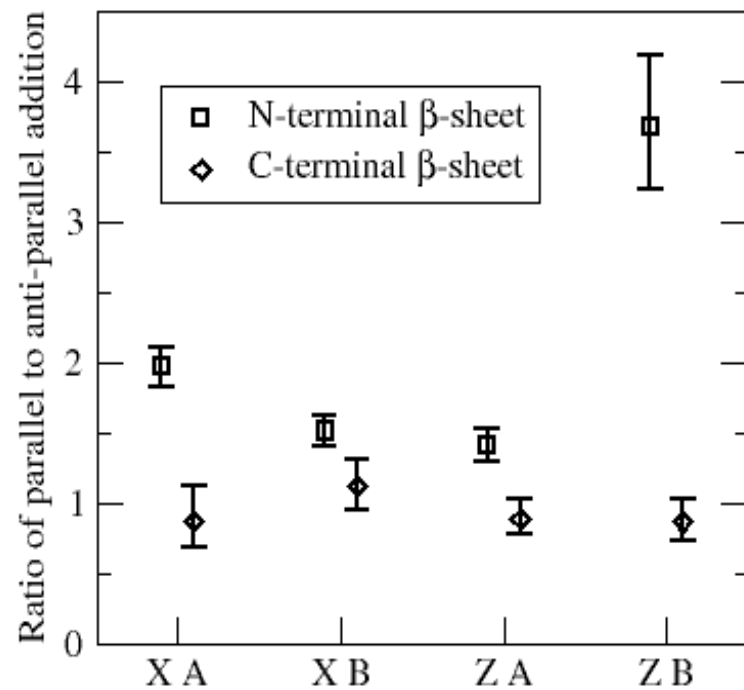
Protofibril Extension Studies



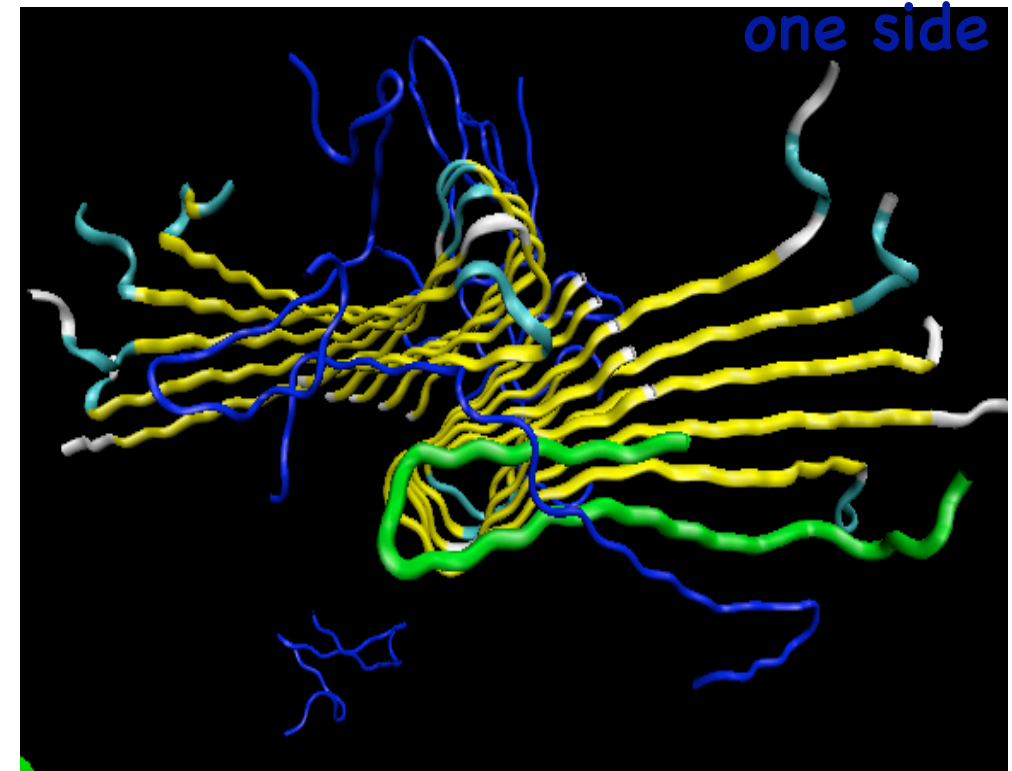
Both C2X and C2Z protofibrils
are capable of addition



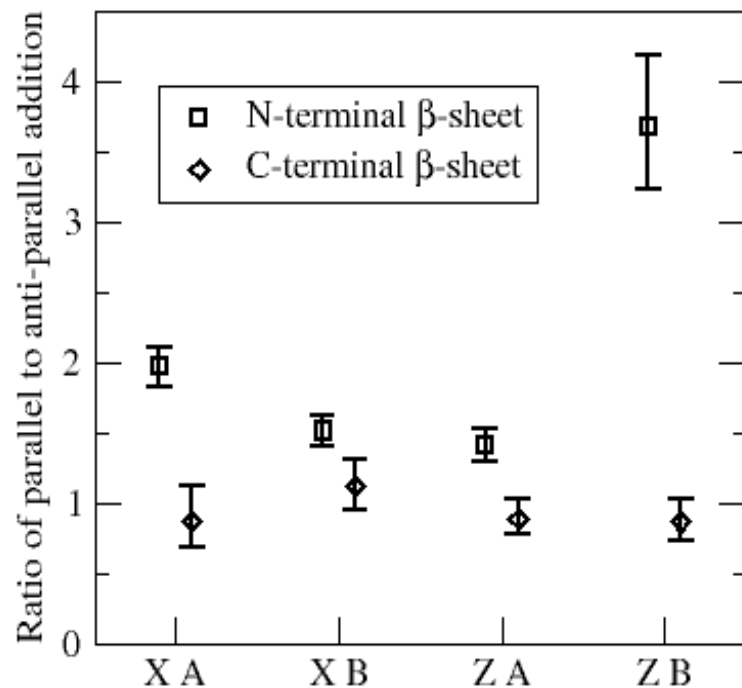
C2X vs.C2Z Protofibril Extension



However, our model finds that C2z has ~2-3 times the rate of "correct" N-terminal extensions on

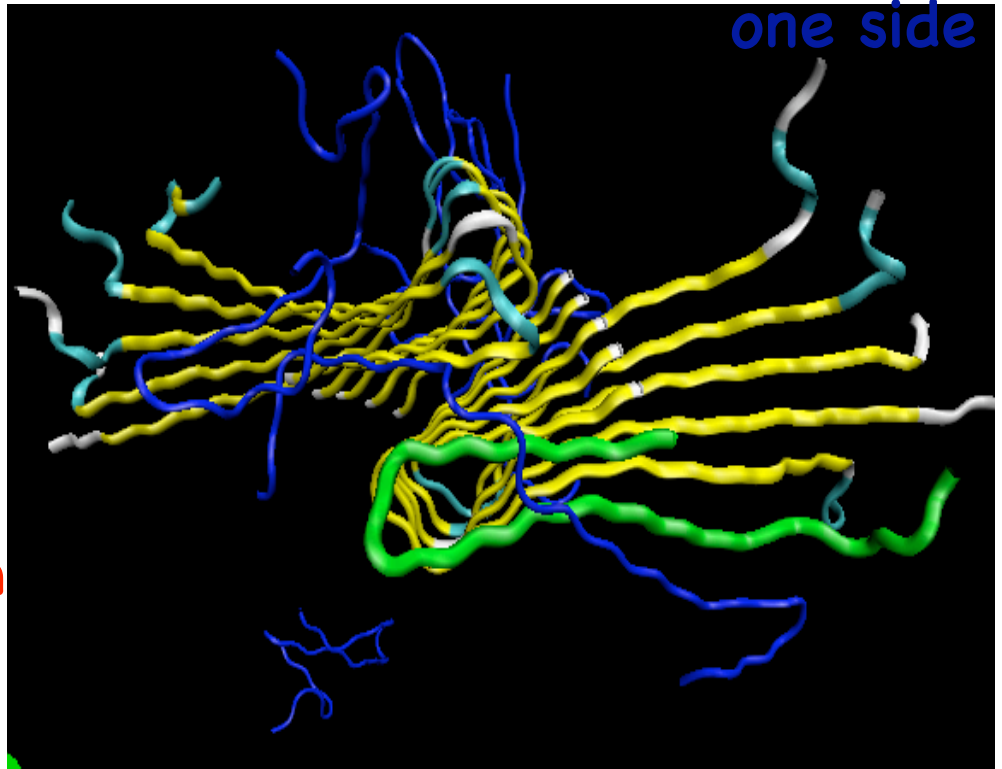


C2X vs. C2Z Protofibril Extension



However, our model finds that C2z has $\sim 2-3$ times the rate of "correct" N-terminal extensions on

one side

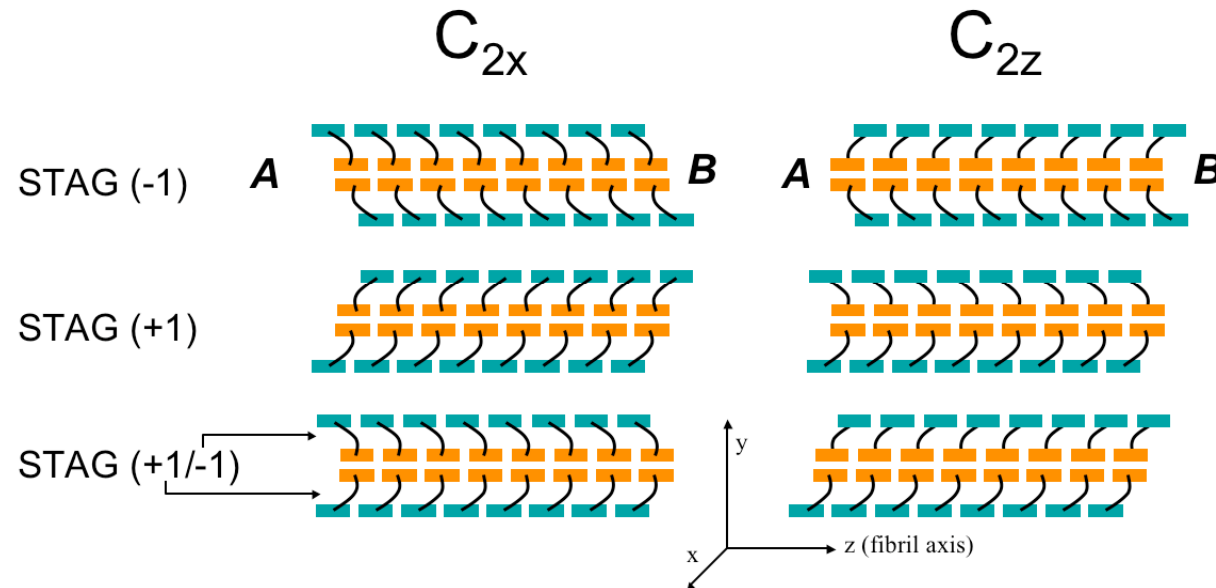


C2z shows unidirectional growth while C2x gives bidirectional growth!

