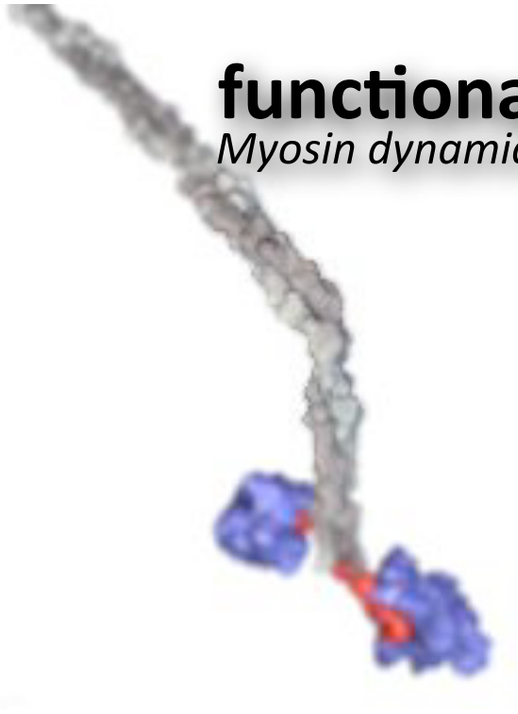
To the ExaScale and Beyond: *Statistically Motivated Linear Scaling for Protein Dynamics*

T. J. Lane

Pande Group (the Folding@home guys)
Department of Chemistry
Stanford University

Dynamics is often critical to function

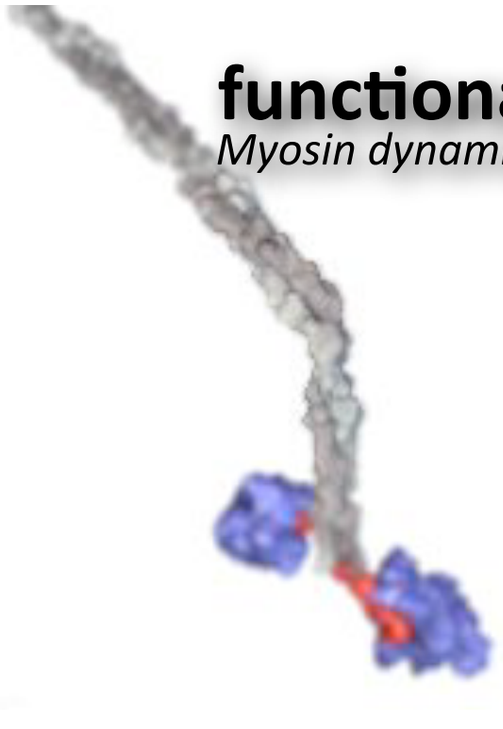
Dynamics is often critical to function



functional motion

Myosin dynamics from Vale lab

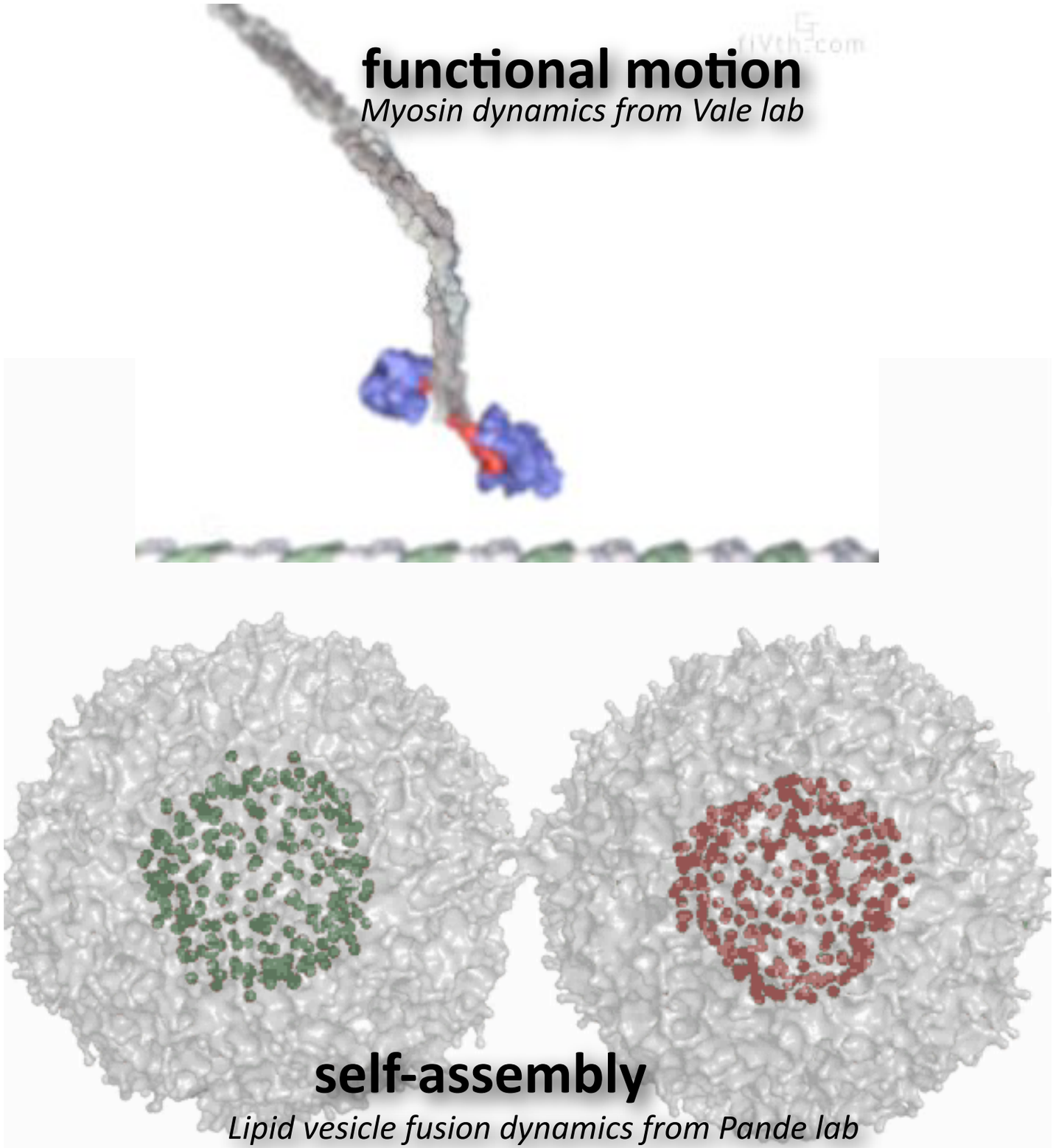
Dynamics is often critical to function



functional motion

Myosin dynamics from Vale lab

Dynamics is often critical to function

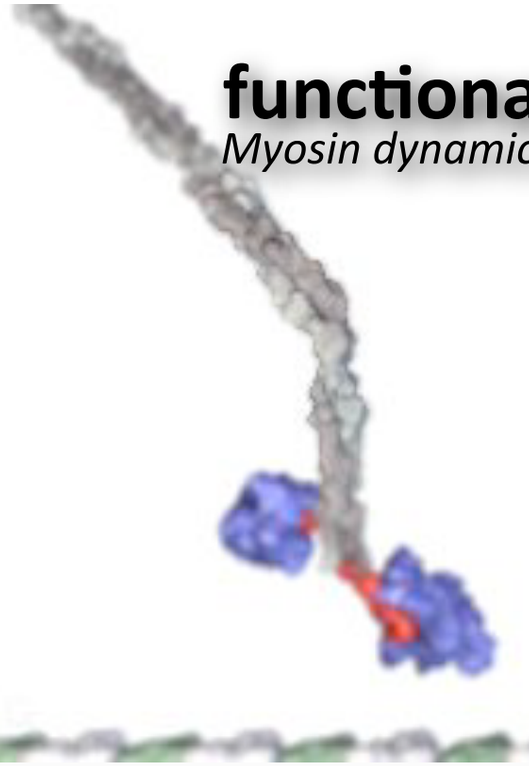


functional motion
Myosin dynamics from Vale lab

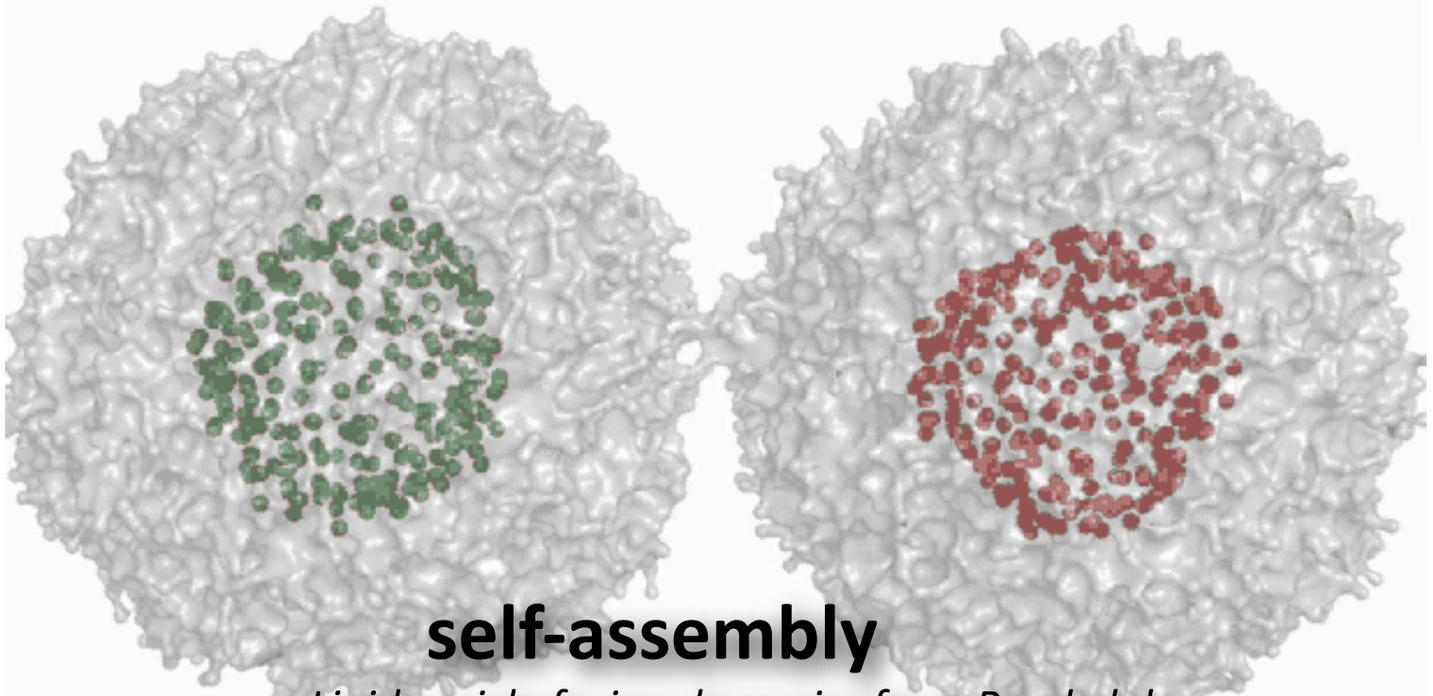
self-assembly
Lipid vesicle fusion dynamics from Pande lab

allostery
GroEL dynamics from Saibil lab

Dynamics is often critical to function

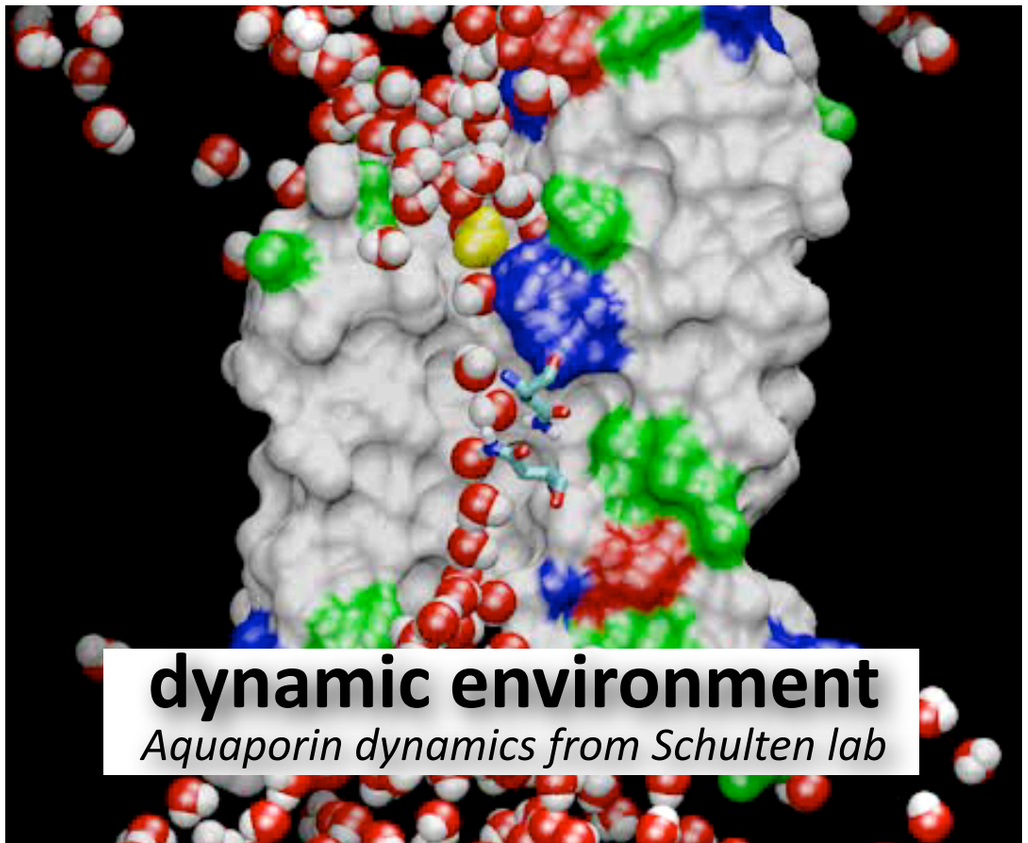


functional motion
Myosin dynamics from Vale lab



self-assembly
Lipid vesicle fusion dynamics from Pande lab

allostery
GroEL dynamics from Saibil lab



dynamic environment
Aquaporin dynamics from Schulten lab

What is protein folding?