May result in different morphologies of the mature fiber

C2X vs. C2Z Protofibril Extension

SS-NMR does not rule out possibility that there is a mixed stagger

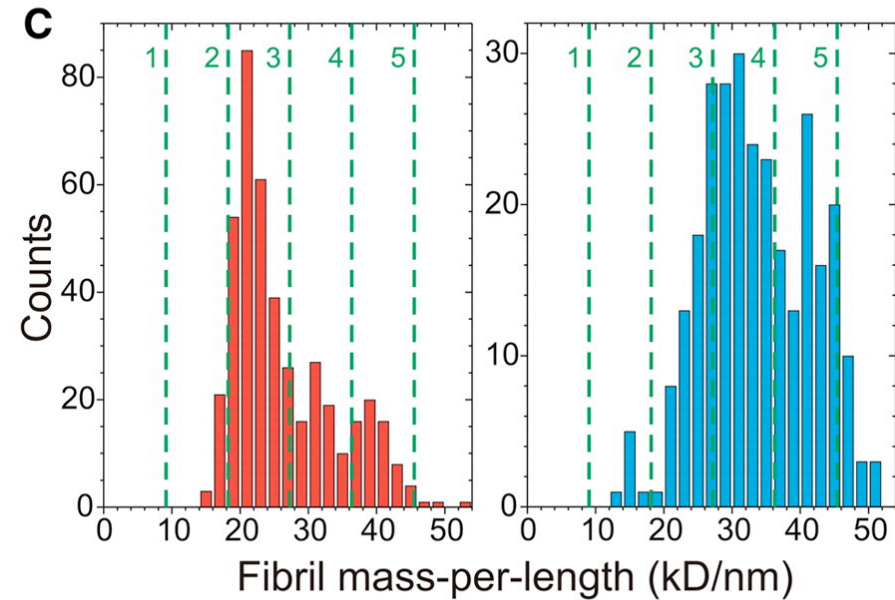
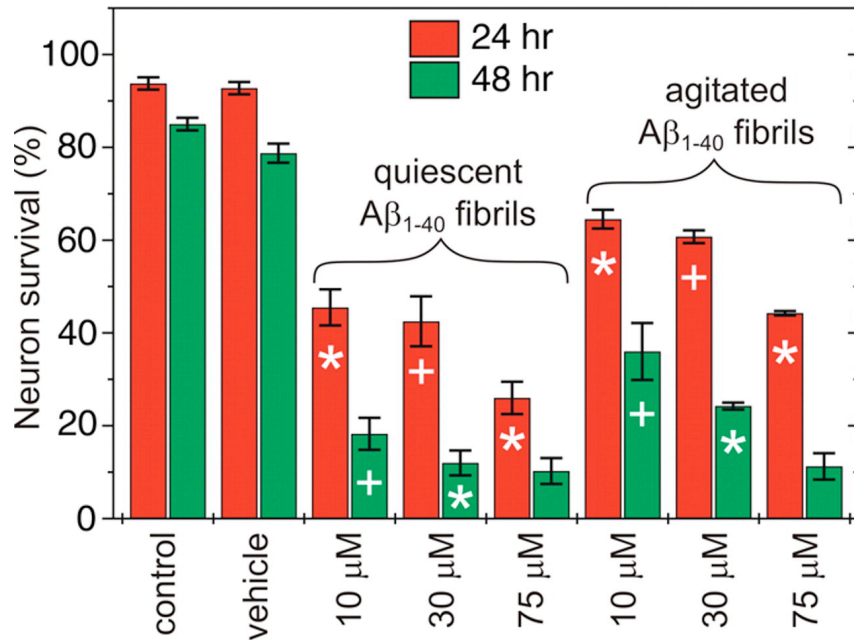


C_{2z} structure (or C_{2x} under mixed stagger) exposes n-terminus region which has sequence patterning which promotes correct association

Polymorphism and Disease

Polymorphism results in different degrees of toxicity in neuronal cell cultures.

Petkova and co-workers, Science 2005



Differences in preparation are manifest in different mass per length distributions.

Certain strains may weaken the correlation between disease and fibrils as an artifact in favor of soluble oligomers as cytotoxic species

Familial Alzheimer's Disease

Well-studied FAD mutants of amyloid β include Dutch (E22Q), Flemish (A21G), Arctic (E22G), all of which have been characterized for both $A\beta_{1-40}$ and $A\beta_{1-42}$ both *in vitro* and *in vivo*.

Dutch (E22Q): cerebral amyloid angiopathy (CAA), deposits in brain vasculature, fibril is toxic, nucleates, fibrillizes faster than WT

Familial Alzheimer's Disease

Well-studied FAD mutants of amyloid β include Dutch (E22Q), Flemish (A21G), Arctic (E22G), all of which have been characterized for both $A\beta_{1-40}$ and $A\beta_{1-42}$ both *in vitro* and *in vivo*.

Dutch (E22Q): cerebral amyloid angiopathy (CAA), deposits in brain vasculature, fibril is toxic, nucleates, fibrillizes faster than WT

Arctic (E22G): classic AD dementia symptoms ; deposits in brain tissue, oligomer is toxic, nucleates more readily than WT, fibril kinetic rate comparable

Familial Alzheimer's Disease

Well-studied FAD mutants of amyloid β include Dutch (E22Q), Flemish (A21G), Arctic (E22G), all of which have been characterized for both $A\beta_{1-40}$ and $A\beta_{1-42}$ both *in vitro* and *in vivo*.

Dutch (E22Q): cerebral amyloid angiopathy (CAA), deposits in brain vasculature, fibril is toxic, nucleates, fibrillizes faster than WT

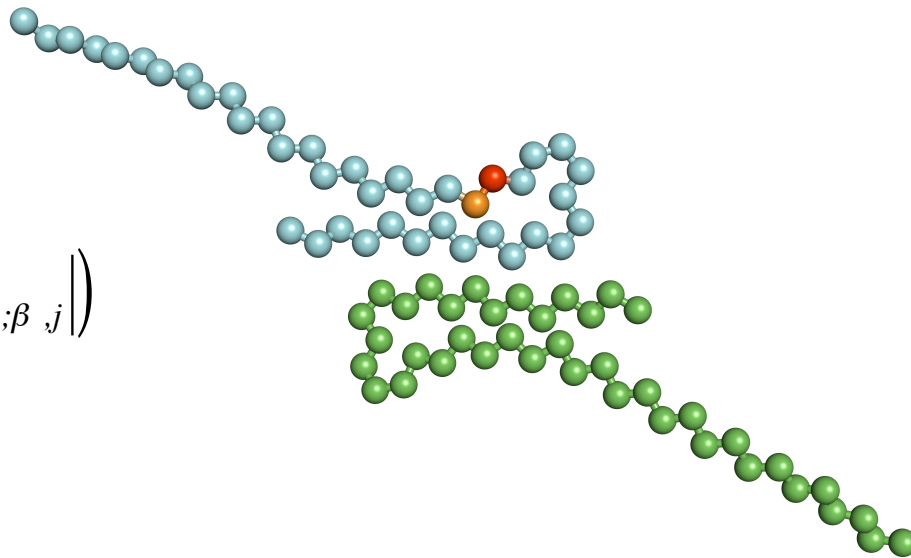
Arctic (E22G): classic AD dementia symptoms ; deposits in brain tissue, oligomer is toxic, nucleates more readily than WT, fibril kinetic rate comparable

Flemish (A21G): CAA and AD; deposits in brain vasculature; deposits are amorphous, largest of all FAD, WT sequences; kinetics slower for nucleation and fibrillization compared to WT

FAD Mutants of Amyloid- β

P_f measures order over *protofilament* (blue or green).

$$\chi = \frac{1}{M} \sum_{\alpha=1}^{N_c} \sum_{\beta>c}^{N_c} \sum_i^N \sum_j^N h\left(\varepsilon - \left| r_{\alpha,i;\beta,j} - r_{\alpha,i;\beta,j}^0 \right| \right)$$



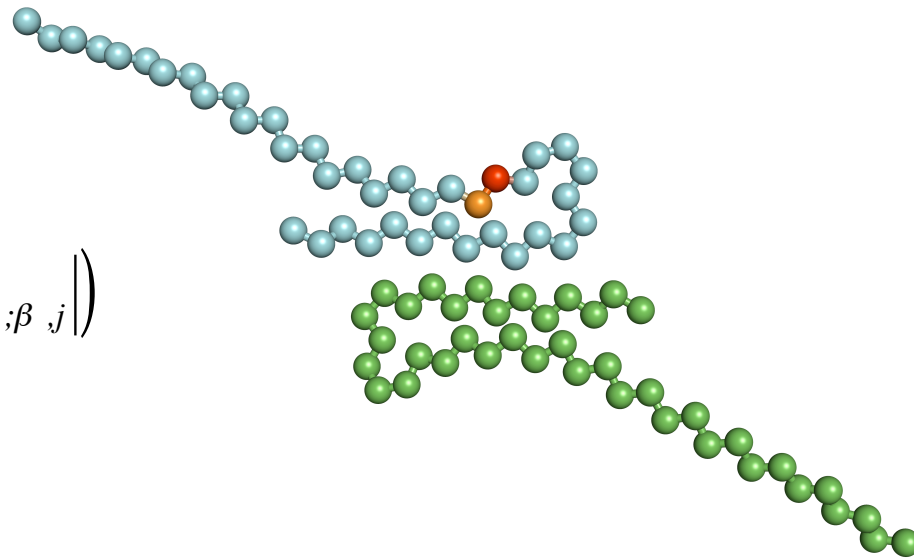
χ_f measures order over *protofibril* (blue and green).

Fawzi, Kohlstedt, Okabe, THG (2008) Biophys. J. 94 2007-16.

FAD Mutants of Amyloid- β

P_f measures order over *protofilament* (blue or green).

$$\chi = \frac{1}{M} \sum_{\alpha=1}^{N_c} \sum_{\beta>c}^{N_c} \sum_i^N \sum_j^N h\left(\varepsilon - \left| r_{\alpha,i;\beta,j} - r_{\alpha,i;\beta,j}^0 \right| \right)$$

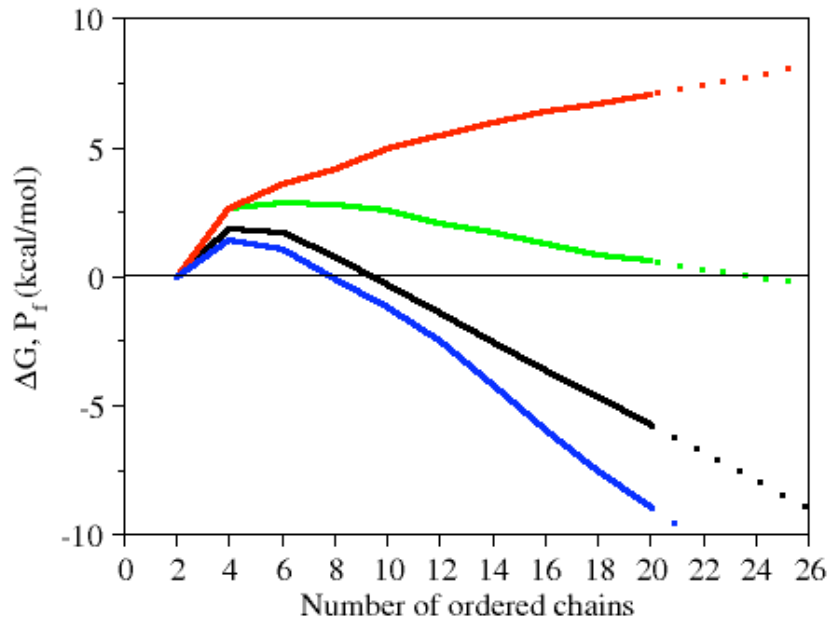
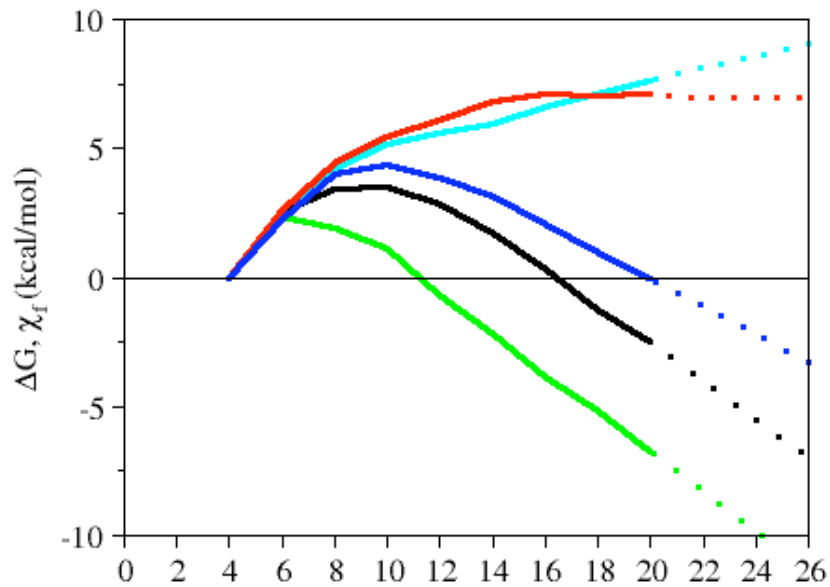


χ_f measures order over *protofibril* (blue and green).

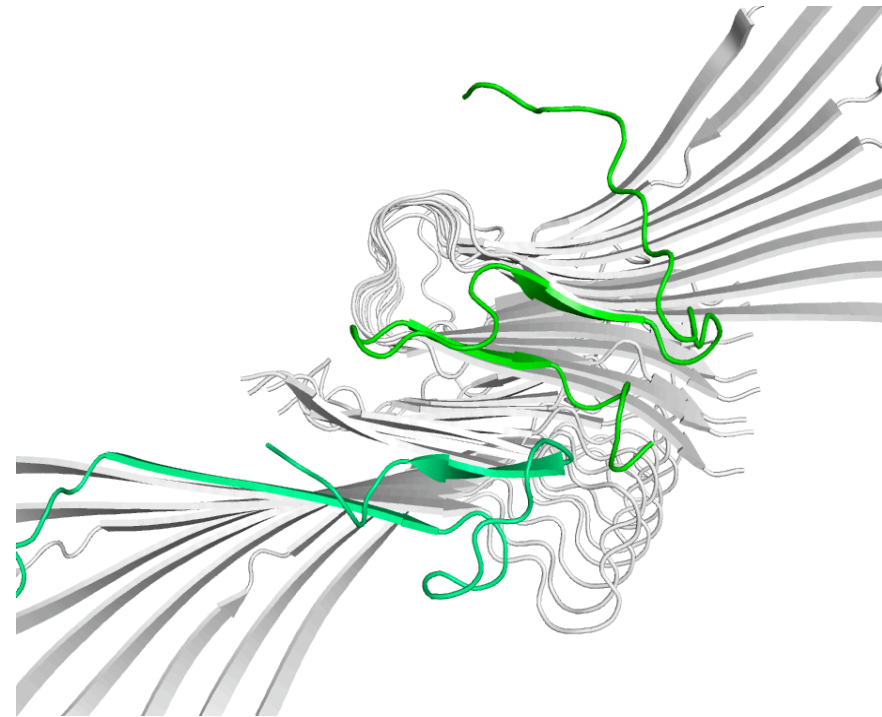
Note that kinetic probes using congo red or thioflavin T only measure cross β -sheet order, but unable to distinguish between Pf and χ_f

Fawzi, Kohlstedt, Okabe, THG (2008) Biophys. J. 94 2007-16.

FAD Mutant: Arctic (E22G)

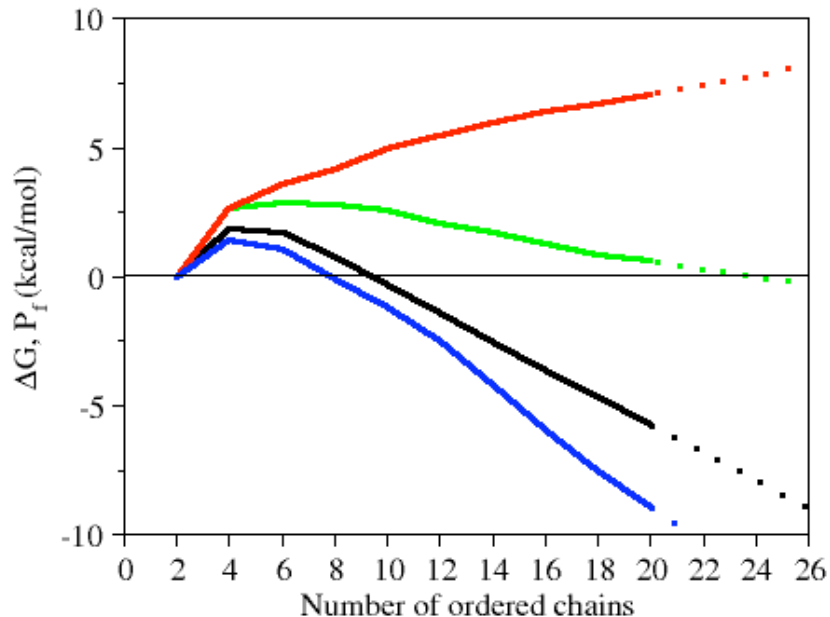
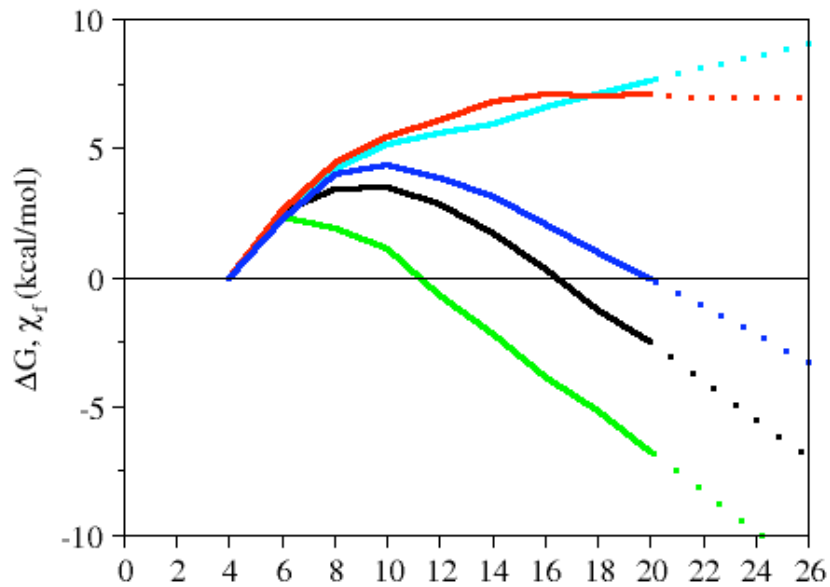