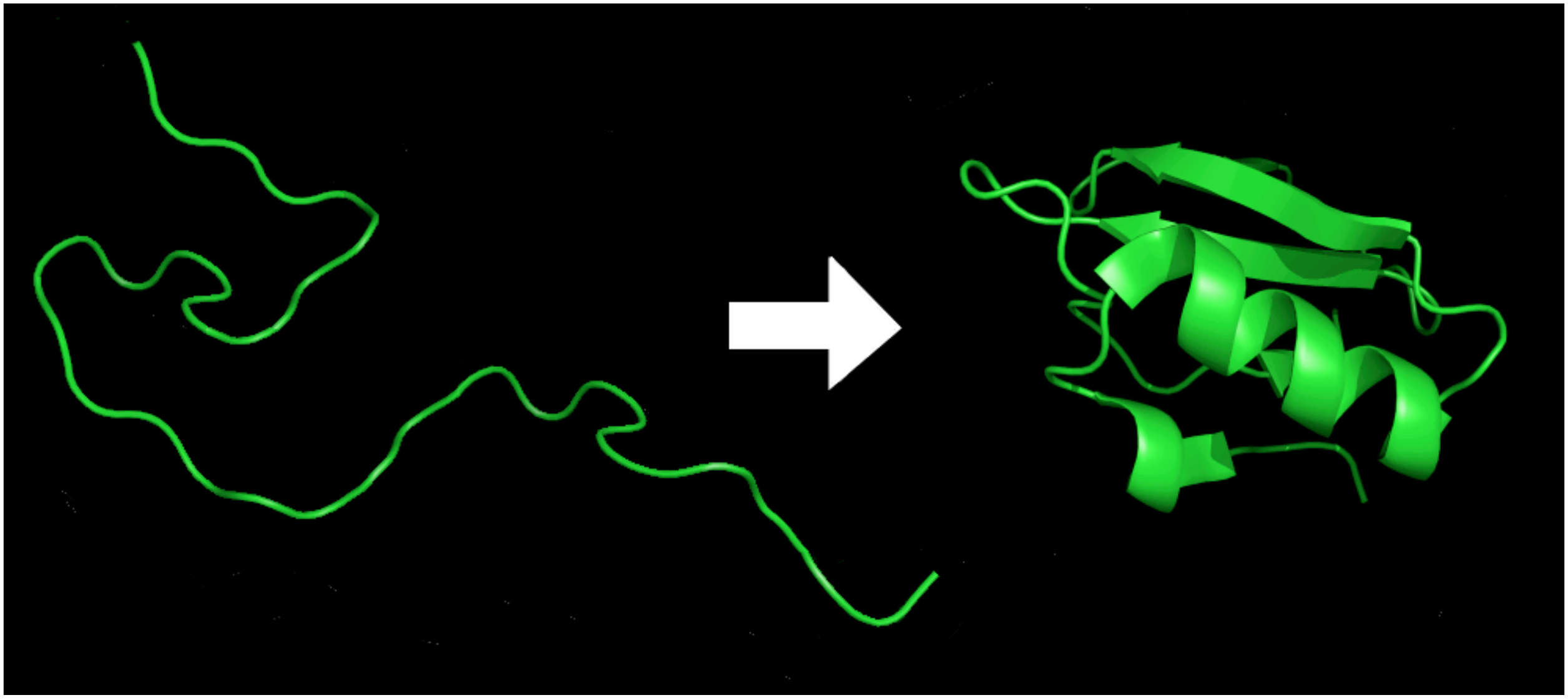
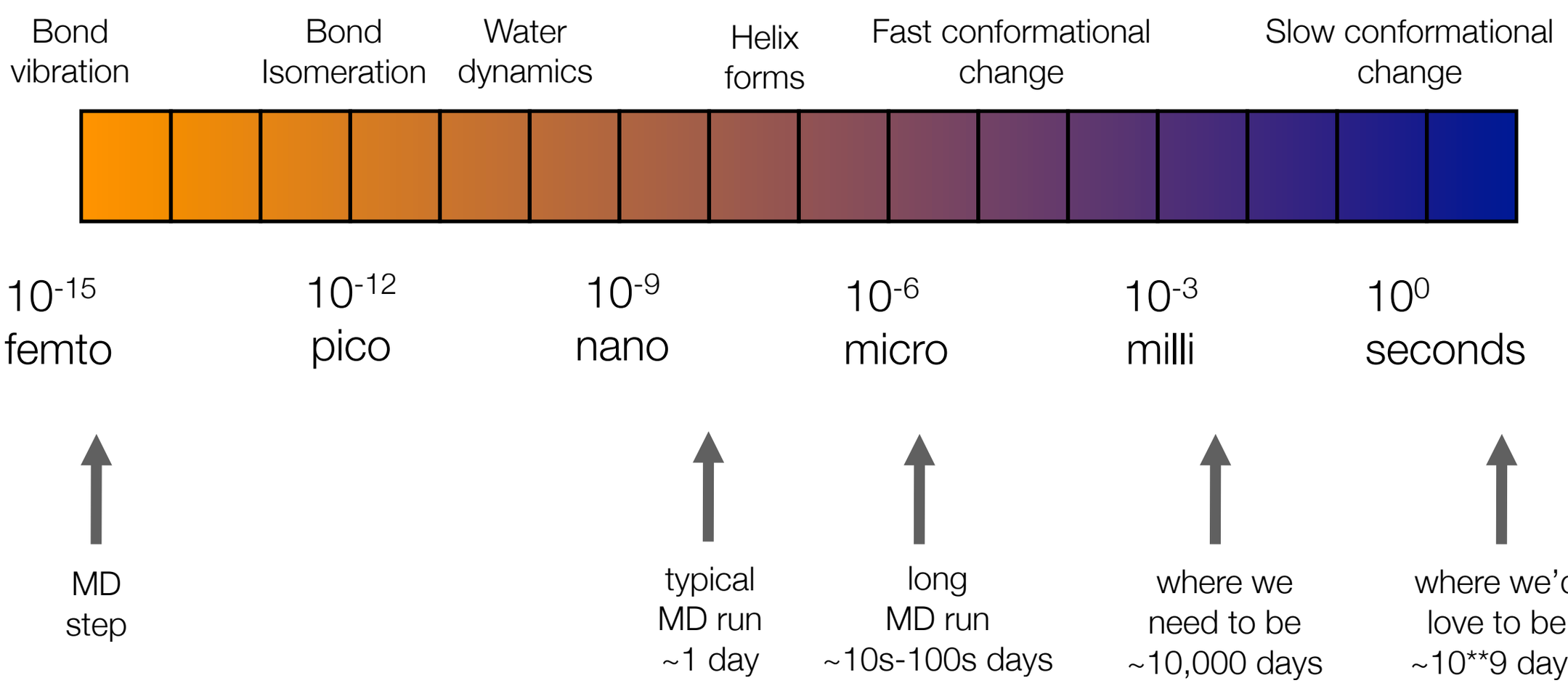


image: wikipedia

The nightmare of simulating long time scales



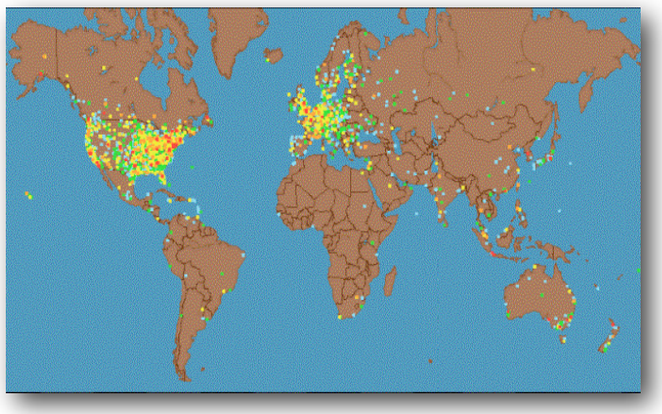
How do you break a billion-fold impasse?

Combine multiple, powerful, complementary technologies

How do you break a billion-fold impasse?

Combine multiple, powerful, complementary technologies

1) Folding@home:
very large-scale
distributed computing



Most powerful computer cluster in the world (~8 petaflops)

10^4x to 10^5x

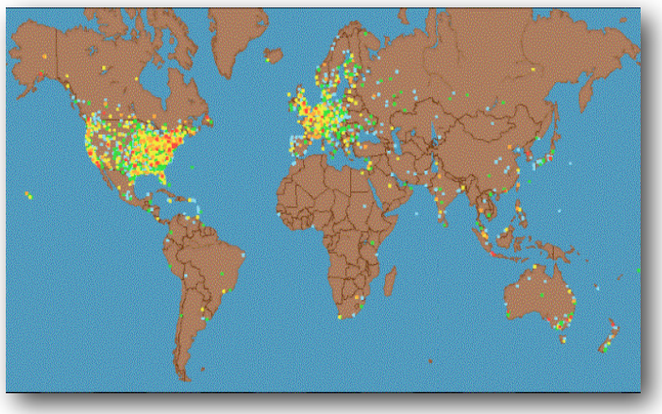
http://folding.stanford.edu

Voelz, *et al*, *JACS* (2010)
Ensign *et al*, *JMB (2007)*
Shirts and Pande, *Science* (2000)

How do you break a billion-fold impasse?

Combine multiple, powerful, complementary technologies

1) Folding@home:
very large-scale
distributed computing



AlMost powerful computer cluster in the world (~8 petaflops)

10^4x to 10^5x

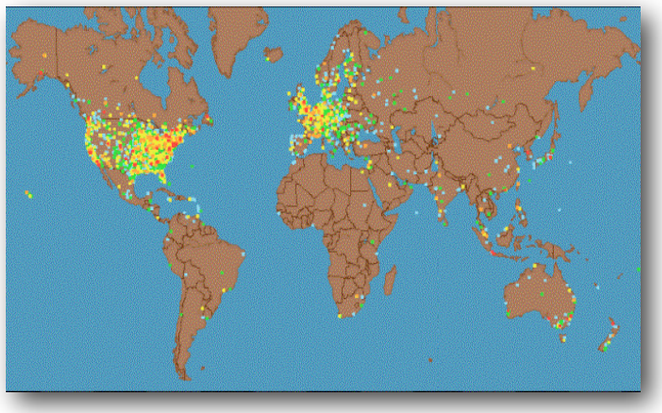
http://folding.stanford.edu

Voelz, *et al*, *JACS* (2010)
Ensign *et al*, *JMB (2007)*
Shirts and Pande, *Science* (2000)

How do you break a billion-fold impasse?

Combine multiple, powerful, complementary technologies

1) Folding@home:
very large-scale
distributed computing



AlMost powerful computer cluster in the world (~8 petaflops)

10^4x to 10^5x

http://folding.stanford.edu

Voelz, *et al*, *JACS* (2010)
Ensign *et al*, *JMB (2007)*
Shirts and Pande, *Science* (2000)

2) OpenMM: Very
fast MD (~1µs/day)
on GPUs

~1µs/day for implicit solvent simulaton of small proteins (~40aa)

10^2x to 10^3x

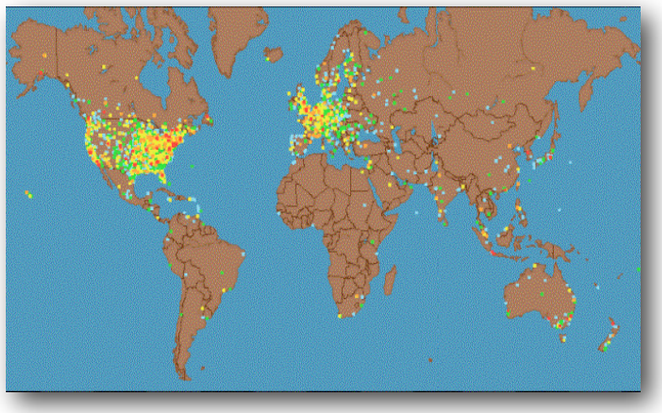
http://simtk.org/home/openmm

Elsen, *et al*. *ACM/IEEE conf. on Supercomputing* (2006)
Friedrichs, *et al*. *J. Comp. Chem*., (2009)
Eastman and Pande. *J. Comp. Chem.* (2009)

How do you break a billion-fold impasse?

Combine multiple, powerful, complementary technologies

1) Folding@home: very large-scale distributed computing



AlMost powerful computer cluster in the world (~8 petaflops)

10^4x to 10^5x

http://folding.stanford.edu

Voelz, *et al*, *JACS* (2010)
Ensign *et al*, *JMB (2007)*
Shirts and Pande, *Science* (2000)

2) OpenMM: Very fast MD (~1µs/day) on GPUs

~1µs/day for implicit solvent simulaton of small proteins (~40aa)

10^2x to 10^3x

http://simtk.org/home/openmm

Elsen, *et al*. *ACM/IEEE conf. on Supercomputing* (2006)
Friedrichs, *et al*. *J. Comp. Chem*., (2009)
Eastman and Pande. *J. Comp. Chem.* (2009)

3) MSMBuilder: Statistical mechanics of many trajectories

very long timescale dynamics by combining many simulations

10^2x to 10^3x

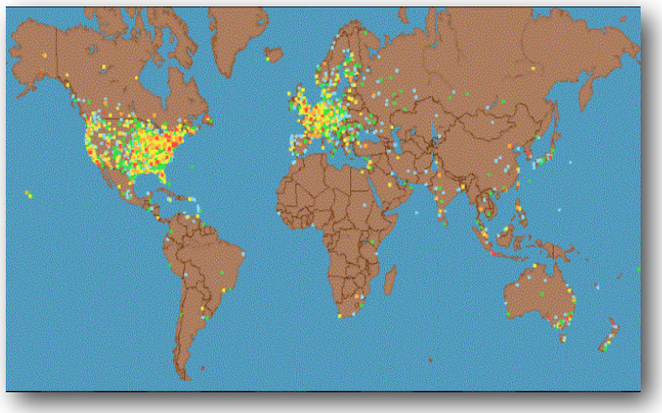
http://simtk.org/home/msmbuilder

Bowman, *et al*, *J. Chem. Phys.* (2009)
Singhal & Pande, *J. Chem. Phys.* (2005)
Singhal, *et al, J. Chem. Phys.* (2004)

How do you break a billion-fold impasse?

Combine multiple, powerful, complementary technologies

1) Folding@home:
very large-scale distributed computing



AlMost powerful computer cluster in the world (~8 petaflops)

10^4x to 10^5x

http://folding.stanford.edu

Voelz, *et al*, *JACS* (2010)
Ensign *et al*, *JMB (2007)*
Shirts and Pande, *Science* (2000)

2) OpenMM: Very fast MD (~1µs/day) on GPUs

~1µs/day for implicit solvent simulaton of small proteins (~40aa)

10^2x to 10^3x

http://simtk.org/home/openmm

Elsen, *et al*. *ACM/IEEE conf. on Supercomputing* (2006)
Friedrichs, *et al*. *J. Comp. Chem*., (2009)
Eastman and Pande. *J. Comp. Chem.* (2009)

3) MSMBuilder:
Statistical mechanics of many trajectories

very long timescale dynamics by combining many simulations

10^2x to 10^3x

http://simtk.org/home/msmbuilder

Bowman, *et al*, *J. Chem. Phys.* (2009)
Singhal & Pande, *J. Chem. Phys.* (2005)
Singhal, *et al, J. Chem. Phys.* (2004)