Resulting free energy barrier is reduced and shifted to smaller critical nucleus for protofibrils

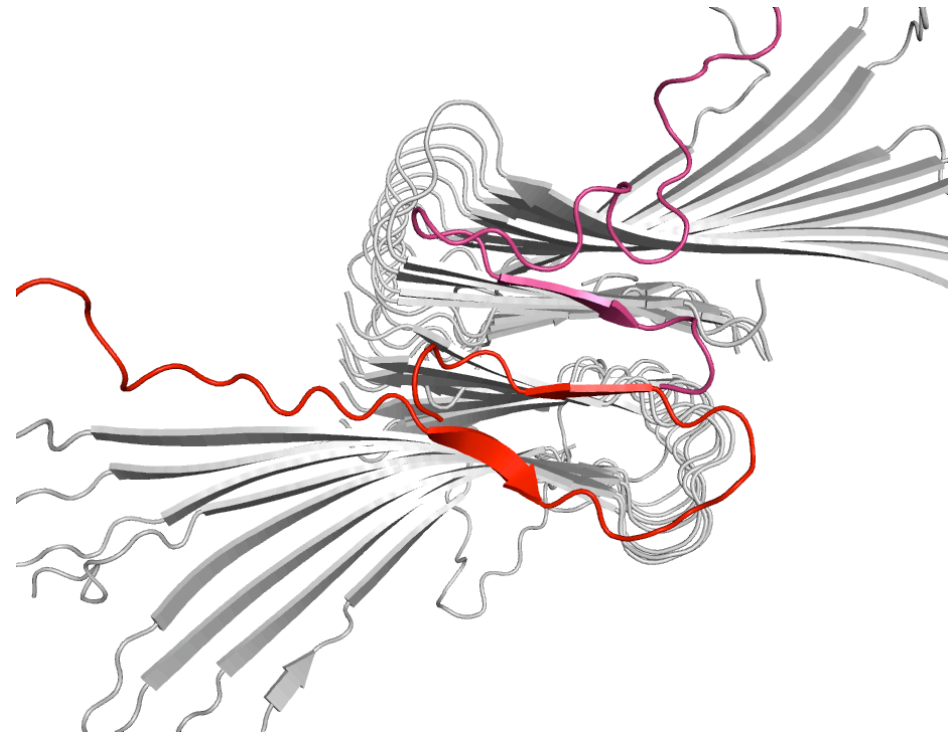


Protofilaments less ordered; but disordered turns stabilize fibril axis

FAD Mutant: Flemish (A21G)

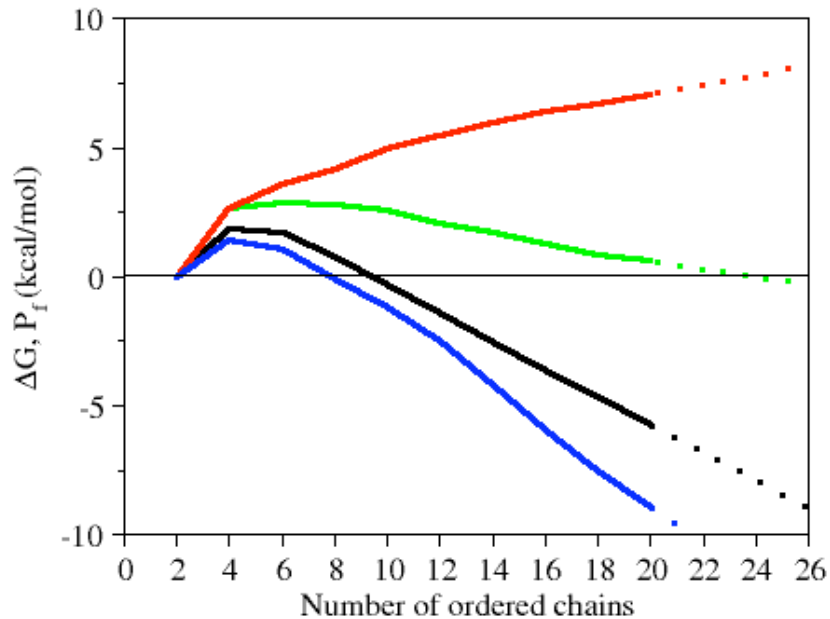
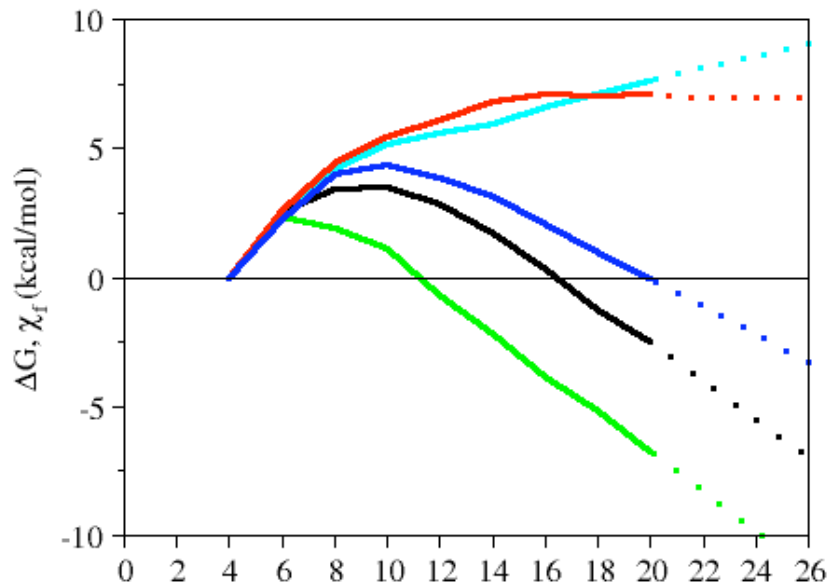


Resulting free energy barrier is increased regardless of order parameter

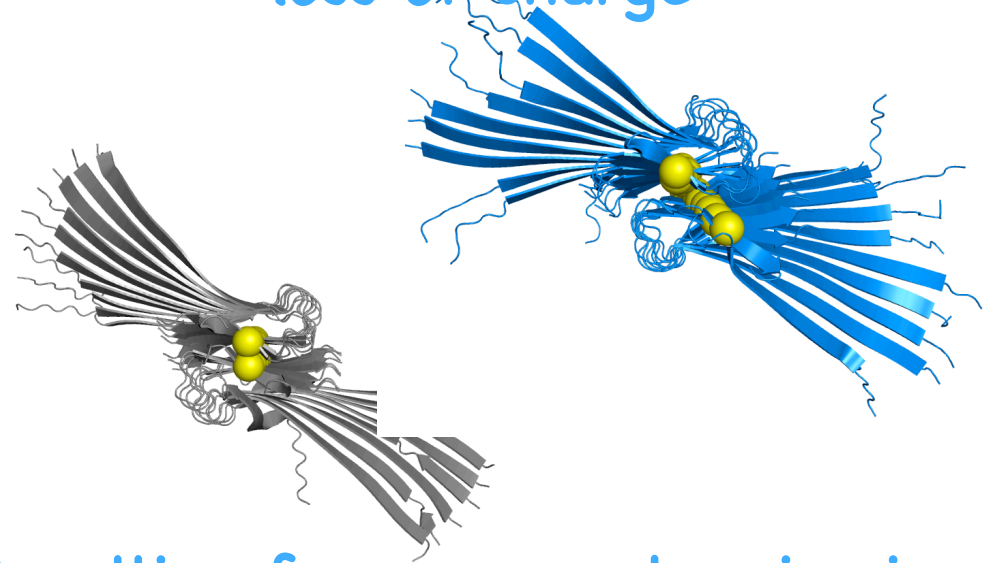


Critical nucleus is much larger than WT (beyond characterized 40-chain protofibril)

FAD Mutant: Dutch (E22Q)



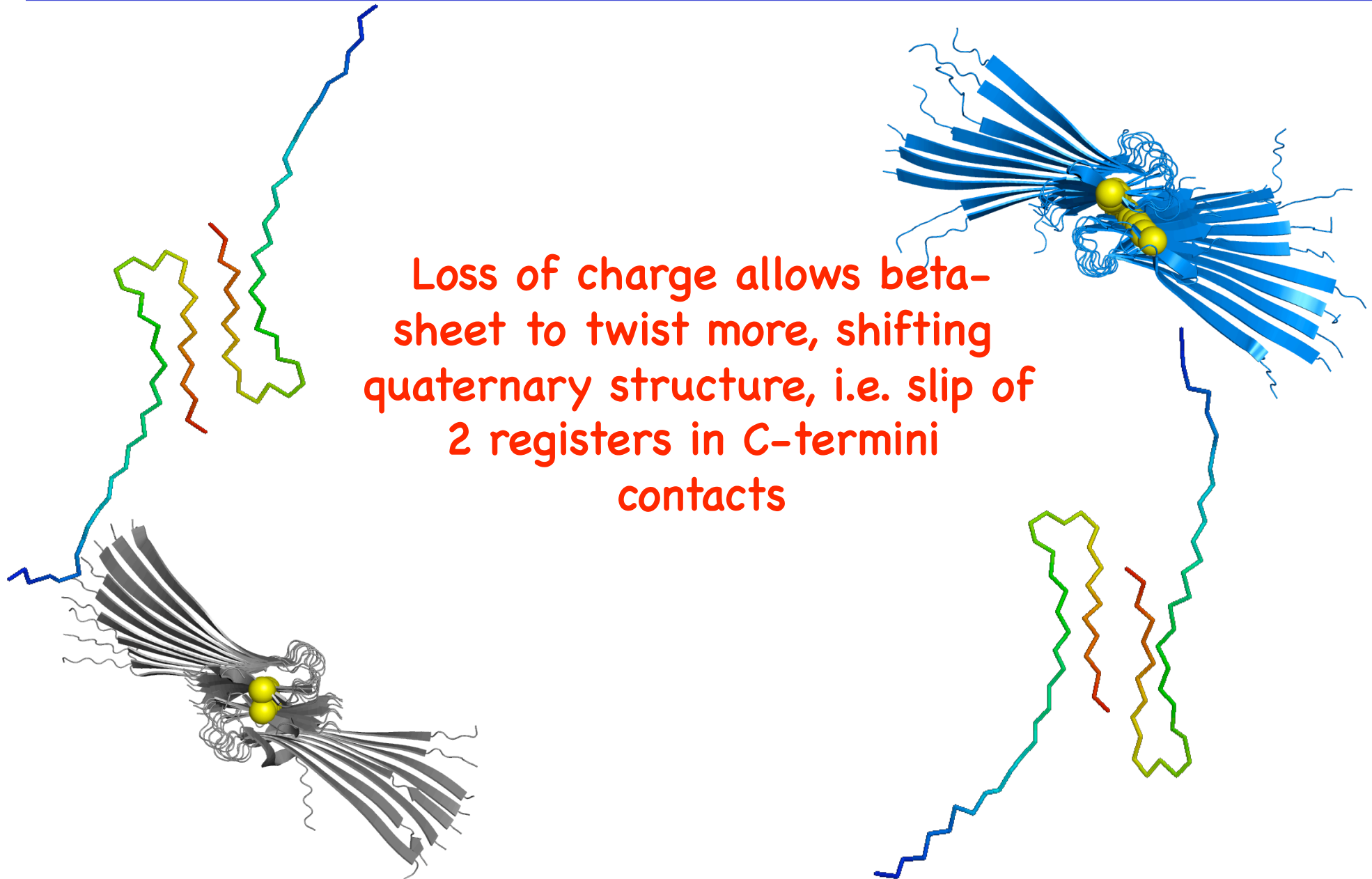
Protofilaments more stable due to loss of charge



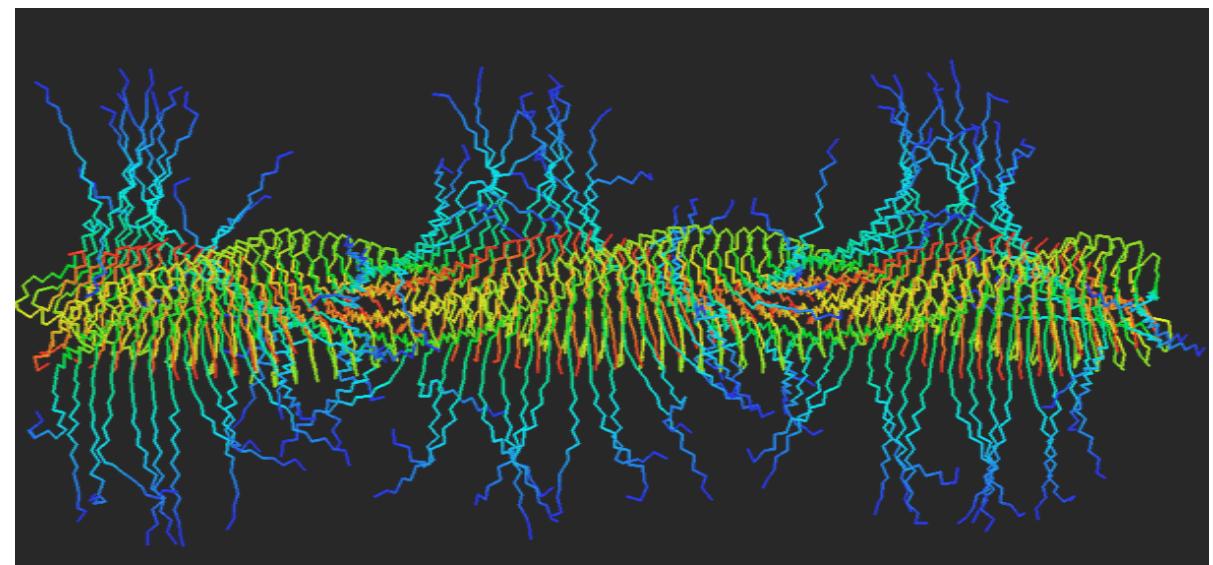
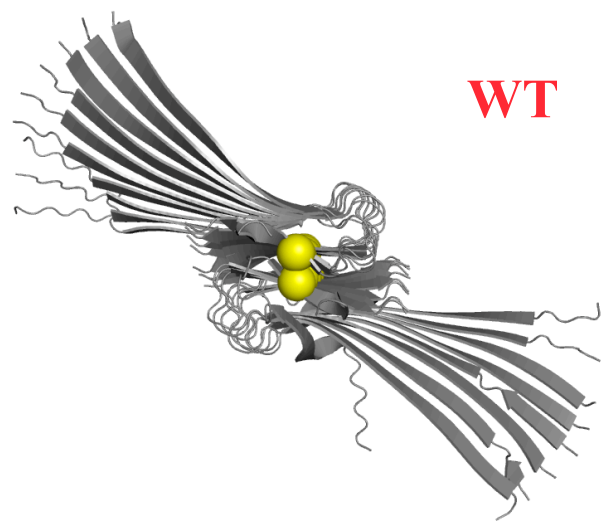
Resulting free energy barrier is increased for protofibril due to appearance of new polymorph

If I make new polymorph the reference, protofibrils stable

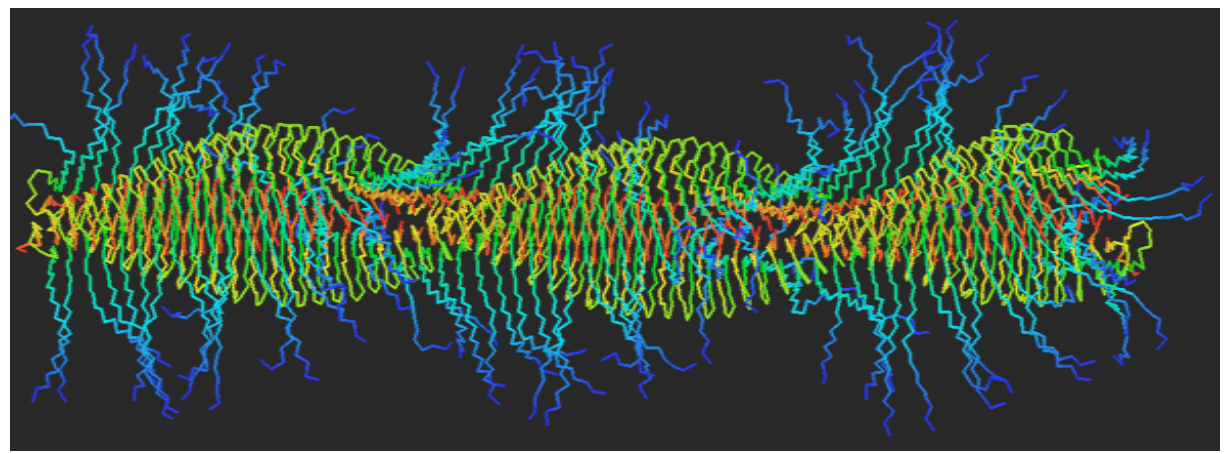
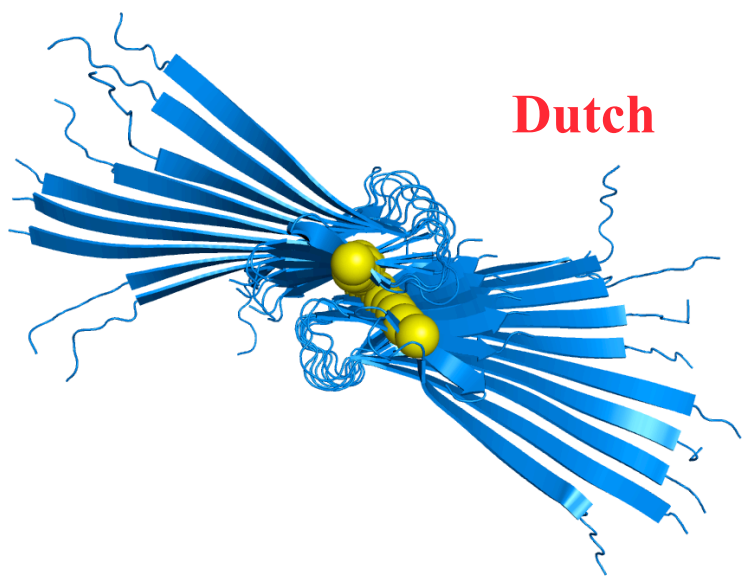
FAD Mutant: Dutch (E22Q)