“States and rates” is a familiar paradigm

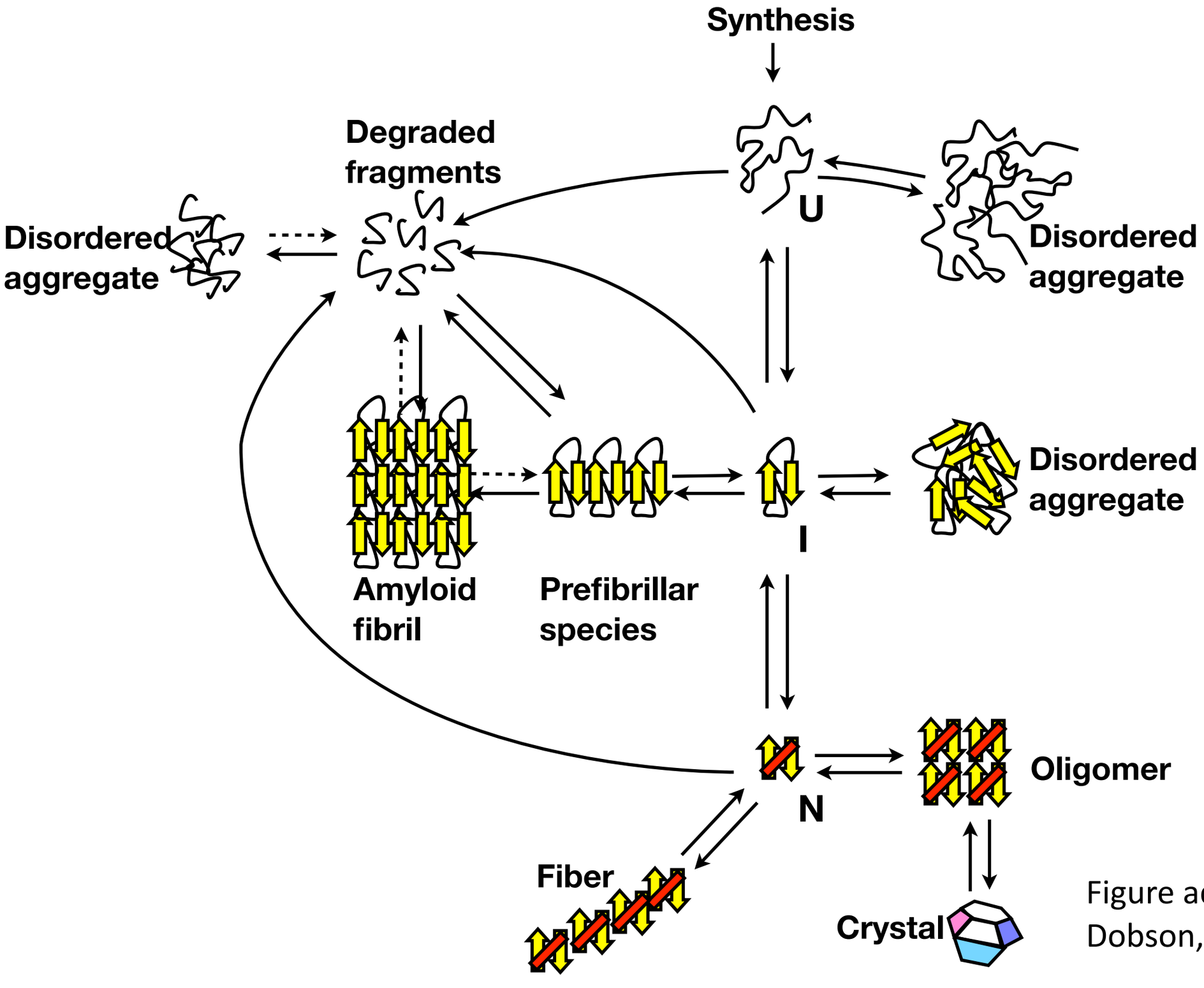


Figure adapted from Dobson, et al, *Nature*

MSMs coarse grain conformation space (to ~3Å) to build a Master equation

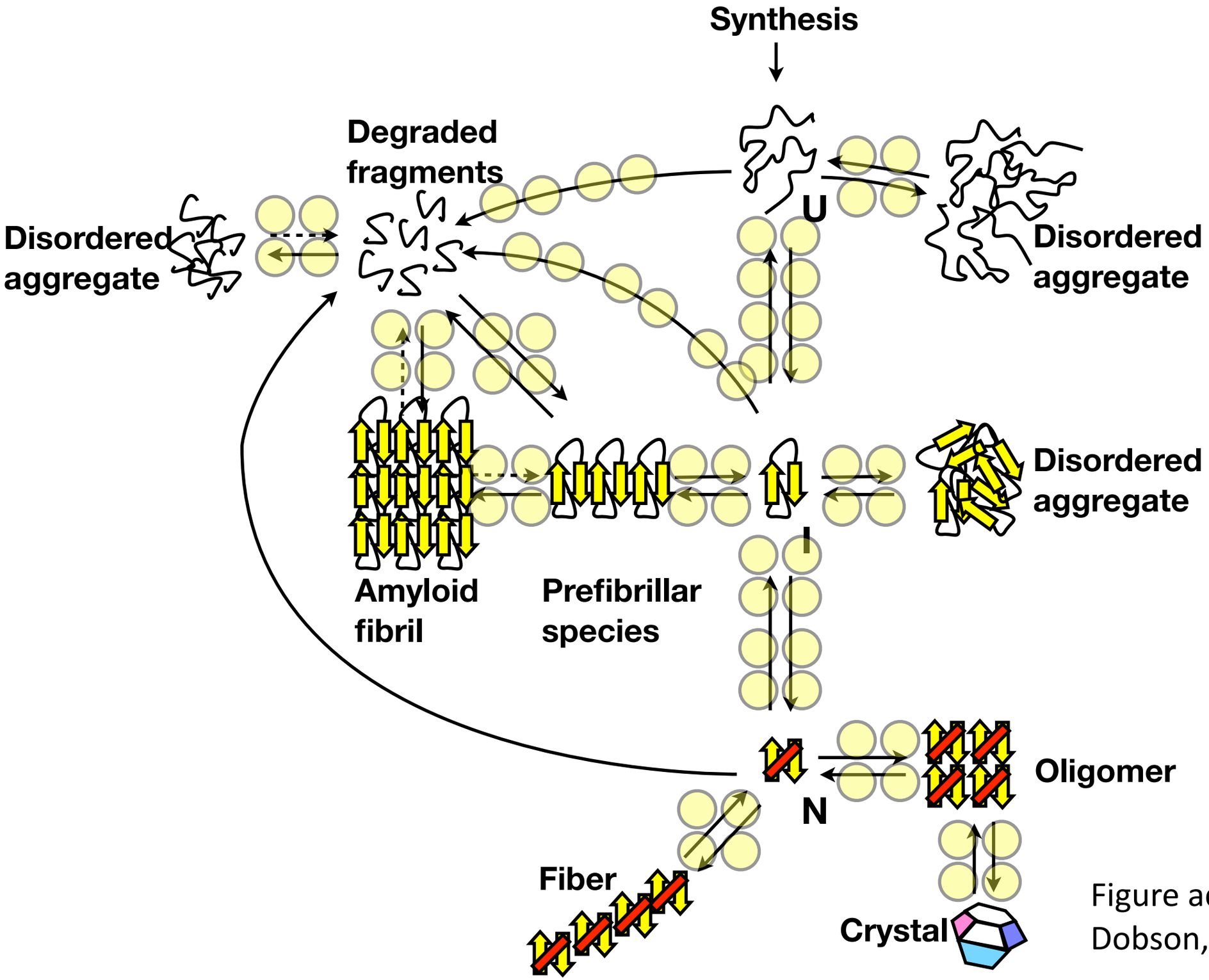


Figure adapted from Dobson, et al, *Nature*

but also derive a coarser view for human consumption

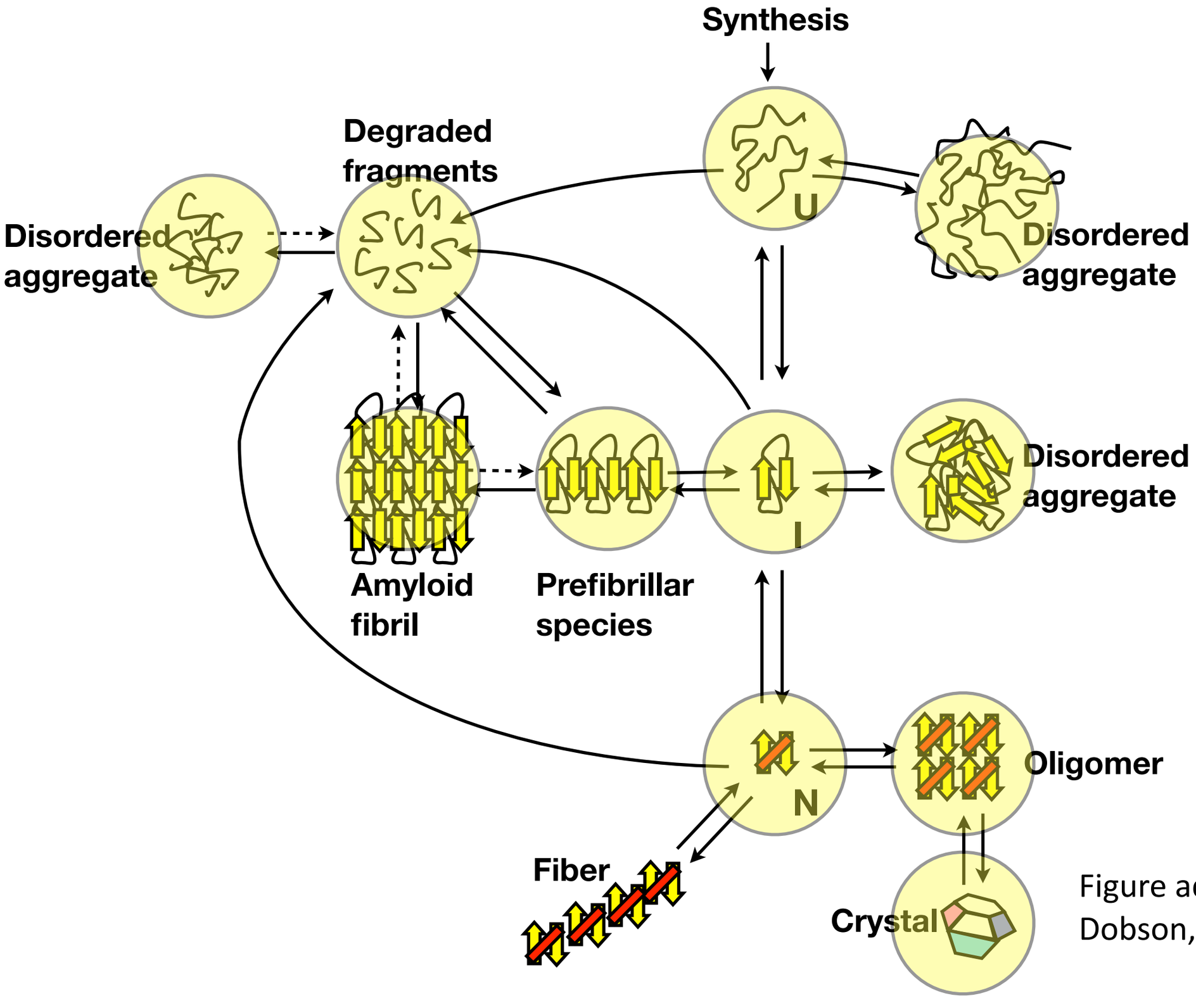
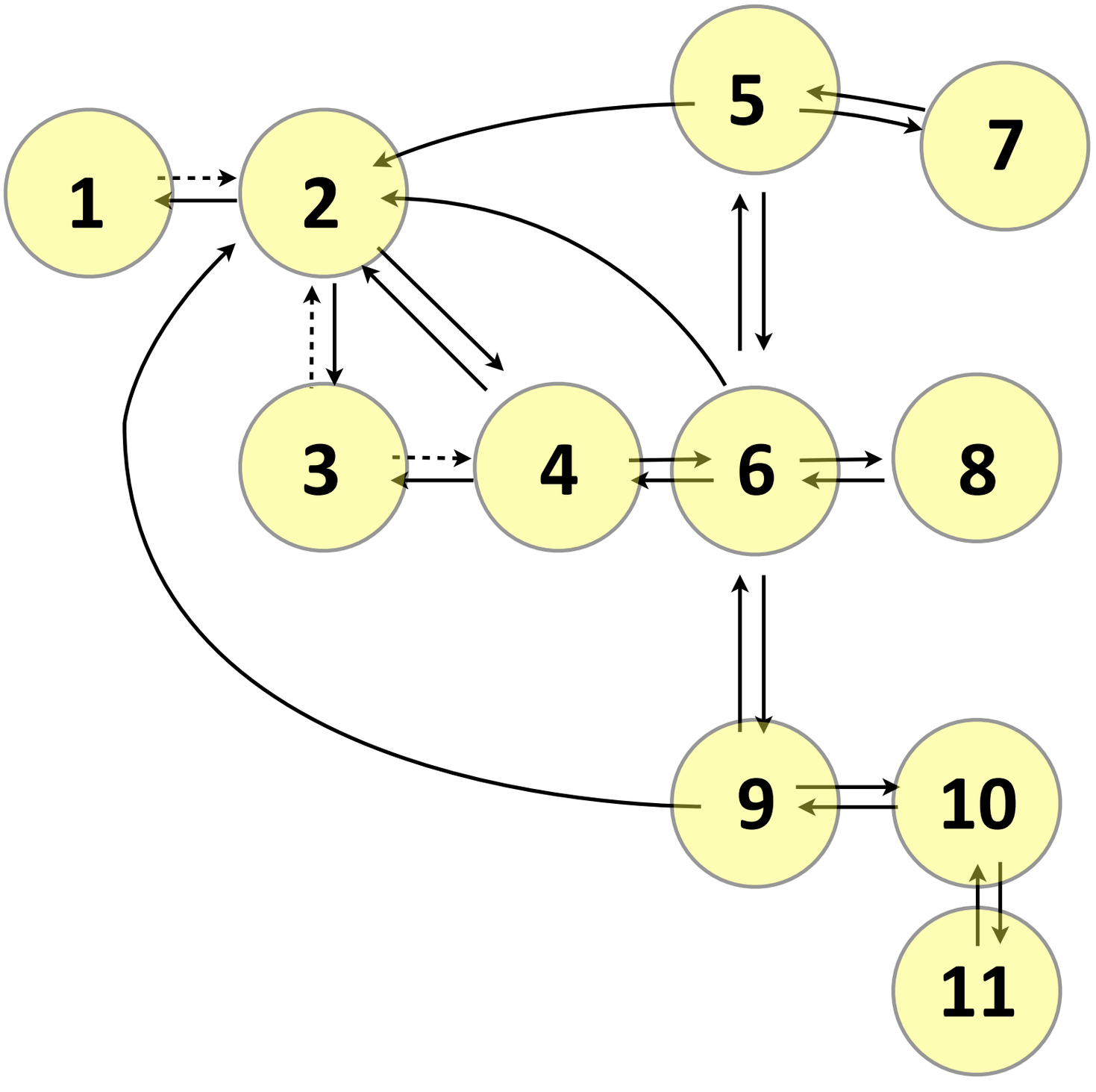


Figure adapted from Dobson, et al, *Nature*

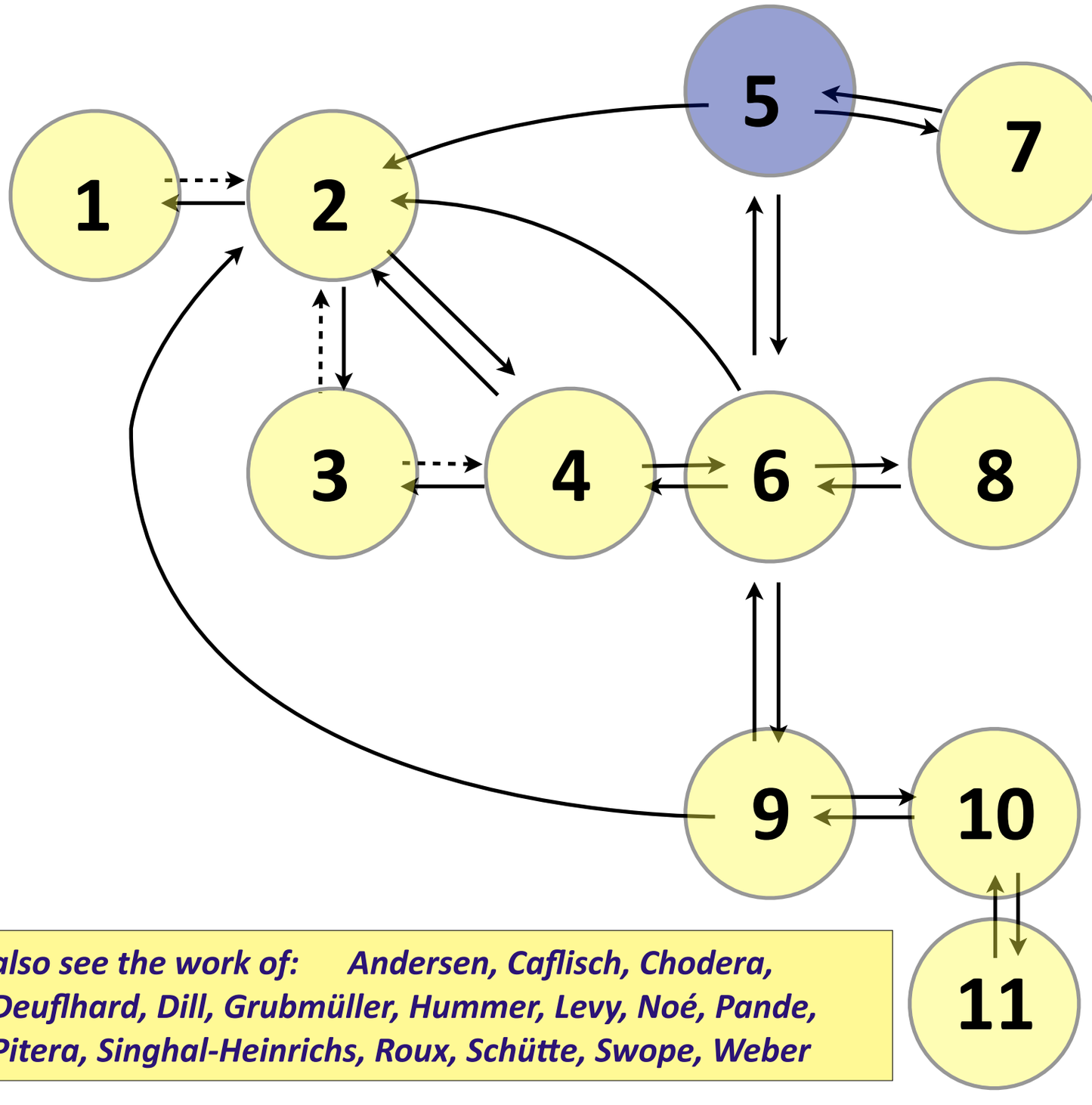
How does one build an MSM?



How does one build an MSM? MSMBuilder

Use Active Learning-like adaptive methods to infer p_{ij}

http://simtk.org/home/msmbuilder



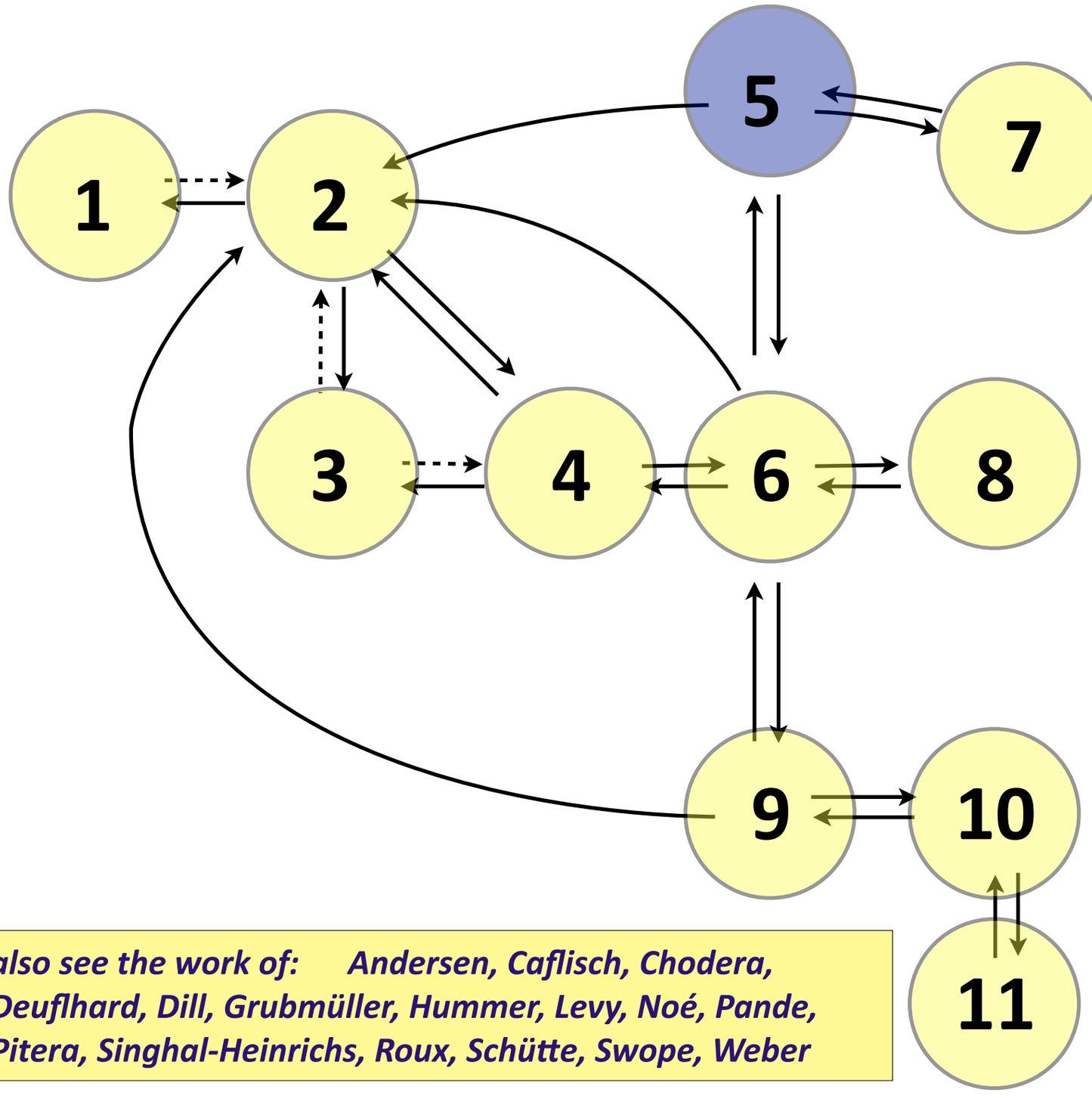
also see the work of: Andersen, Caflisch, Chodera, Deuflhard, Dill, Grubmüller, Hummer, Levy, Noé, Pande, Pitera, Singhal-Heinrichs, Roux, Schütte, Swope, Weber