Polymorphism in our Coarse-Grained Model

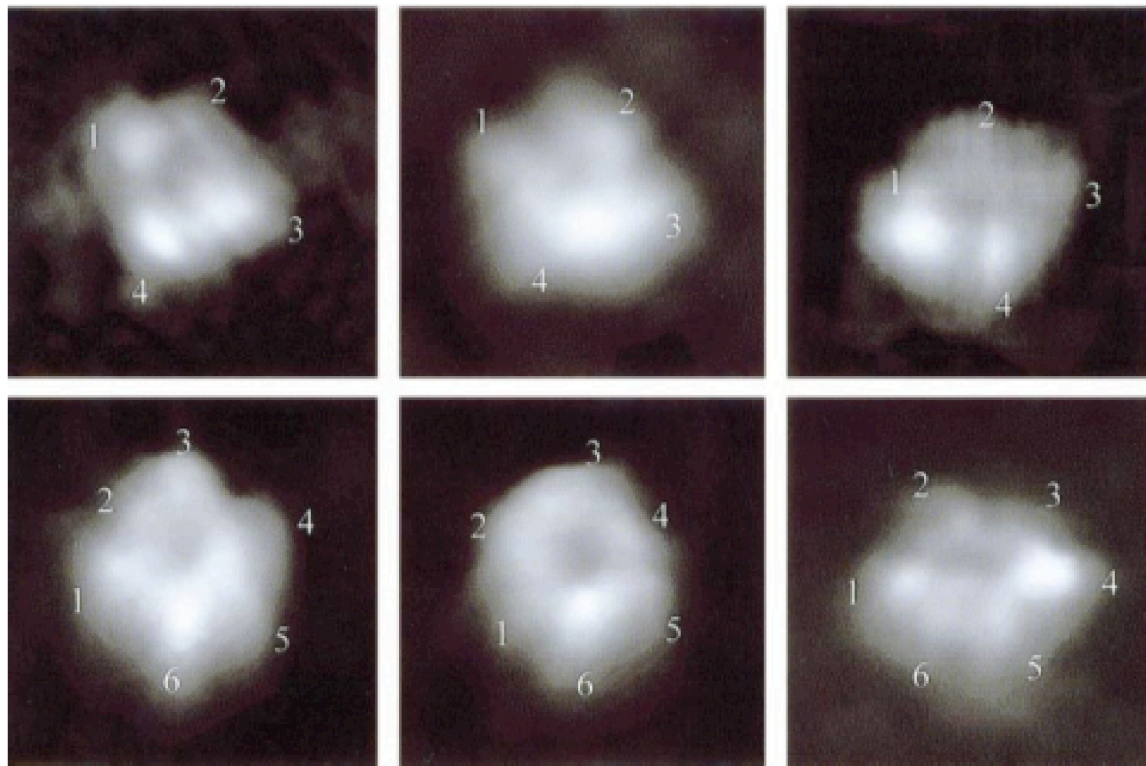


$\sim 0.03\mu$

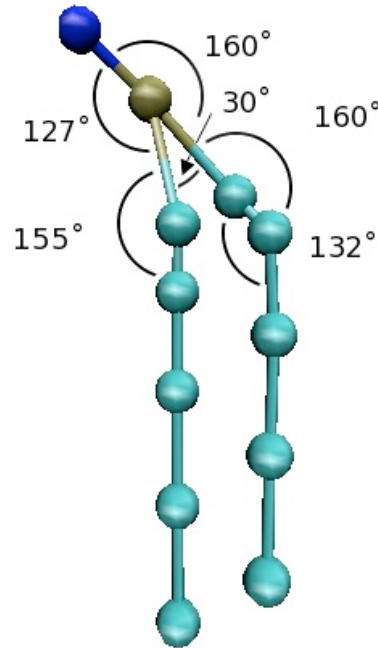


Cytotoxic A β Pores in Membranes

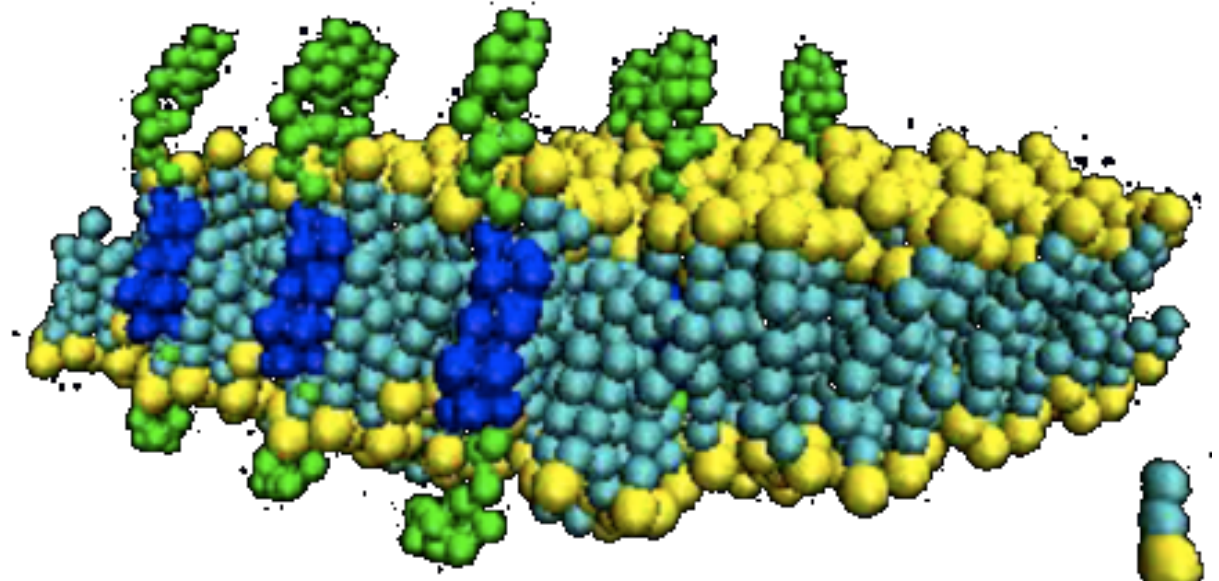
A β_{1-40} and A β_{1-42} can form ion channels that are selective for Ca⁺² in planar lipid bilayers, suggesting that the ability of A β to form such channels upset homeostasis, triggering a signaling cascade for apoptosis and thus amyloid neurotoxicity.



Cytotoxic A β Pores in Membranes

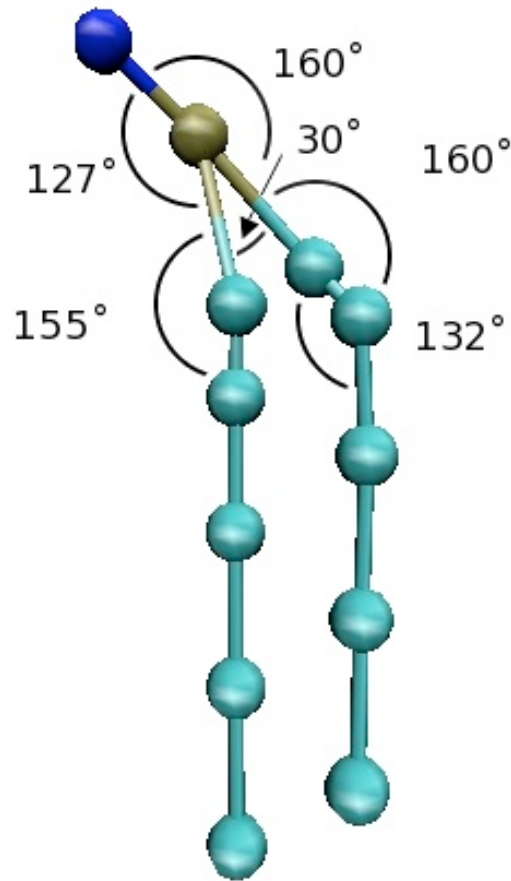


We have formulated a new membrane model by adapting our protein CG model with modified long-range attractions (ala Cooke, Kremer & Deserno) to promote fluid phase of membrane to understand pore formation of A β_{1-40} .



Sodt & THG (2008). In preparation

Coarse-Grained Membrane Model



We take the CG approach that water is dominant cost whose effect on membrane properties can be represented implicitly

In turn we take a far less aggressive strategy on coarse-graining the lipid (DPPC) than other CG membrane models

Approach is generalizable to other lipid types

Sodt & THG (2009). Biophysical J. submitted

Coarse-Grained Membrane Model

Our protein CG model is modified long-range attractions (ala Cooke, Kremer & Deserno) to promote fluid phase of membrane

$$V_{tail-tail}(r) = \begin{cases} 4\varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] & r < r_c \\ -\varepsilon & r_c < r < r_c + w_f \\ 4\varepsilon \left[\left(\frac{\sigma}{r - w_f} \right)^{12} - \left(\frac{\sigma}{r - w_f} \right)^6 \right] & r_c + w_f < r < w_{cut} \end{cases}$$

Sodt & THG (2009). Biophysical J. submitted

Coarse-Grained Membrane Model

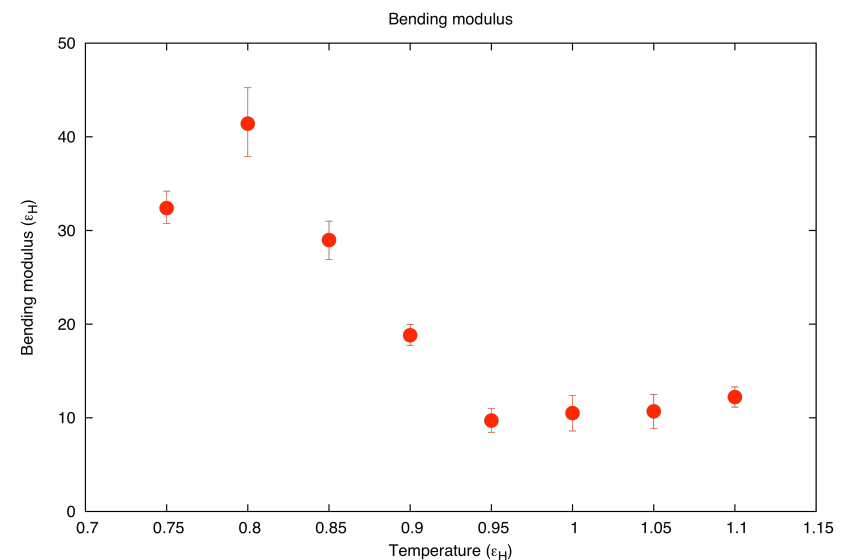
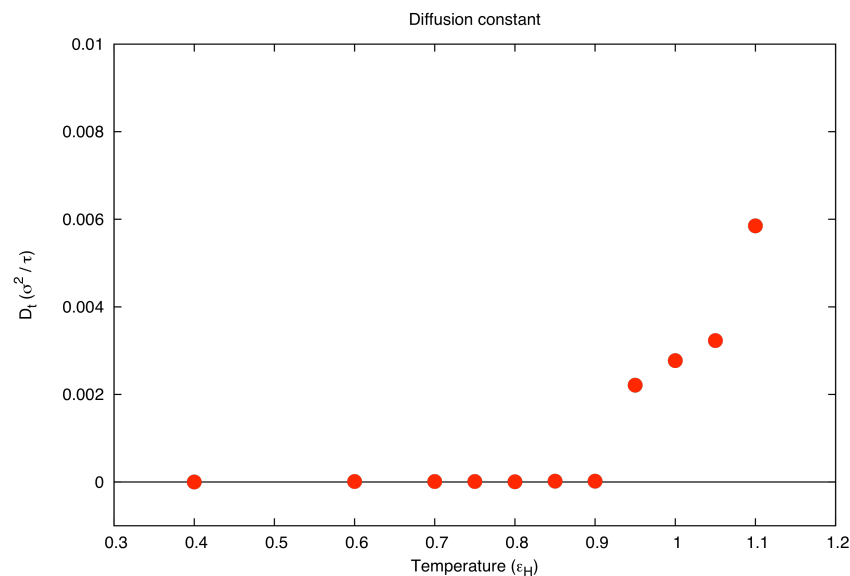
Our protein CG model is modified long-range attractions (ala Cooke, Kremer & Deserno) to promote fluid phase of membrane

$$4\varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] \quad r < r_c$$

$$V_{tail-tail}(r) = \quad -\varepsilon \quad r_c < r < r_c + w_f$$

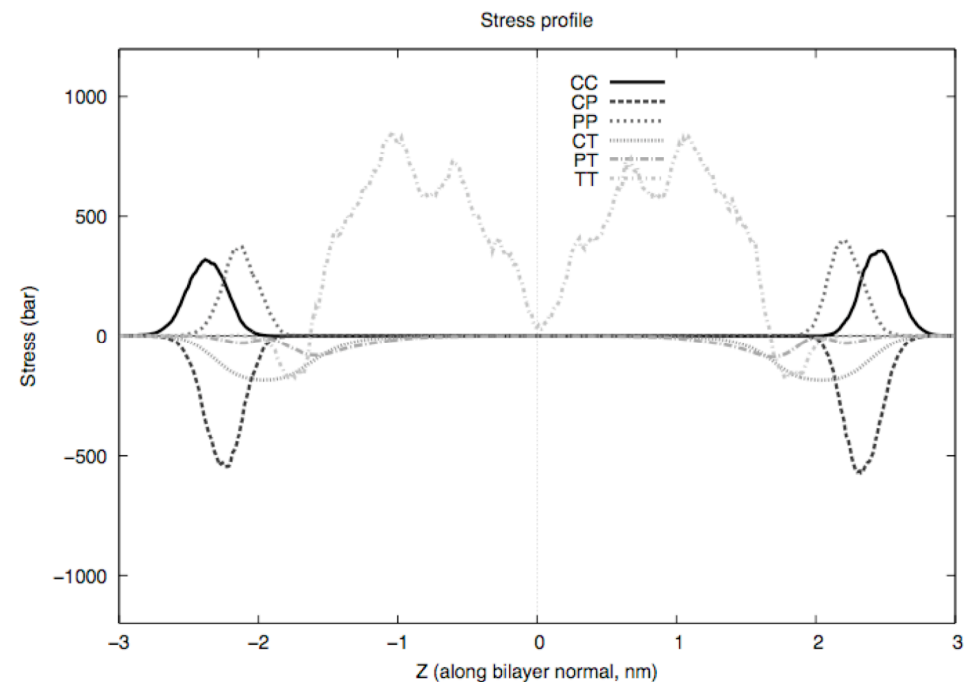
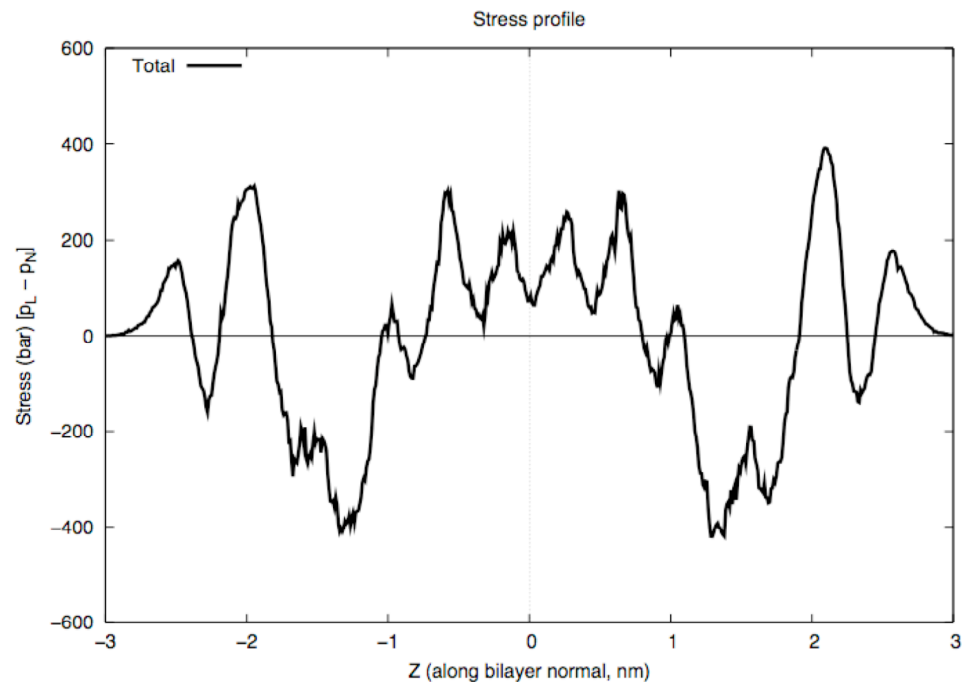
$$4\varepsilon \left[\left(\frac{\sigma}{r - w_f} \right)^{12} - \left(\frac{\sigma}{r - w_f} \right)^6 \right] \quad r_c + w_f < r < w_{cut}$$

Sodt & THG (2009). Biophysical J. submitted



Coarse-Grained Membrane Model

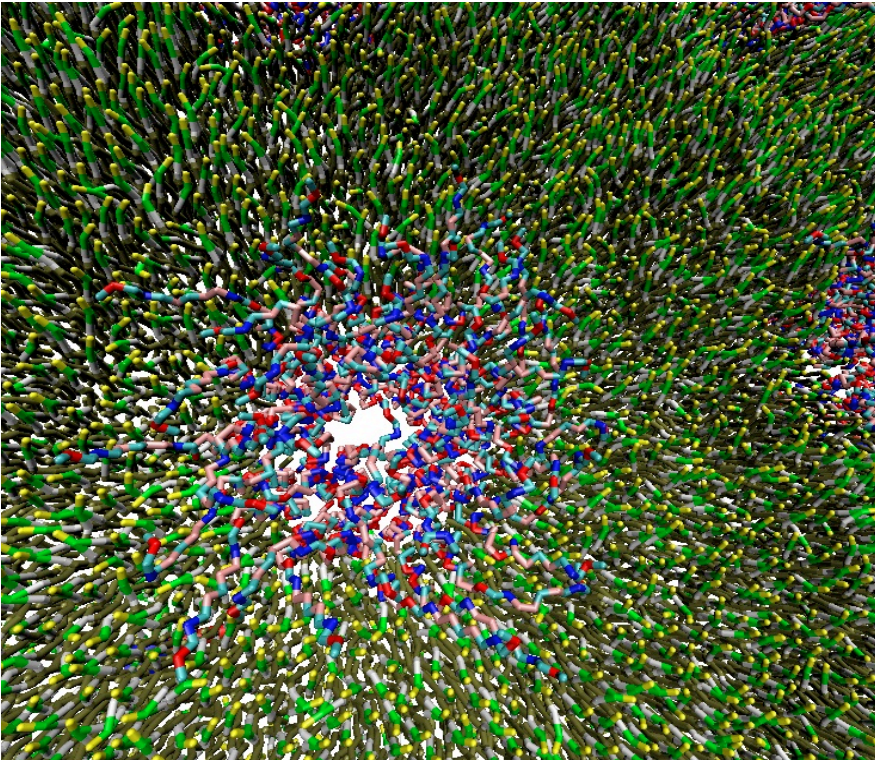
Previous implicit solvent model had trouble getting the stress profile right and getting it right for the right reasons



Our stress profile matches all-atom explicit solvent models very well

Sodt & THG (2009). Biophysical J. submitted

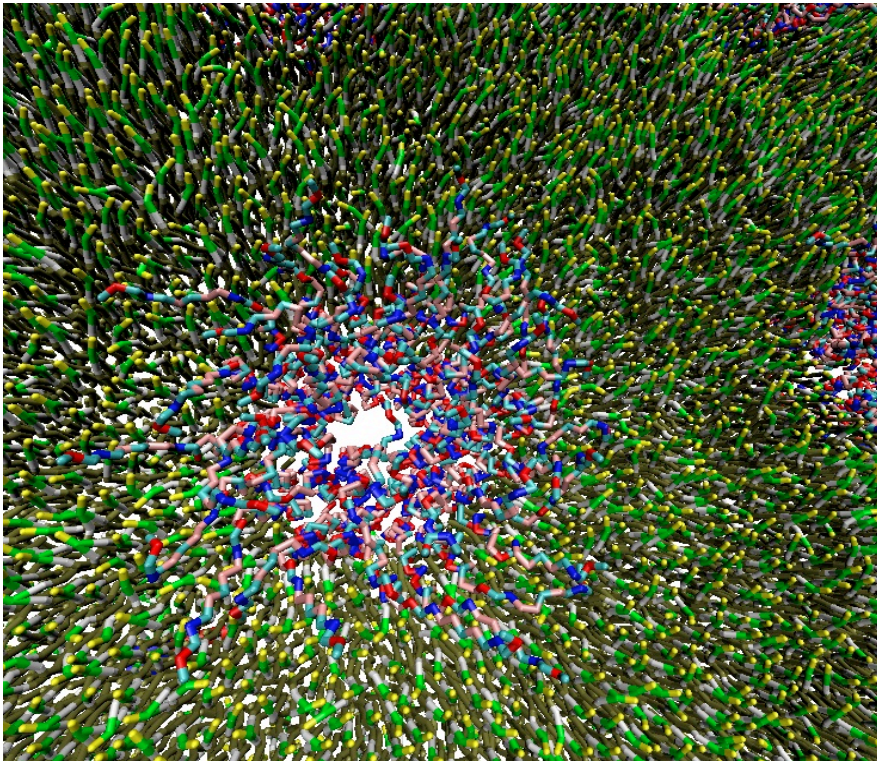
Cytotoxic A β Pores in Membranes



We use our CG membrane with our protein model to understand pore formation of A β_{1-40} .