How does one build an MSM? MSMBuilder

Use Active Learning-like adaptive methods to infer p_{ij}

1. **Kinetic clustering:** structures are combined into a state if they rapidly interconvert

http://simtk.org/home/msmbuilder



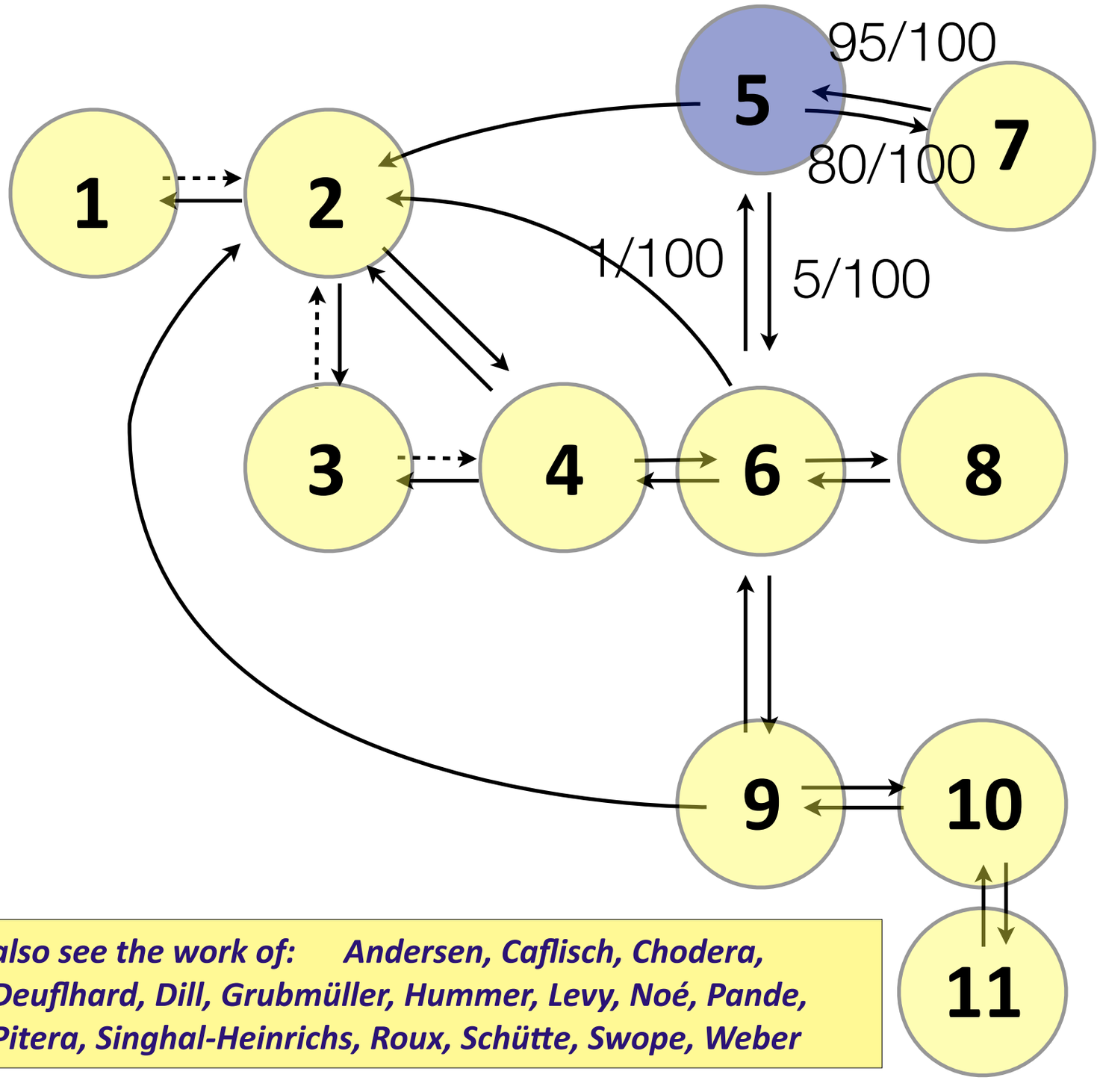
also see the work of: Andersen, Caflisch, Chodera, Deuflhard, Dill, Grubmüller, Hummer, Levy, Noé, Pande, Pitera, Singhal-Heinrichs, Roux, Schütte, Swope, Weber

How does one build an MSM? MSMBuilder

Use Active Learning-like adaptive methods to infer p_{ij}

1. **Kinetic clustering:** structures are combined into a state if they rapidly interconvert

http://simtk.org/home/msmbuilder



also see the work of: Andersen, Caflisch, Chodera, Deuflhard, Dill, Grubmüller, Hummer, Levy, Noé, Pande, Pitera, Singhal-Heinrichs, Roux, Schütte, Swope, Weber

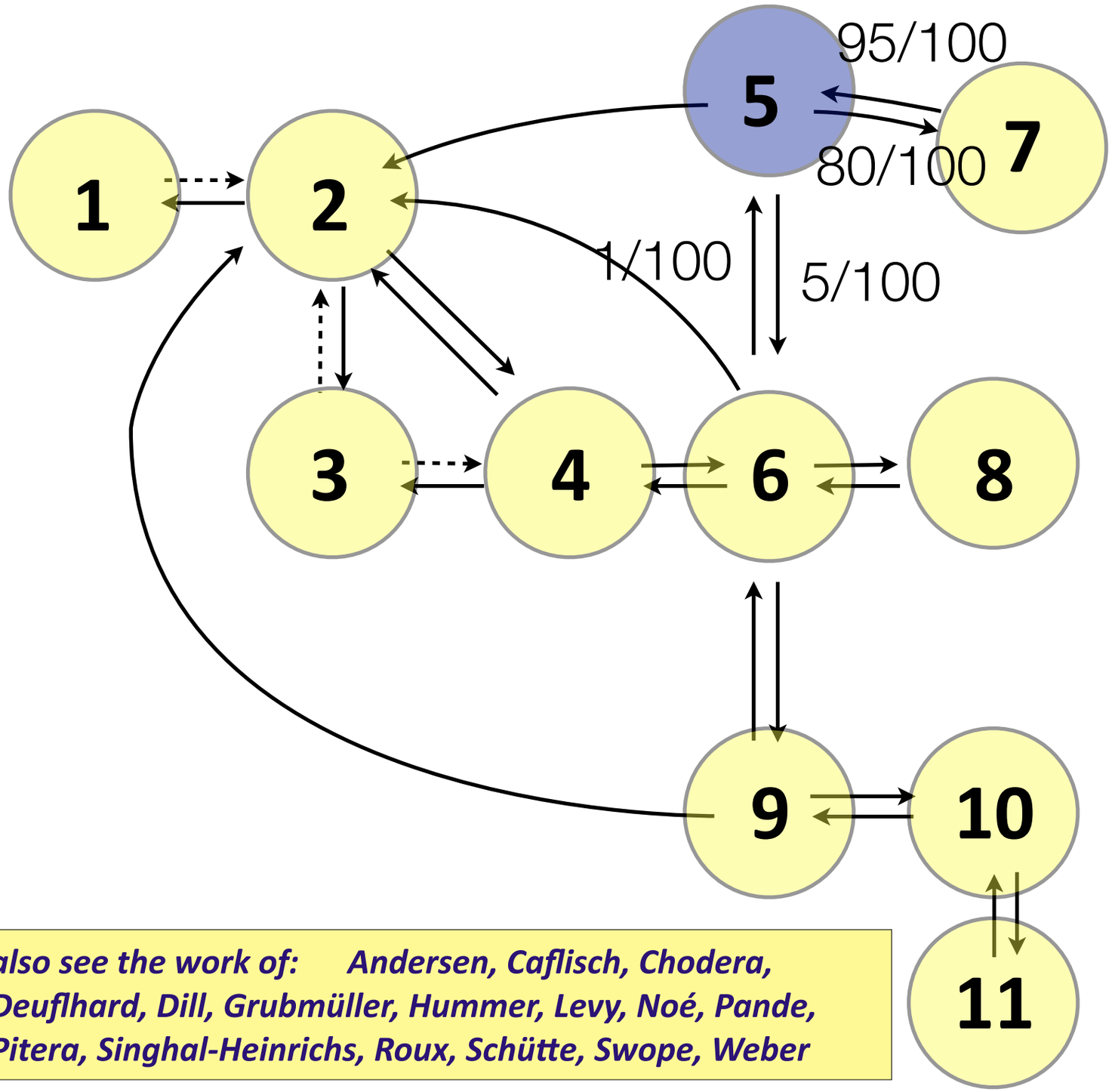
How does one build an MSM? MSMBuilder

Use Active Learning-like adaptive methods to infer p_{ij}

1. **Kinetic clustering:** structures are combined into a state if they rapidly interconvert

2. **Build the best model possible given the data:** Bayesian methods to infer p_{ij} from transition counts

http://simtk.org/home/msmbuilder



also see the work of: Andersen, Caflisch, Chodera, Deuflhard, Dill, Grubmüller, Hummer, Levy, Noé, Pande, Pitera, Singhal-Heinrichs, Roux, Schütte, Swope, Weber

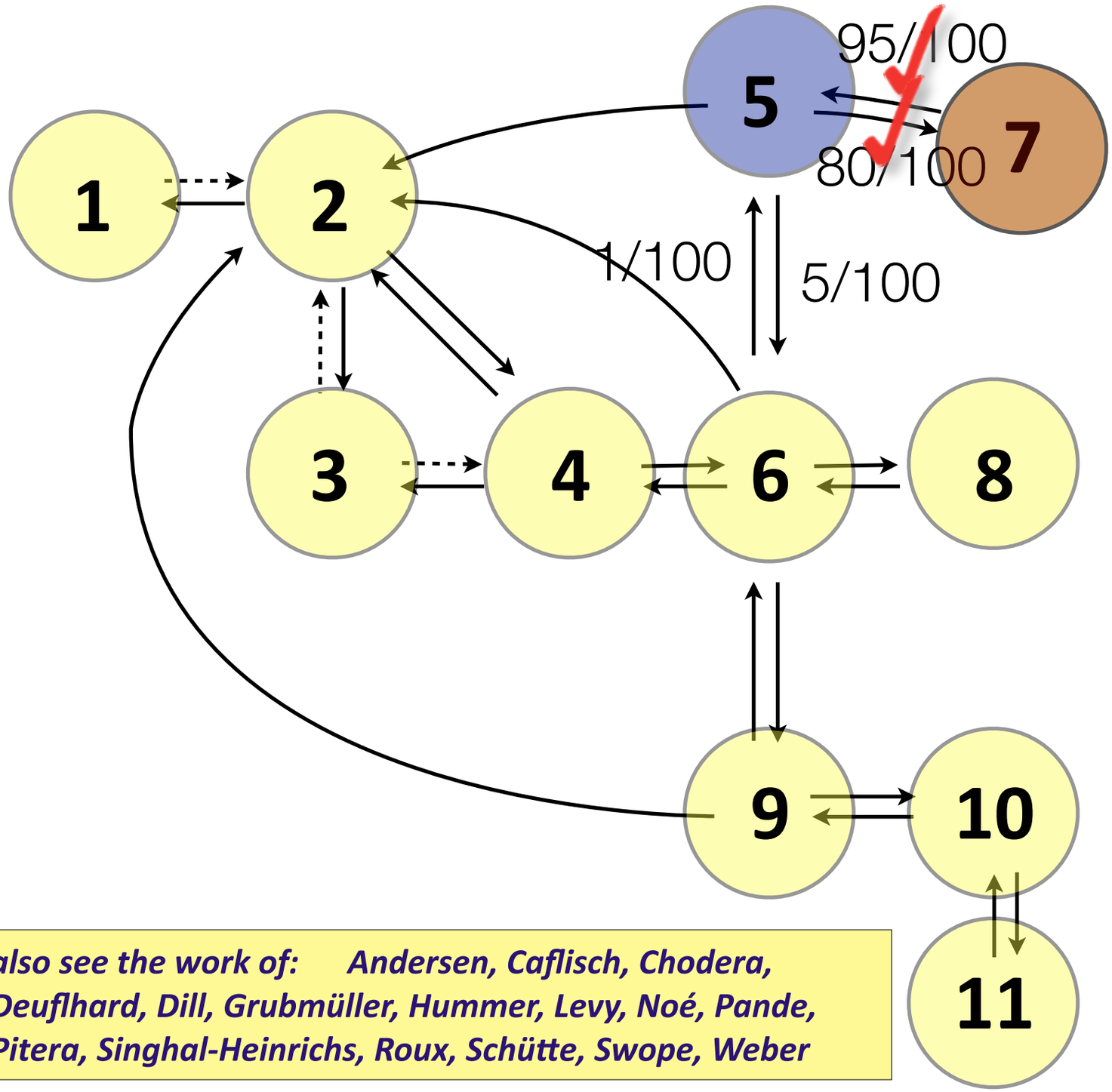
How does one build an MSM? MSMBuilder

Use Active Learning-like adaptive methods to infer p_{ij}

1. **Kinetic clustering:** structures are combined into a state if they rapidly interconvert

2. **Build the best model possible given the data:** Bayesian methods to infer p_{ij} from transition counts

http://simtk.org/home/msmbuilder



also see the work of: Andersen, Caflisch, Chodera, Deuflhard, Dill, Grubmüller, Hummer, Levy, Noé, Pande, Pitera, Singhal-Heinrichs, Roux, Schütte, Swope, Weber

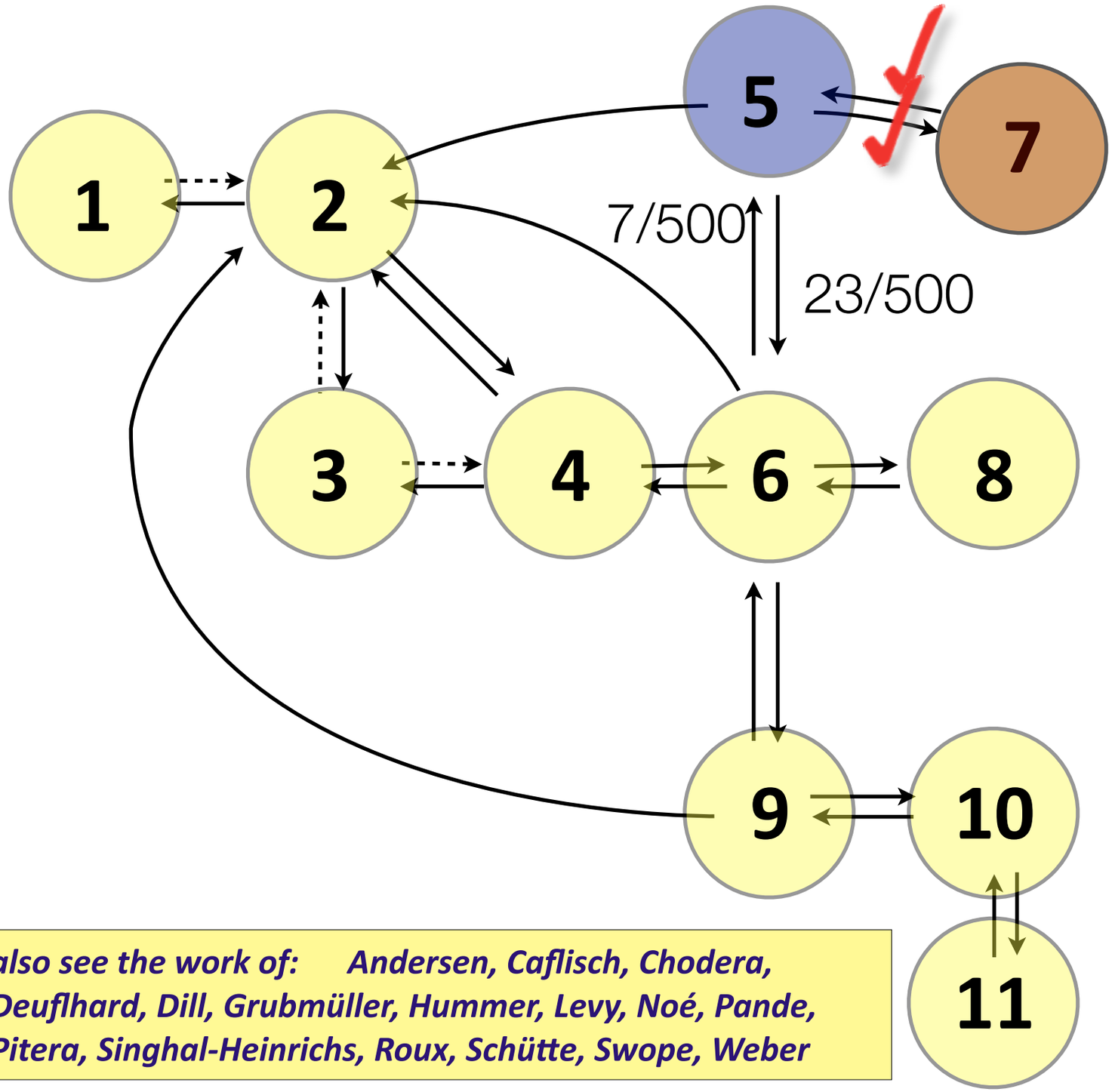
How does one build an MSM? MSMBuilder

Use Active Learning-like adaptive methods to infer p_{ij}

1. **Kinetic clustering:** structures are combined into a state if they rapidly interconvert

2. **Build the best model possible given the data:** Bayesian methods to infer p_{ij} from transition counts

http://simtk.org/home/msmbuilder



also see the work of: Andersen, Caflisch, Chodera, Deuflhard, Dill, Grubmüller, Hummer, Levy, Noé, Pande, Pitera, Singhal-Heinrichs, Roux, Schütte, Swope, Weber

How does one build an MSM? MSMBuilder

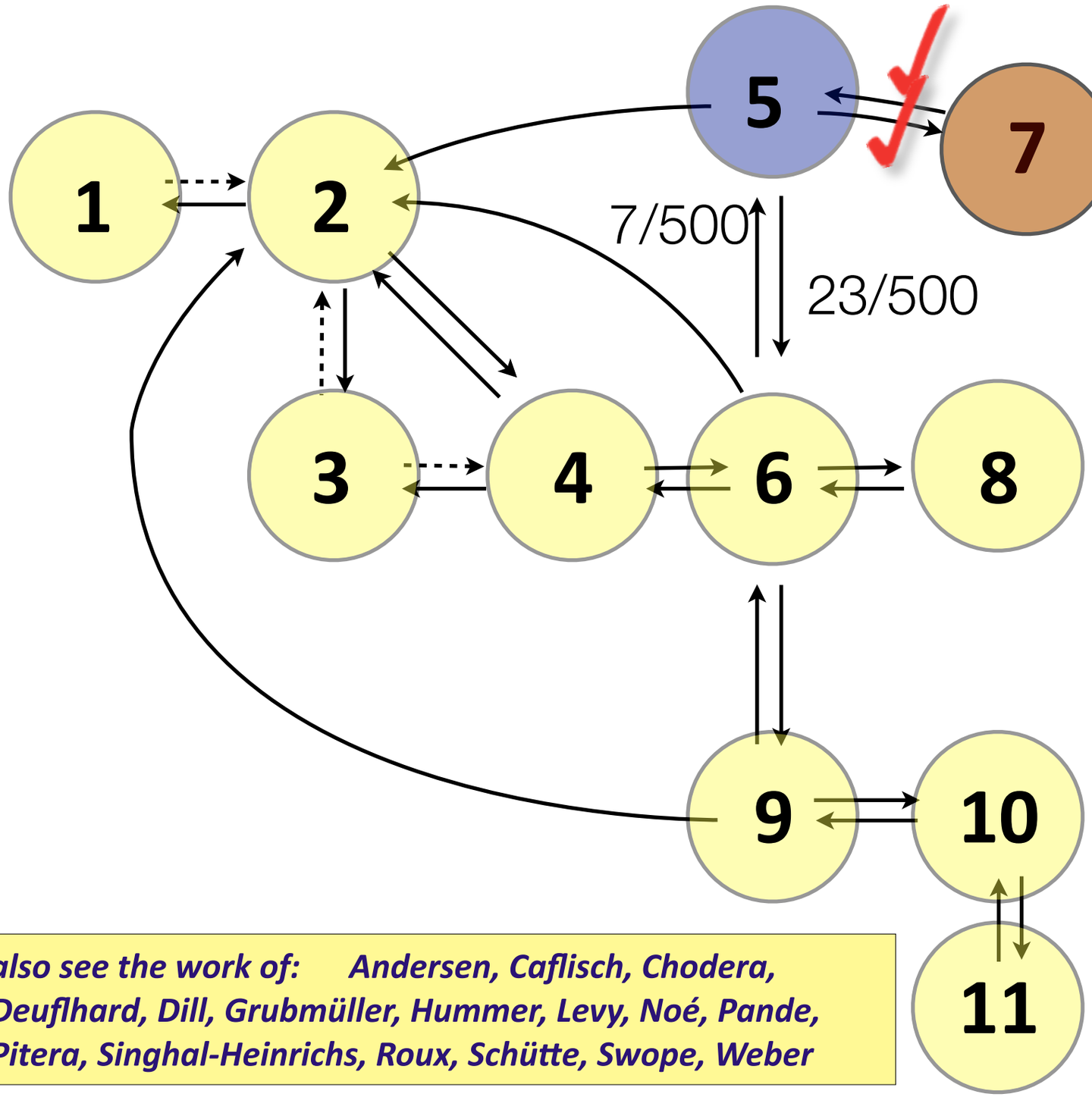
Use Active Learning-like adaptive methods to infer p_{ij}

1. **Kinetic clustering:** structures are combined into a state if they rapidly interconvert

2. **Build the best model possible given the data:** Bayesian methods to infer p_{ij} from transition counts

3. **Simulate proactively:** Use uncertainty to drive new simulations where needed most

http://simtk.org/home/msmbuilder



also see the work of: Andersen, Caflisch, Chodera, Deuflhard, Dill, Grubmüller, Hummer, Levy, Noé, Pande, Pitera, Singhal-Heinrichs, Roux, Schütte, Swope, Weber

How does one build an MSM? MSMBuilder

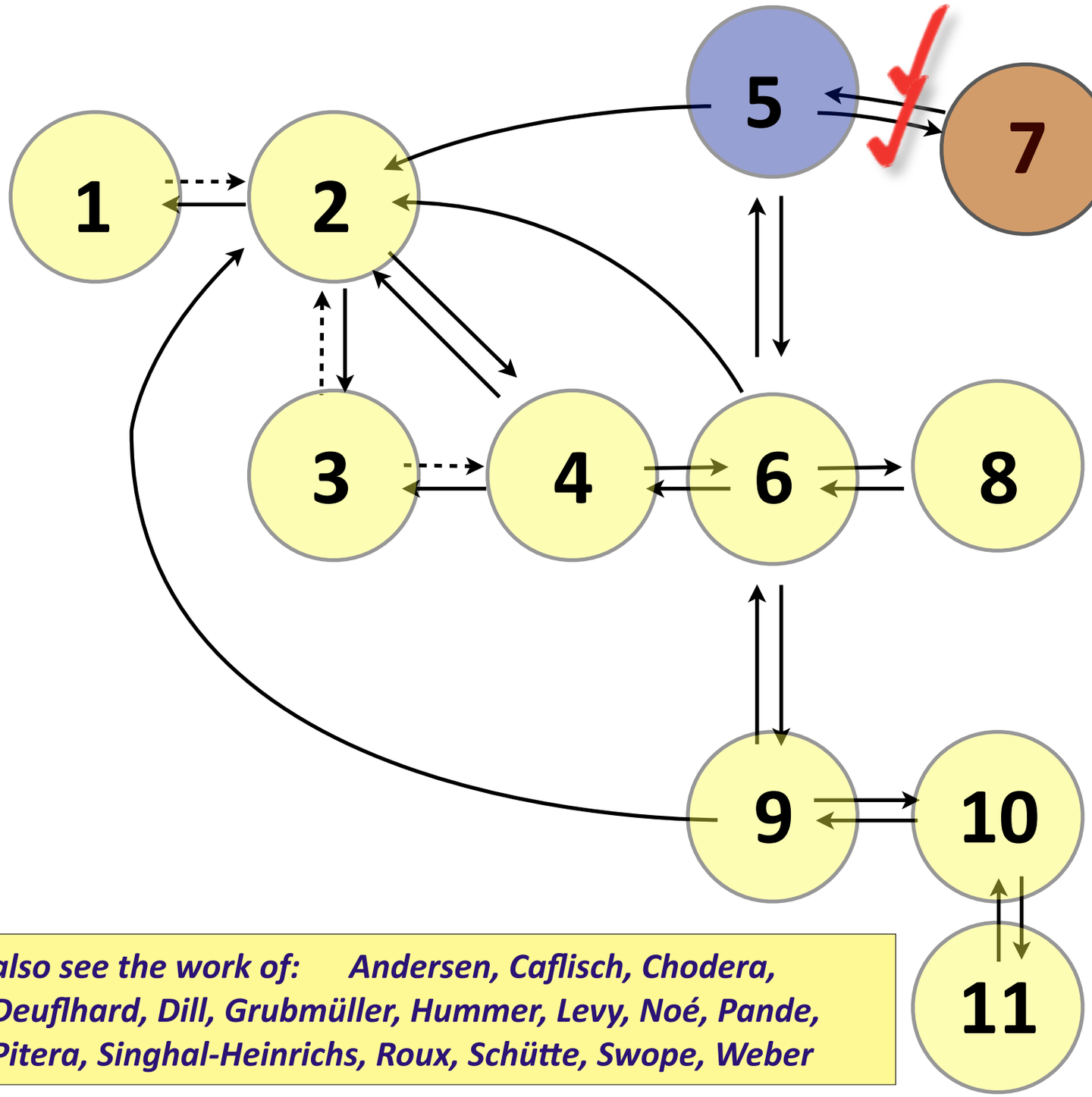
Use Active Learning-like adaptive methods to infer p_{ij}

1. **Kinetic clustering:** structures are combined into a state if they rapidly interconvert

2. **Build the best model possible given the data:** Bayesian methods to infer p_{ij} from transition counts

3. **Simulate proactively:** Use uncertainty to drive new simulations where needed most

http://simtk.org/home/msmbuilder



also see the work of: Andersen, Caflisch, Chodera, Deuflhard, Dill, Grubmüller, Hummer, Levy, Noé, Pande, Pitera, Singhal-Heinrichs, Roux, Schütte, Swope, Weber

How does one build an MSM? MSMBuilder

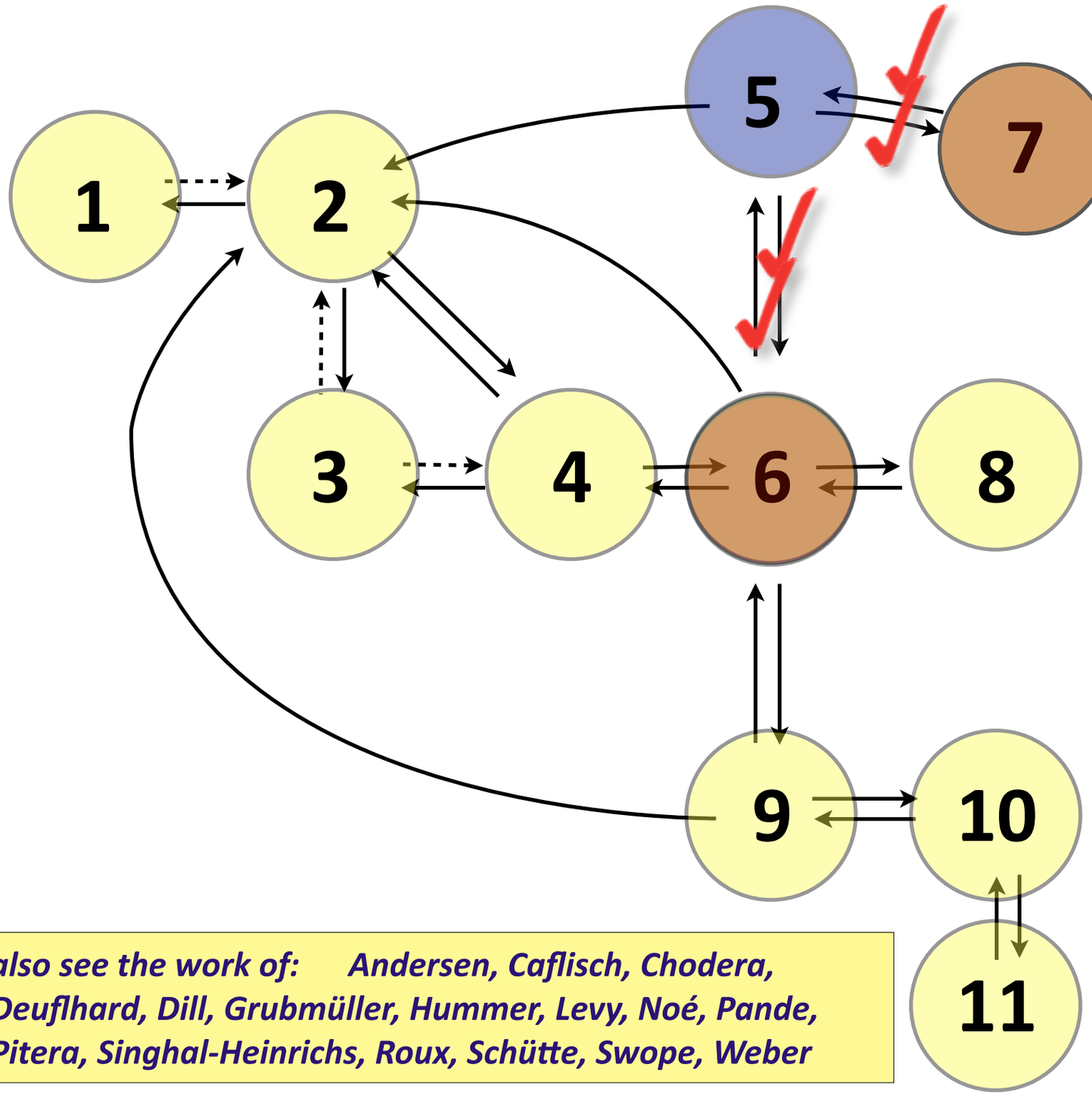
Use Active Learning-like adaptive methods to infer p_{ij}

1. **Kinetic clustering:** structures are combined into a state if they rapidly interconvert

2. **Build the best model possible given the data:** Bayesian methods to infer p_{ij} from transition counts

3. **Simulate proactively:** Use uncertainty to drive new simulations where needed most

http://simtk.org/home/msmbuilder



also see the work of: Andersen, Caflisch, Chodera, Deuflhard, Dill, Grubmüller, Hummer, Levy, Noé, Pande, Pitera, Singhal-Heinrichs, Roux, Schütte, Swope, Weber

How does one build an MSM? MSMBuilder

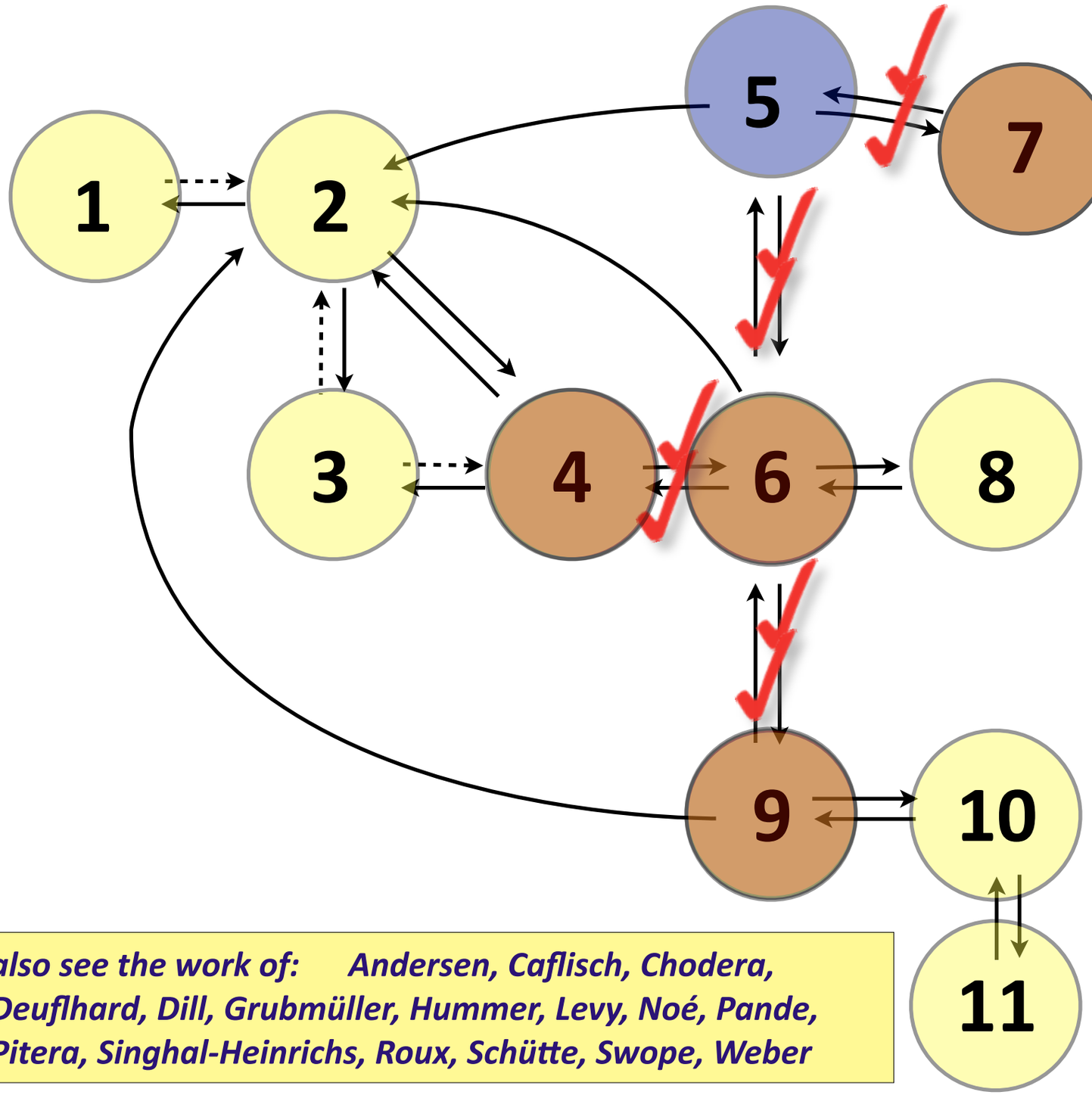
Use Active Learning-like adaptive methods to infer p_{ij}