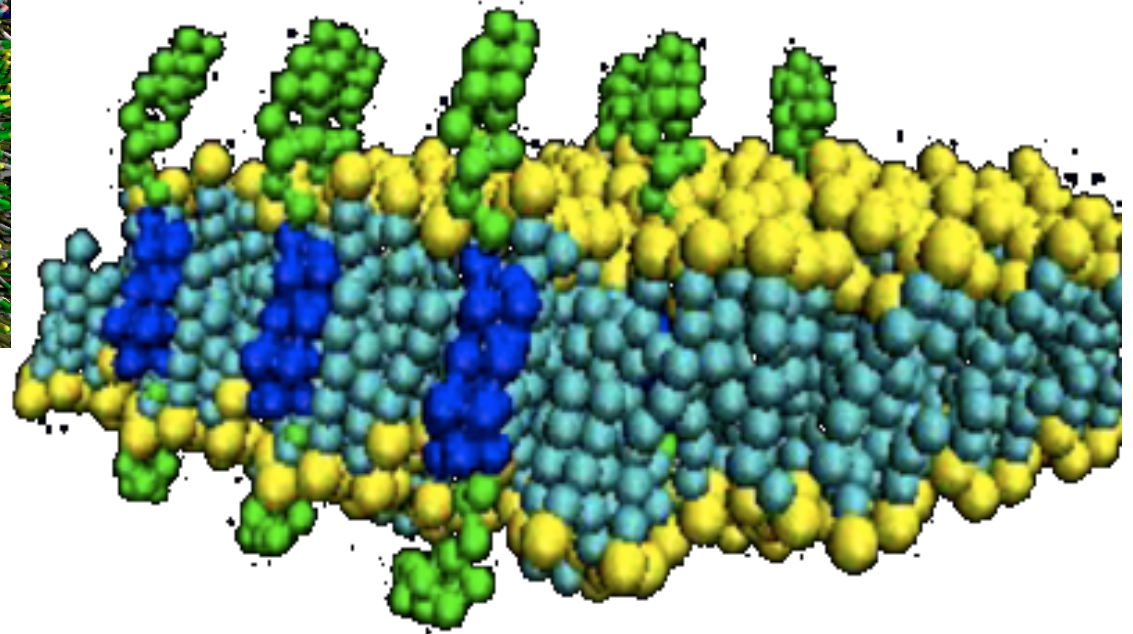
Sodt & THG (2009). In preparation

Cytotoxic A β Pores in Membranes



We use our CG membrane with our protein model to understand pore formation of A β_{1-40} .

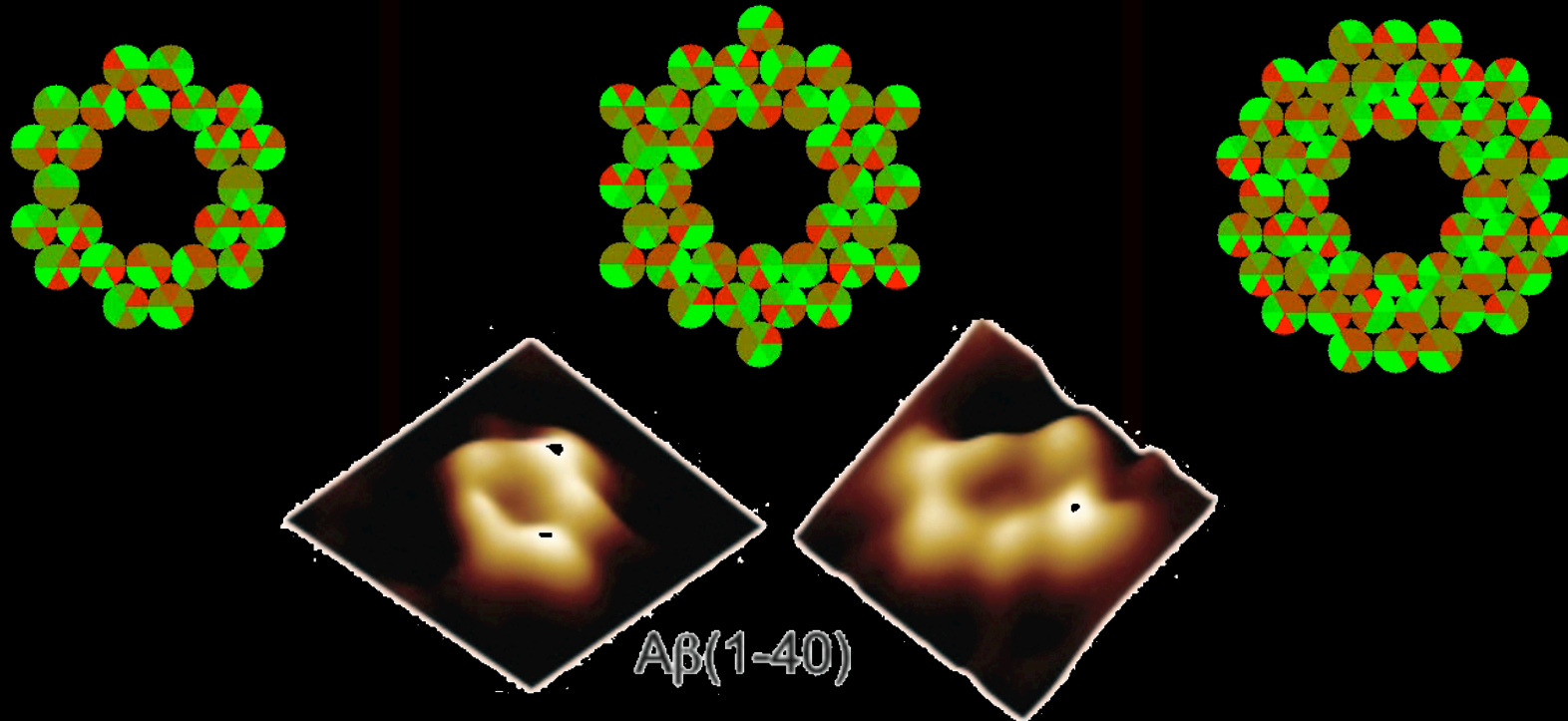
Sodt & THG (2009). In preparation



What α -helical peptide organizations give rise to pore sizes with an inner diameter of 1-2nm and outer diameters of 8-12nm that are stable?

Cytotoxic A β Pores in Membranes

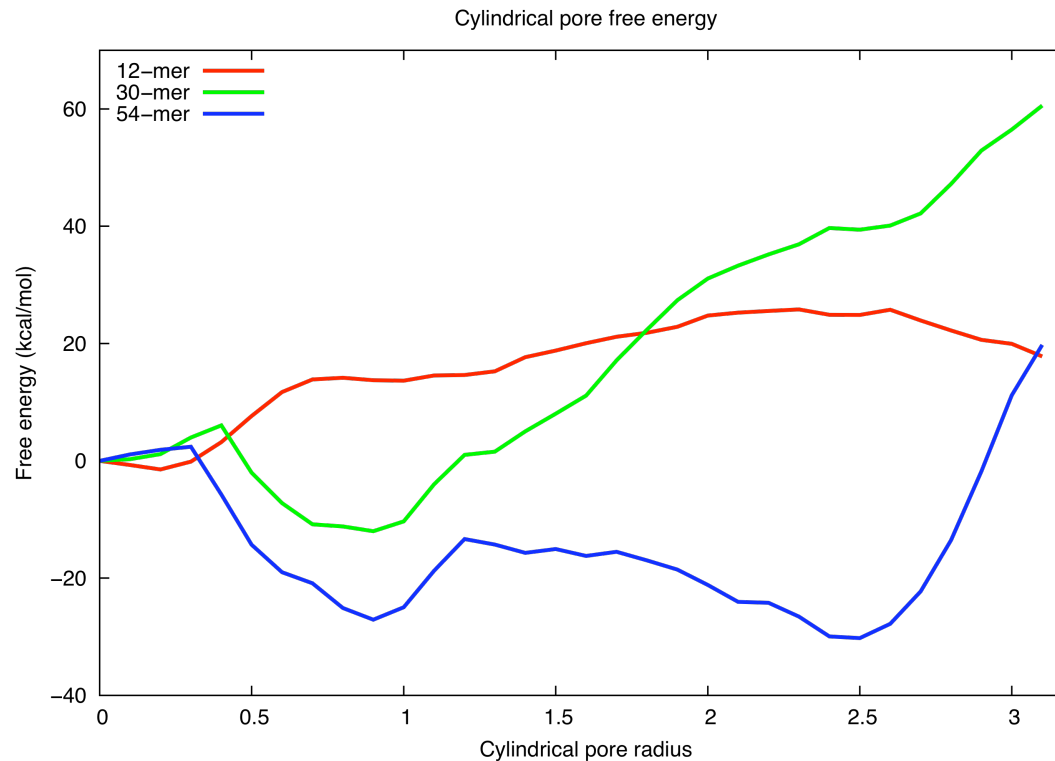
Using coarse-grained model of amphipathic helical peptides in the cell membrane, our working hypothesis is that "hydrophobic matching" influences the monomer assembly and pore stability



Red indicates a face with relatively more hydrophilic residues
Green indicates primarily hydrophobic ones.

Cytotoxic A β Pores in Membranes

Non-equilibrium free energy (BAR) as a function of inner pore diameter



Pores with at least 50 peptides is necessary to stabilize inner pores of 1-2nm seen in AFM/SEM images

Thank You!

I enjoyed my time here in Bangalore!