1. **Kinetic clustering:** structures are combined into a state if they rapidly interconvert

2. **Build the best model possible given the data:** Bayesian methods to infer p_{ij} from transition counts

3. **Simulate proactively:** Use uncertainty to drive new simulations where needed most

http://simtk.org/home/msmbuilder



also see the work of: Andersen, Caflisch, Chodera, Deuflhard, Dill, Grubmüller, Hummer, Levy, Noé, Pande, Pitera, Singhal-Heinrichs, Roux, Schütte, Swope, Weber

How does one build an MSM? MSMBuilder

Use Active Learning-like adaptive methods to infer p_{ij}

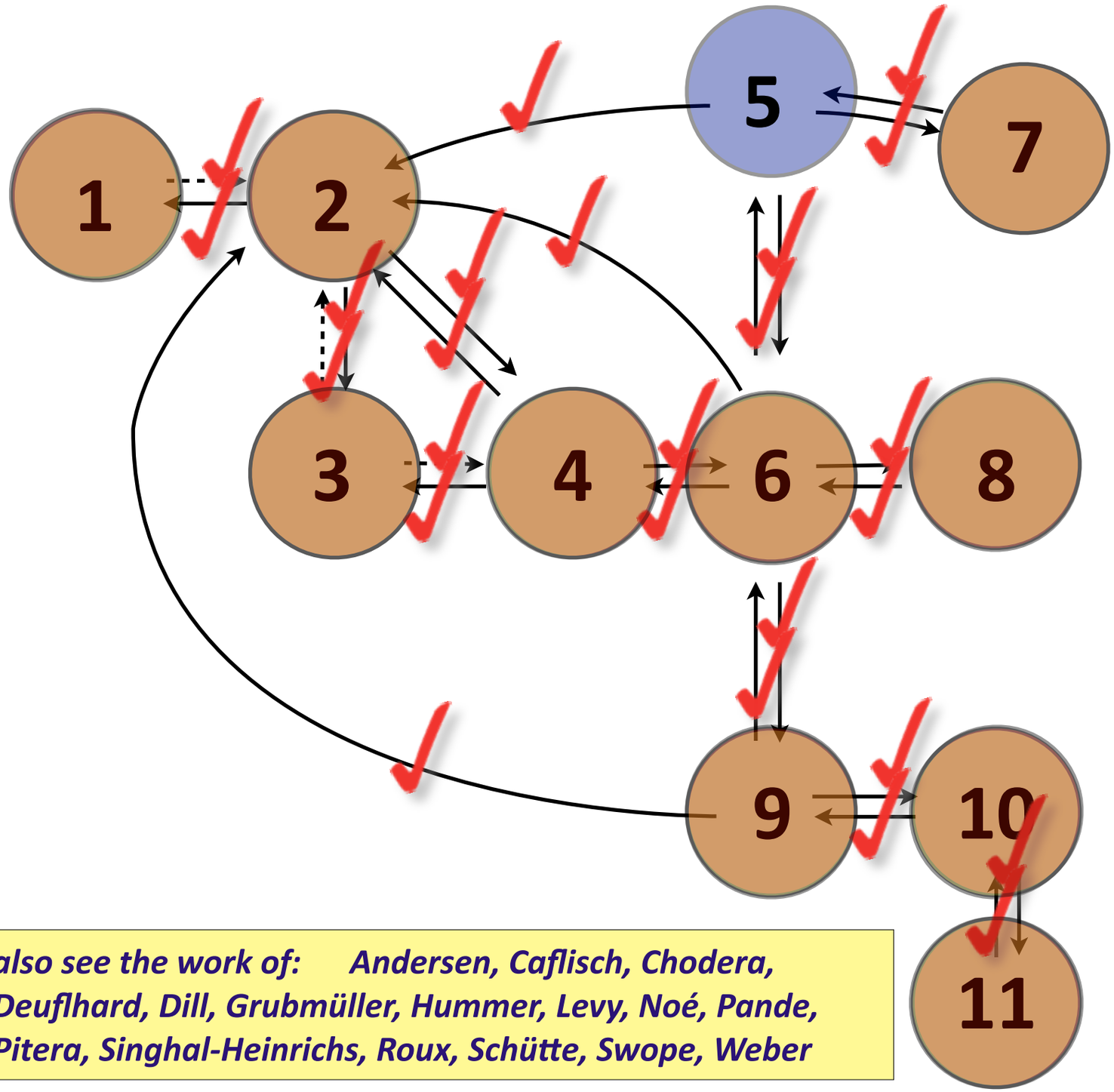
1. **Kinetic clustering:** structures are combined into a state if they rapidly interconvert

2. **Build the best model possible given the data:** Bayesian methods to infer p_{ij} from transition counts

3. **Simulate proactively:** Use uncertainty to drive new simulations where needed most

4. **Repeat until tolerances are met**

http://simtk.org/home/msmbuilder

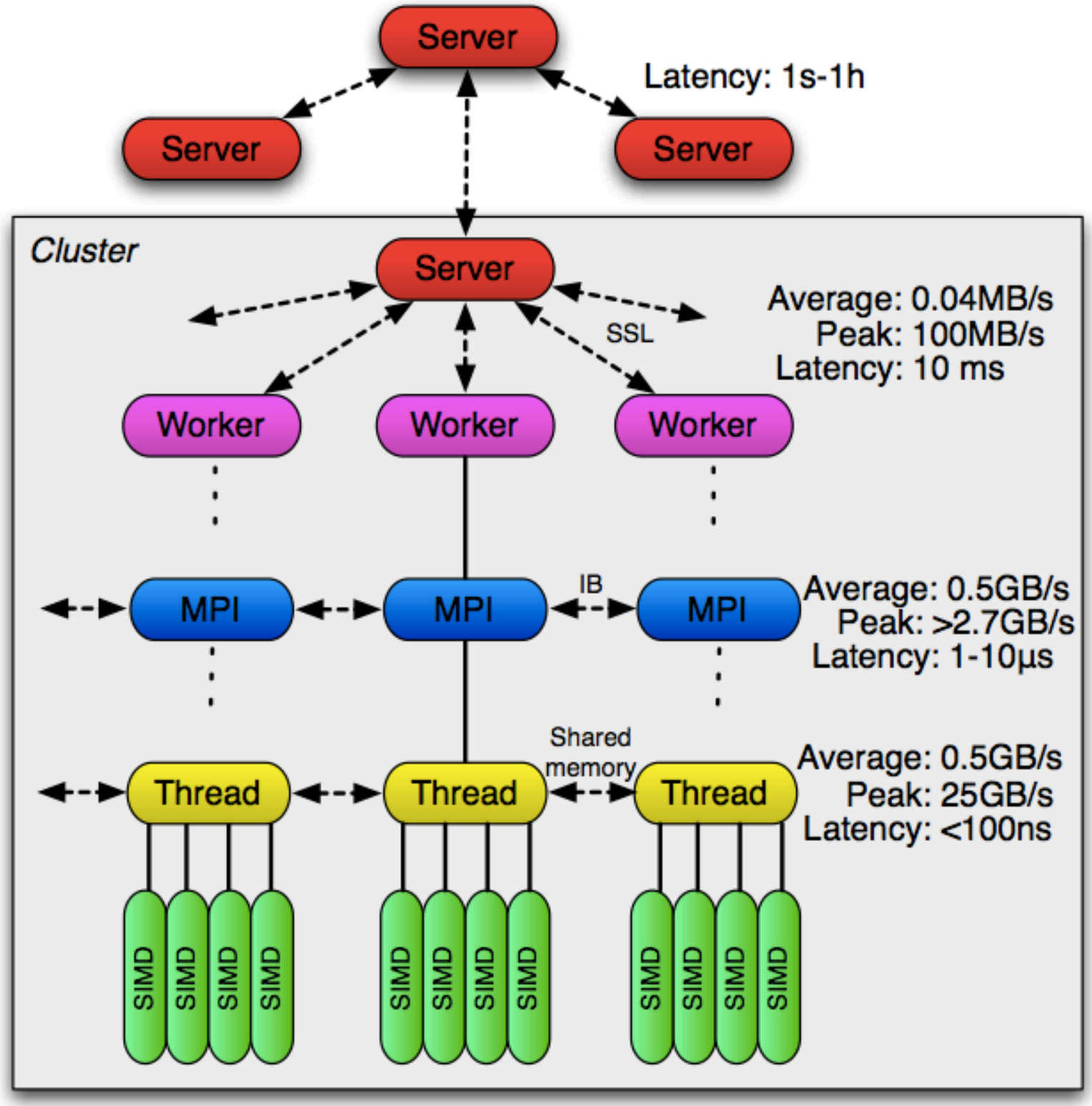


also see the work of: Andersen, Caflisch, Chodera, Deuflhard, Dill, Grubmüller, Hummer, Levy, Noé, Pande, Pitera, Singhal-Heinrichs, Roux, Schütte, Swope, Weber

Linear scaling across multiple HPC machines

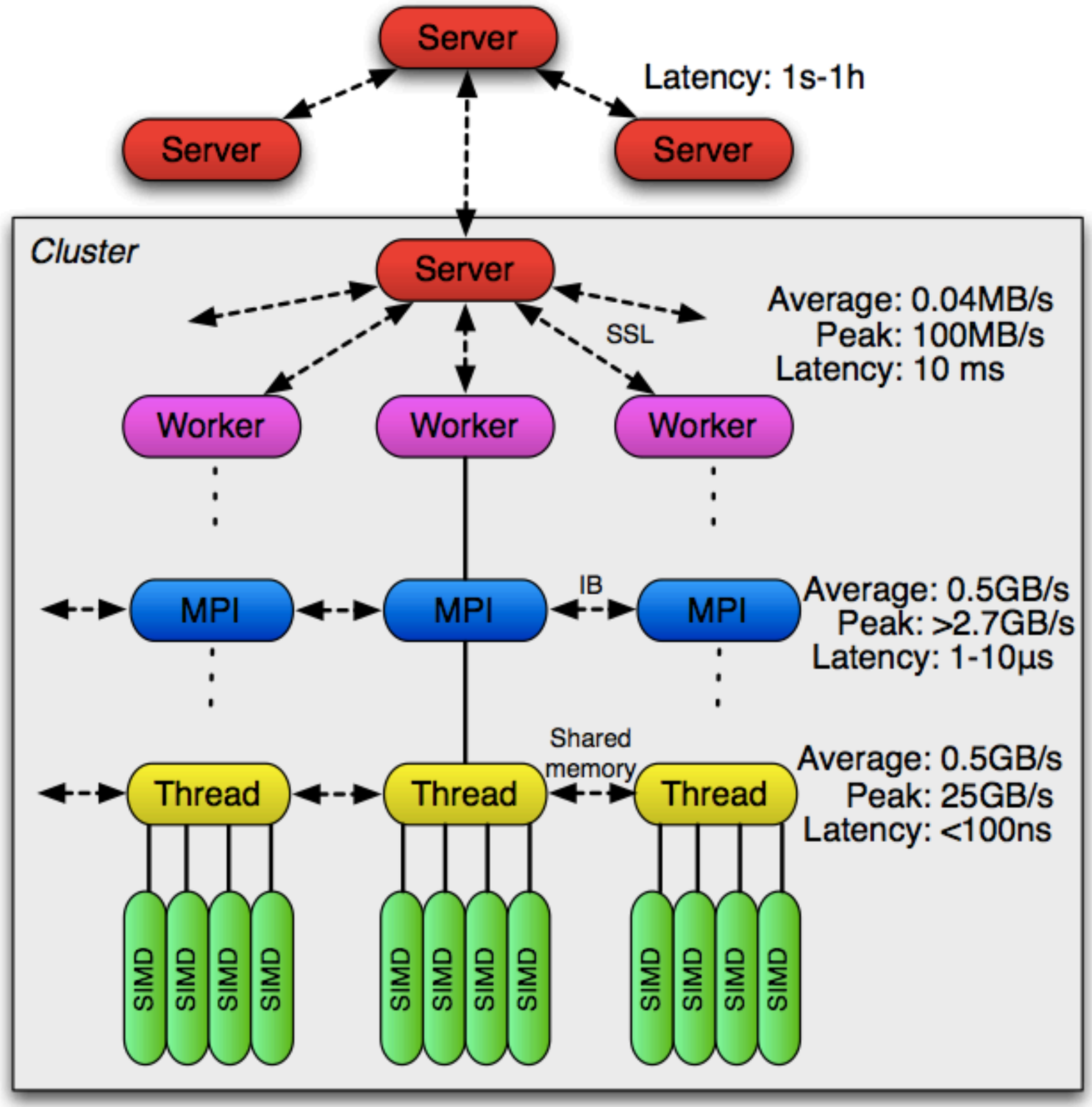
Copernicus

Copernicus: A Massively Parallel Engine for Simulation



Copernicus: A Massively Parallel Engine for Simulation

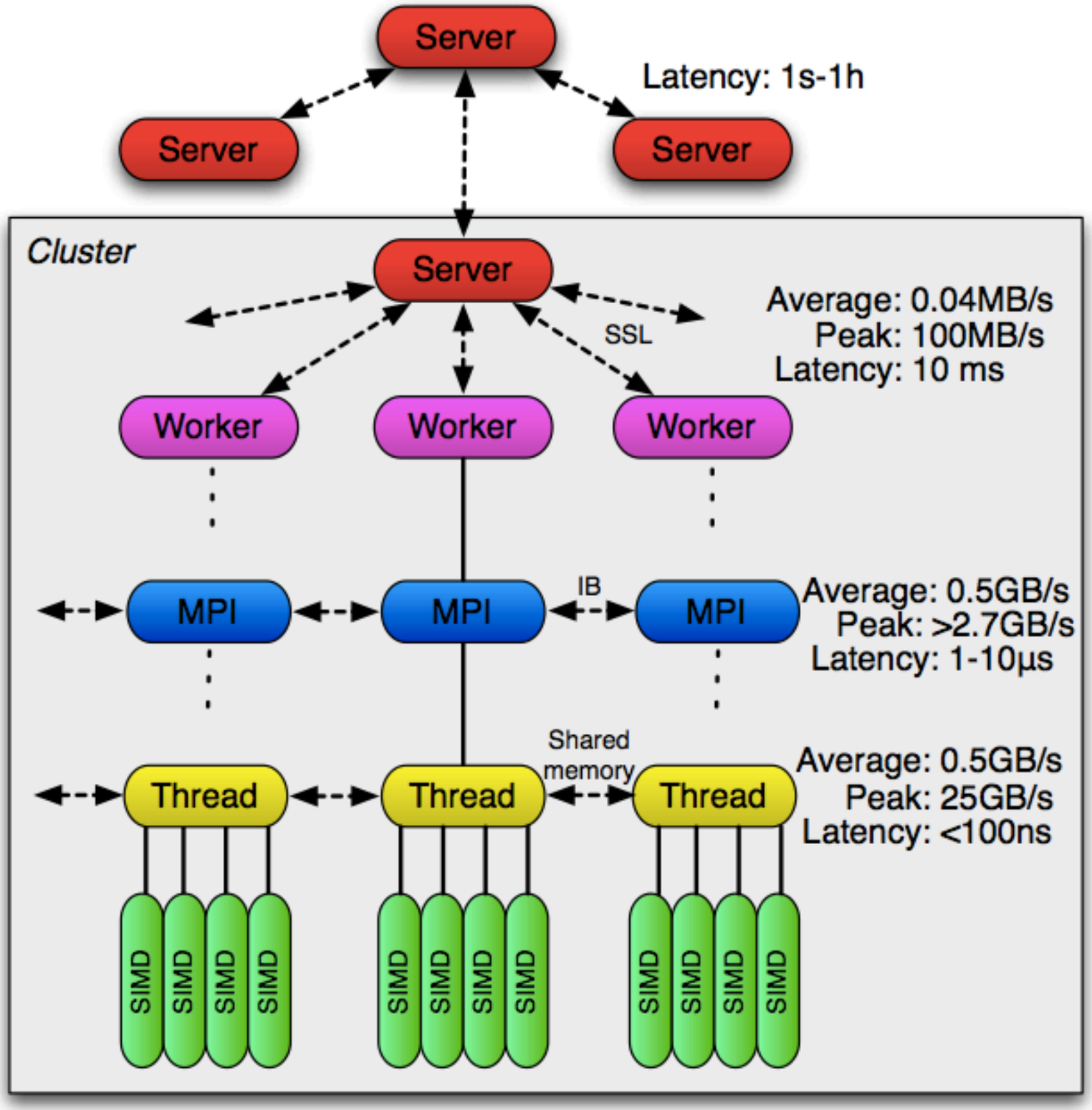
1. Multi-machine paradigm



Copernicus: A Massively Parallel Engine for Simulation

1. Multi-machine paradigm

2. Heirarchical architecture

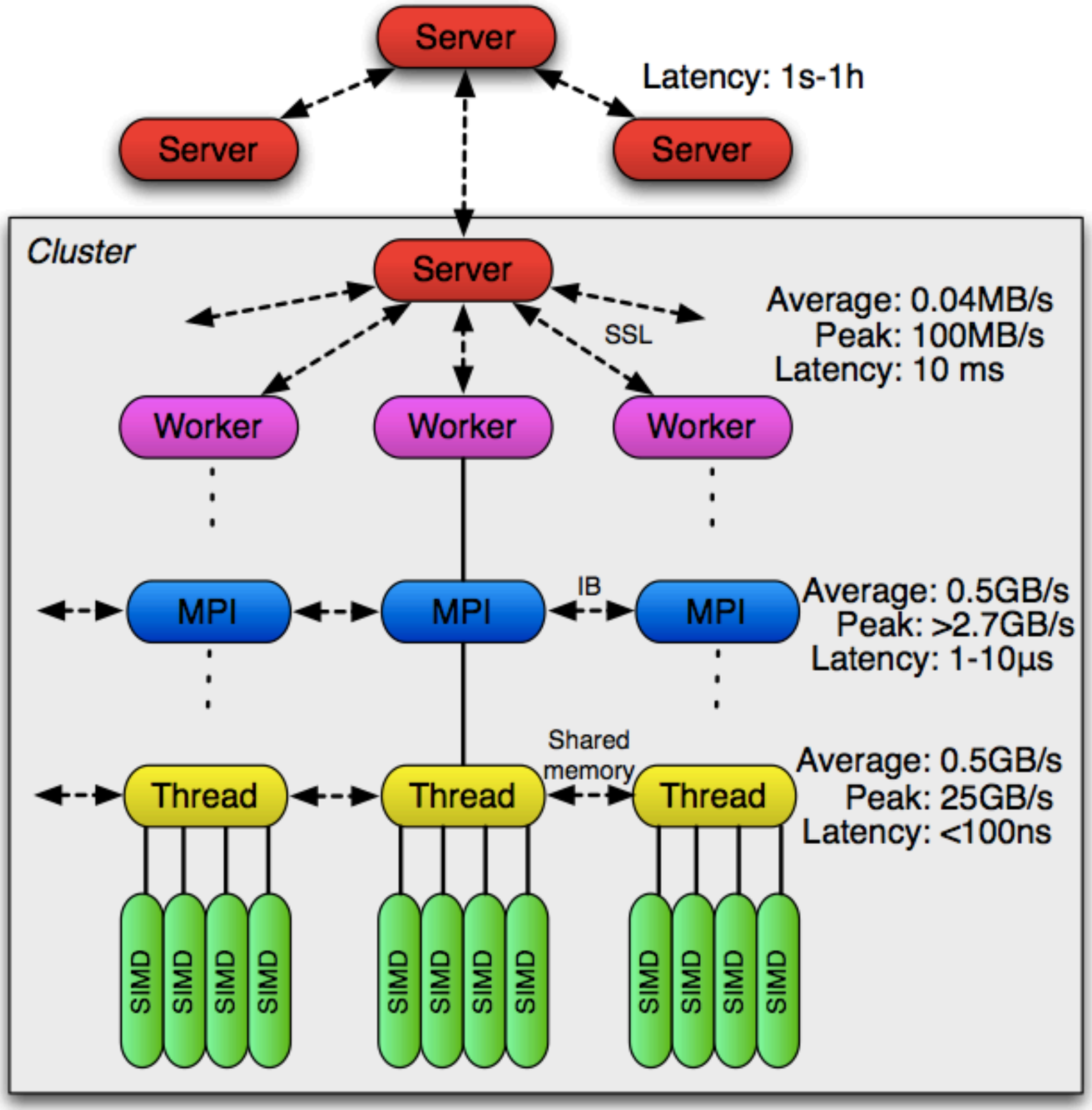


Copernicus: A Massively Parallel Engine for Simulation

1. Multi-machine paradigm

2. Heirarchical architecture

3. Uses real-time clustering to adaptively drive simulations to convergence



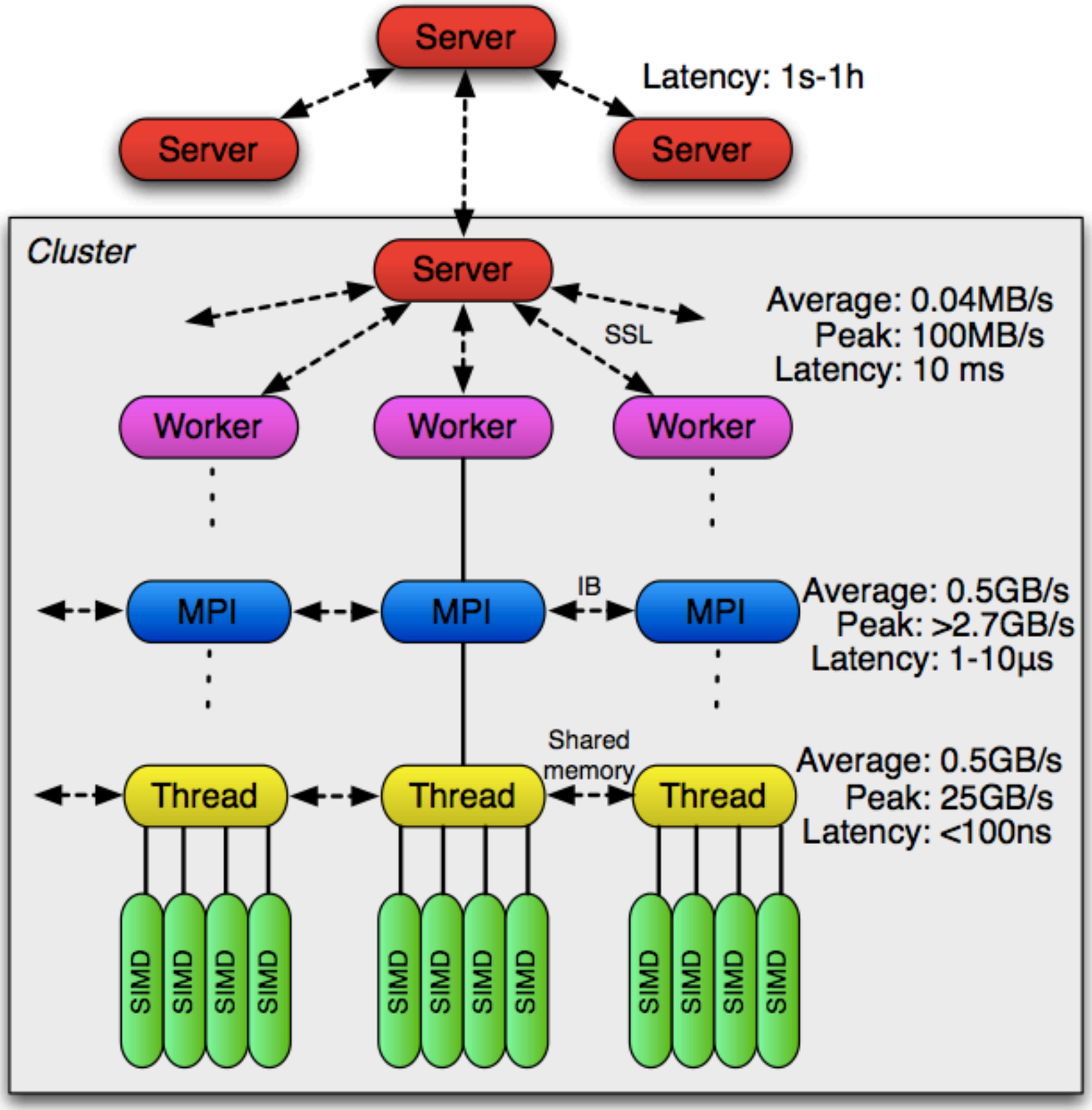
Copernicus: A Massively Parallel Engine for Simulation

1. Multi-machine paradigm

2. Heirarchical architecture

3. Uses real-time clustering to adaptively drive simulations to convergence

4. Results in guaranteed linear scaling in simulation time



Copernicus: A Massively Parallel Engine for Simulation

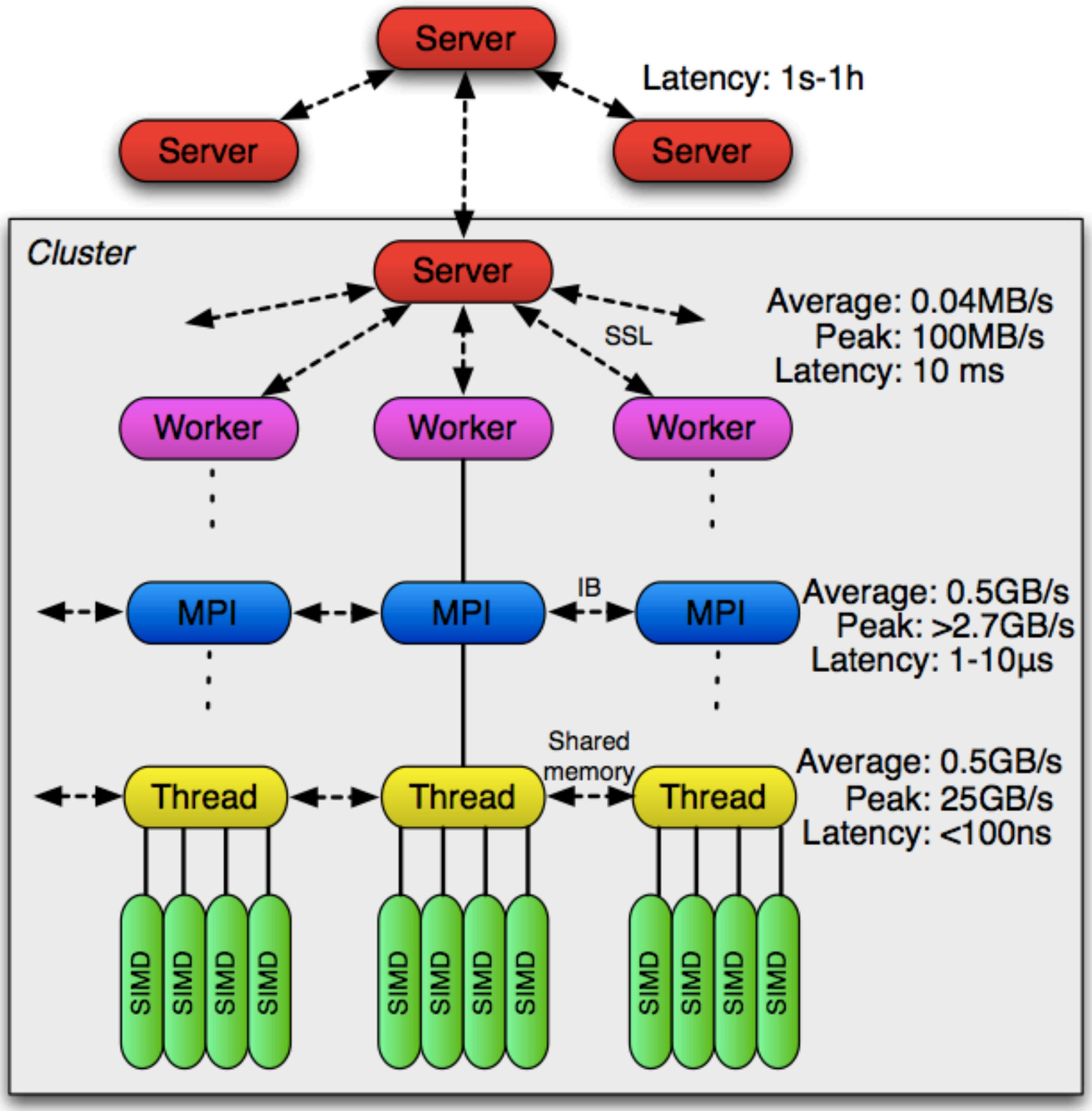
1. Multi-machine paradigm

2. Heirarchical architecture

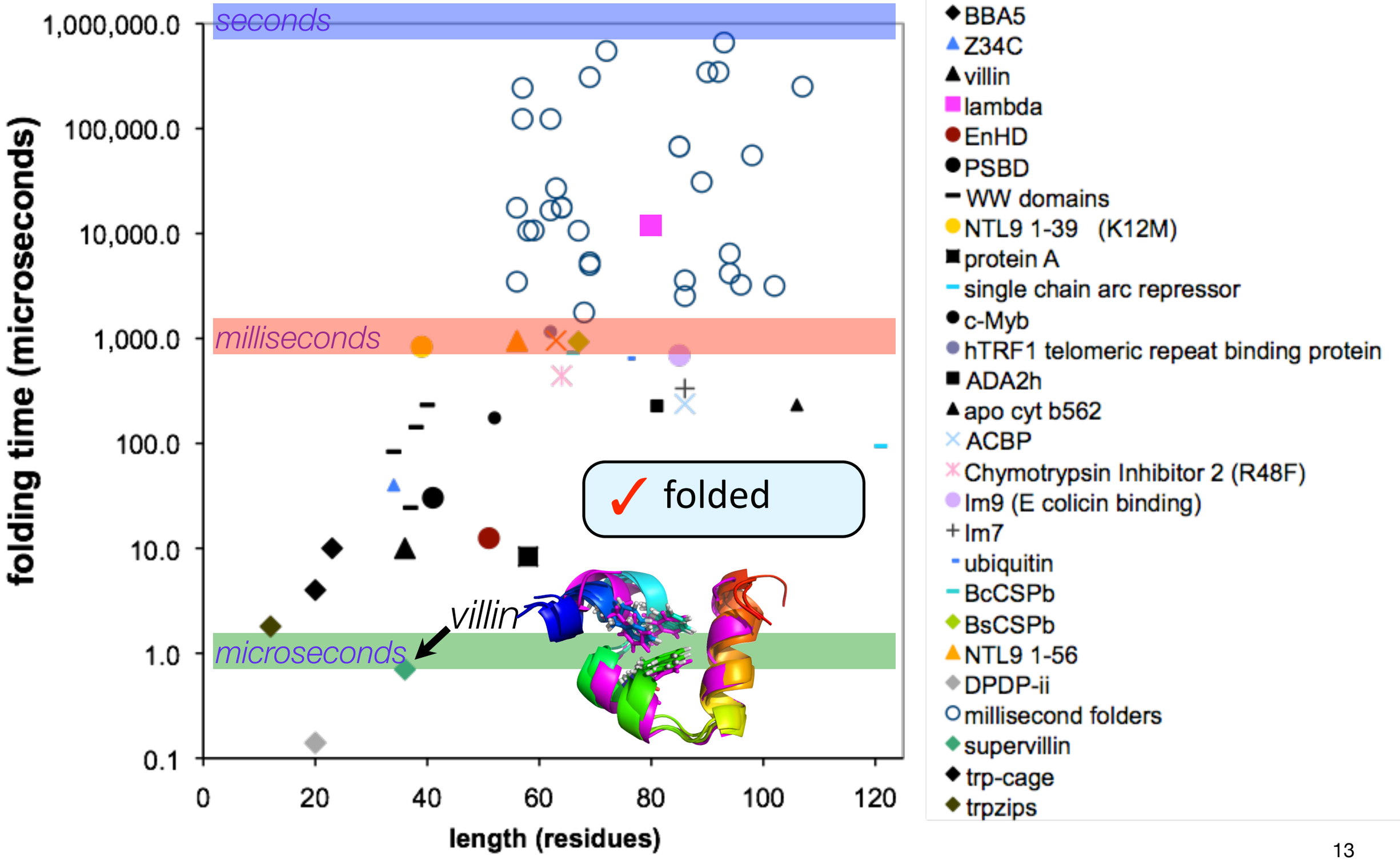
3. Uses real-time clustering to adaptively drive simulations to convergence

4. Results in guaranteed linear scaling in simulation time

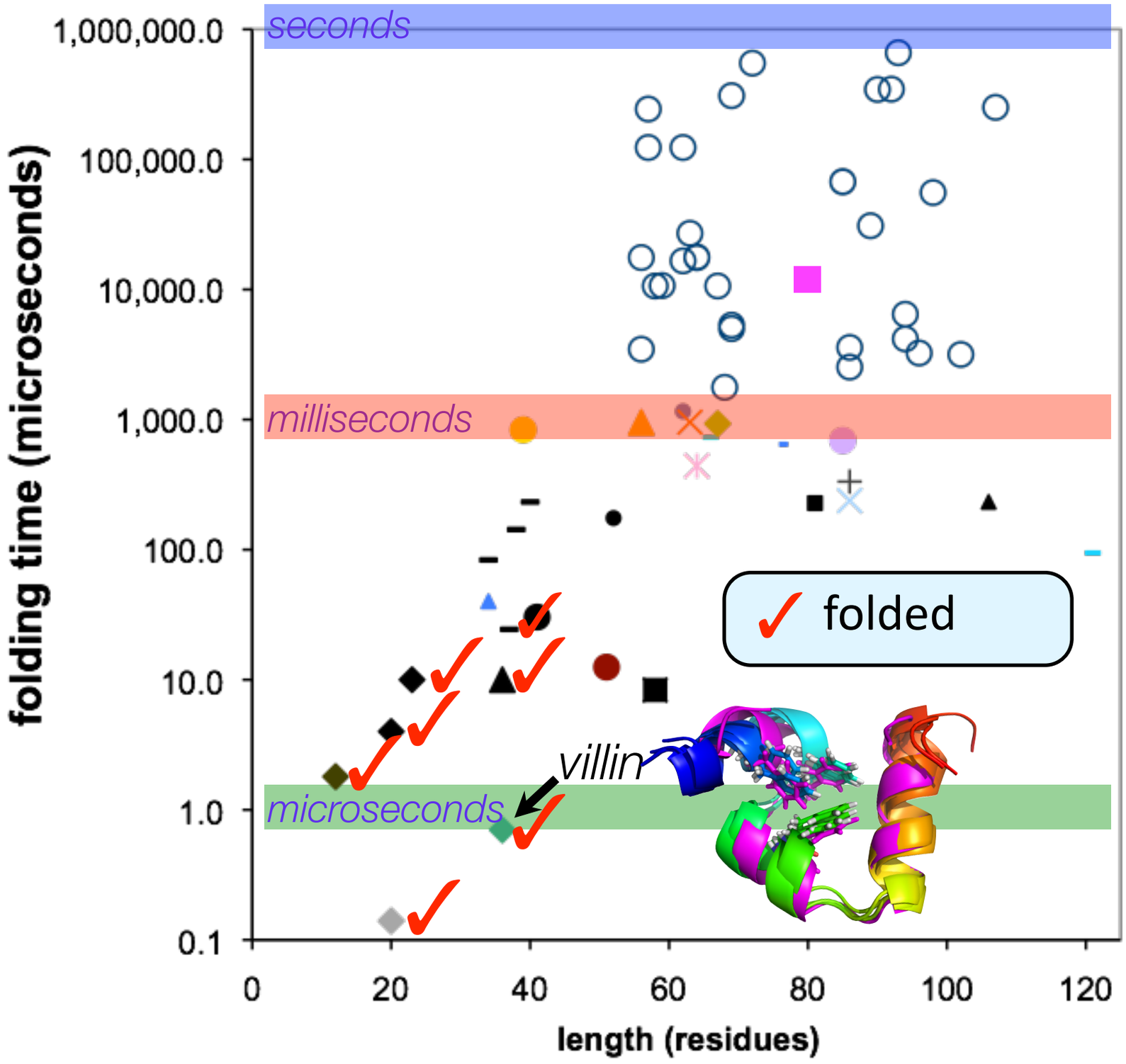
5. Near linear scaling in time-to-solution (folded protein)

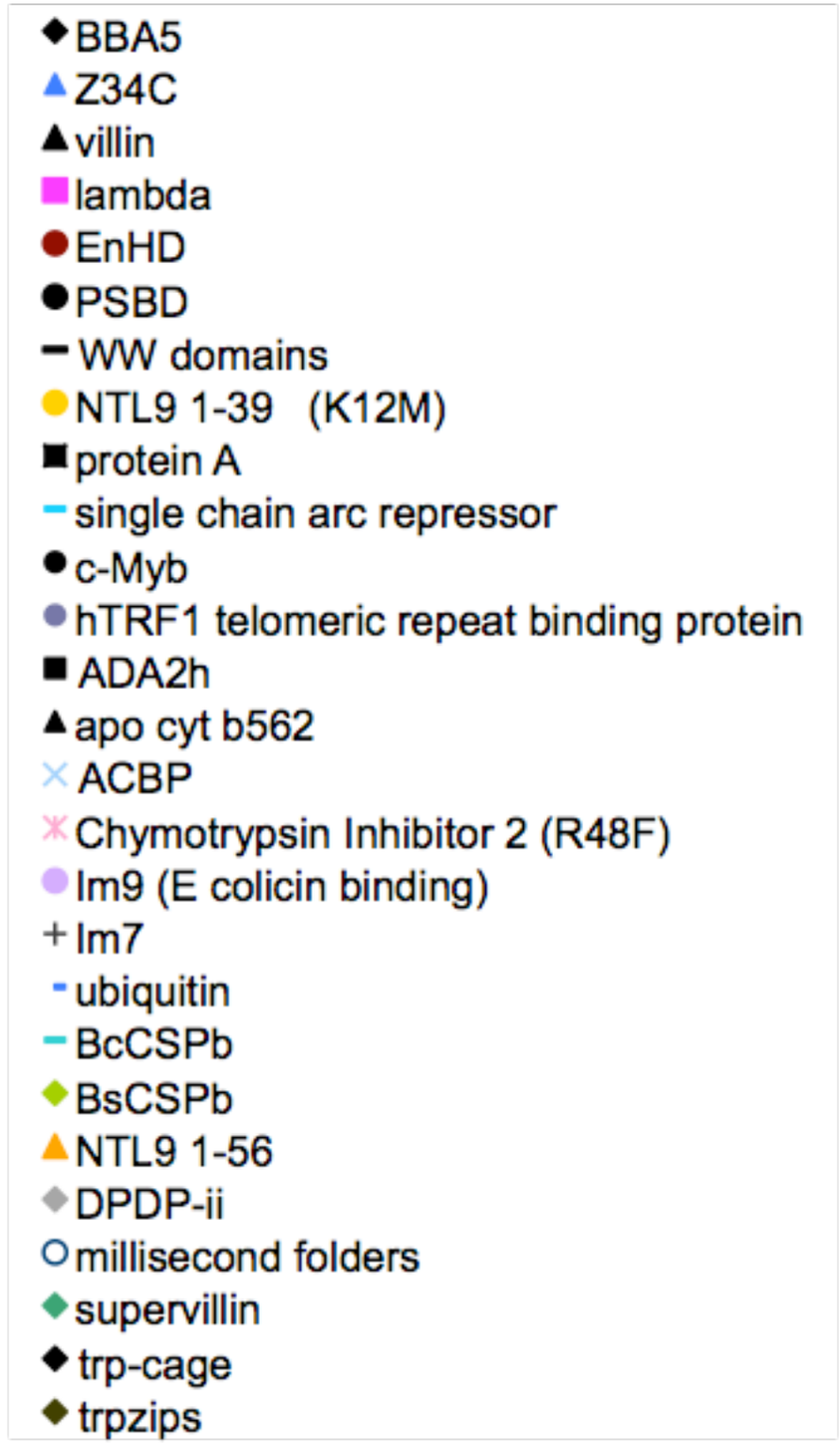


Application: folding on the millisecond timescale

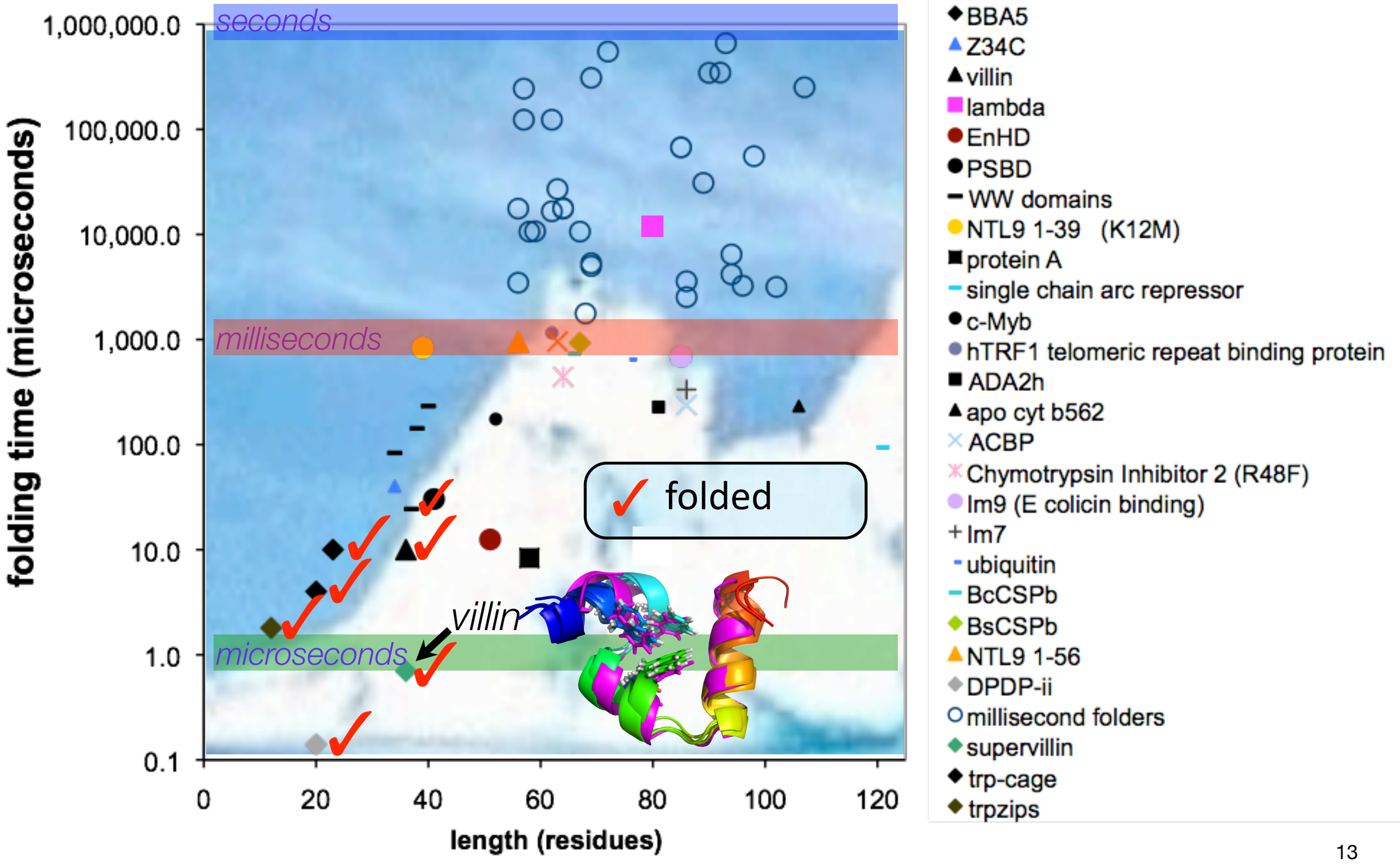


Application: folding on the millisecond timescale

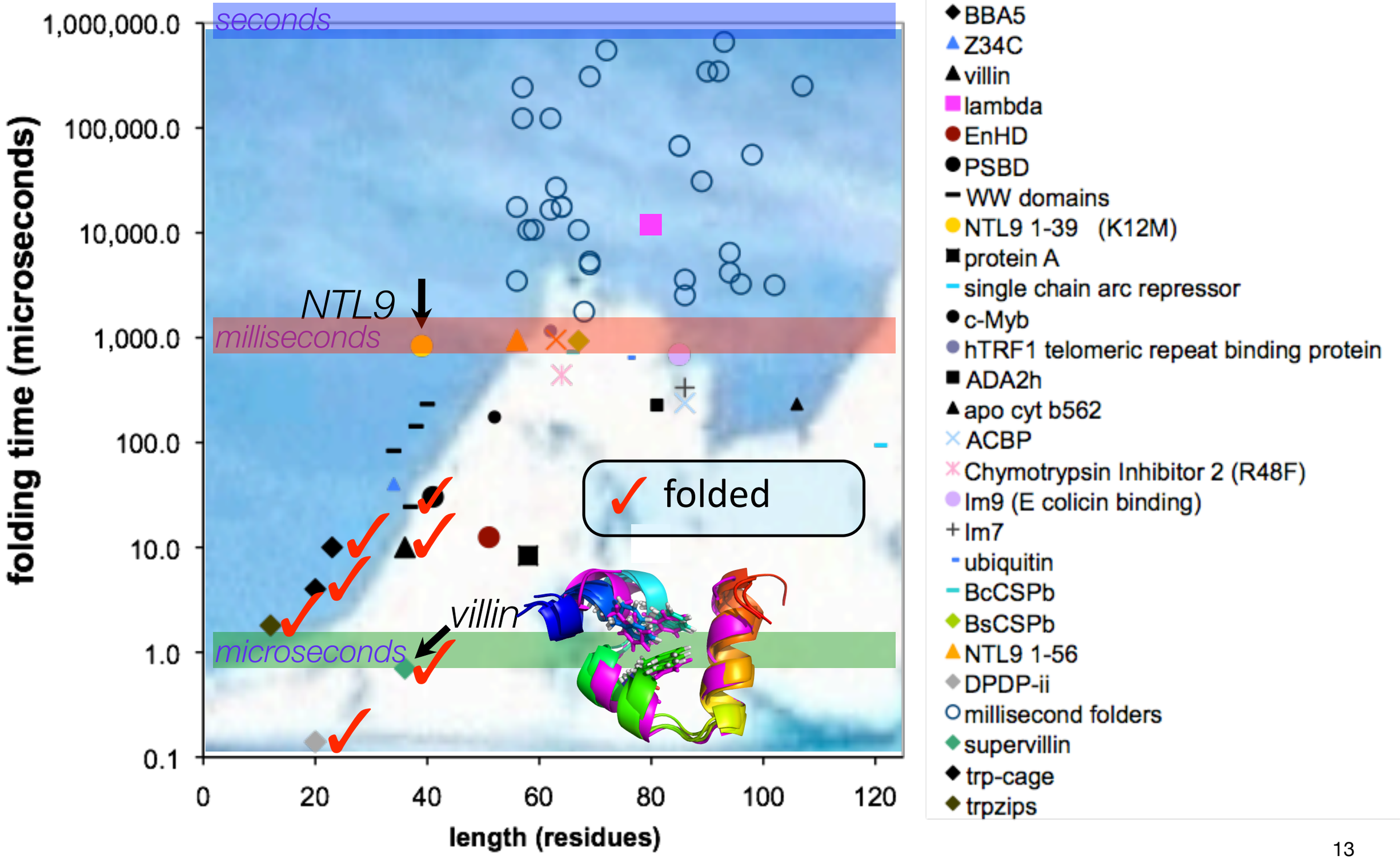




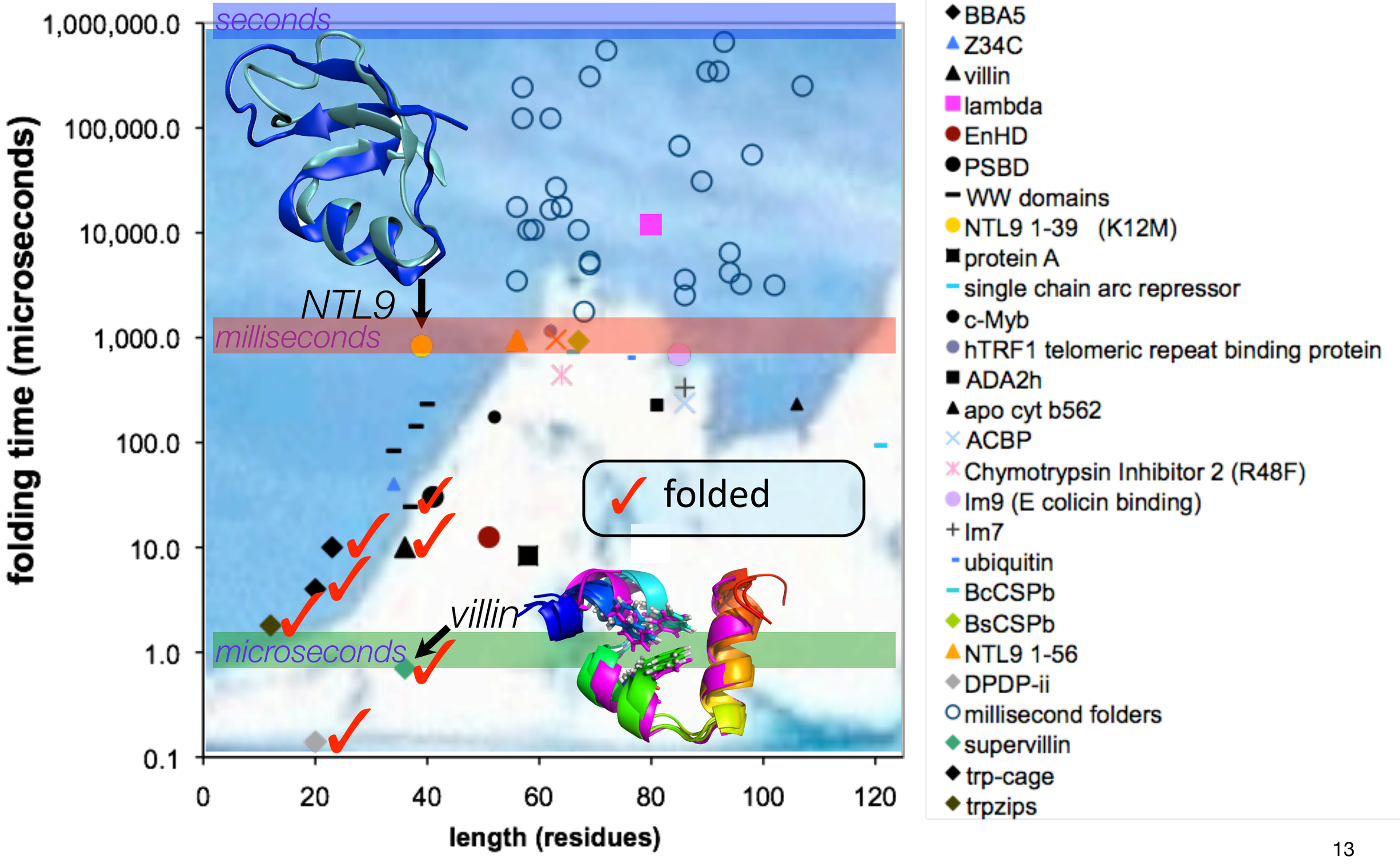
Application: folding on the millisecond timescale



Application: folding on the millisecond timescale

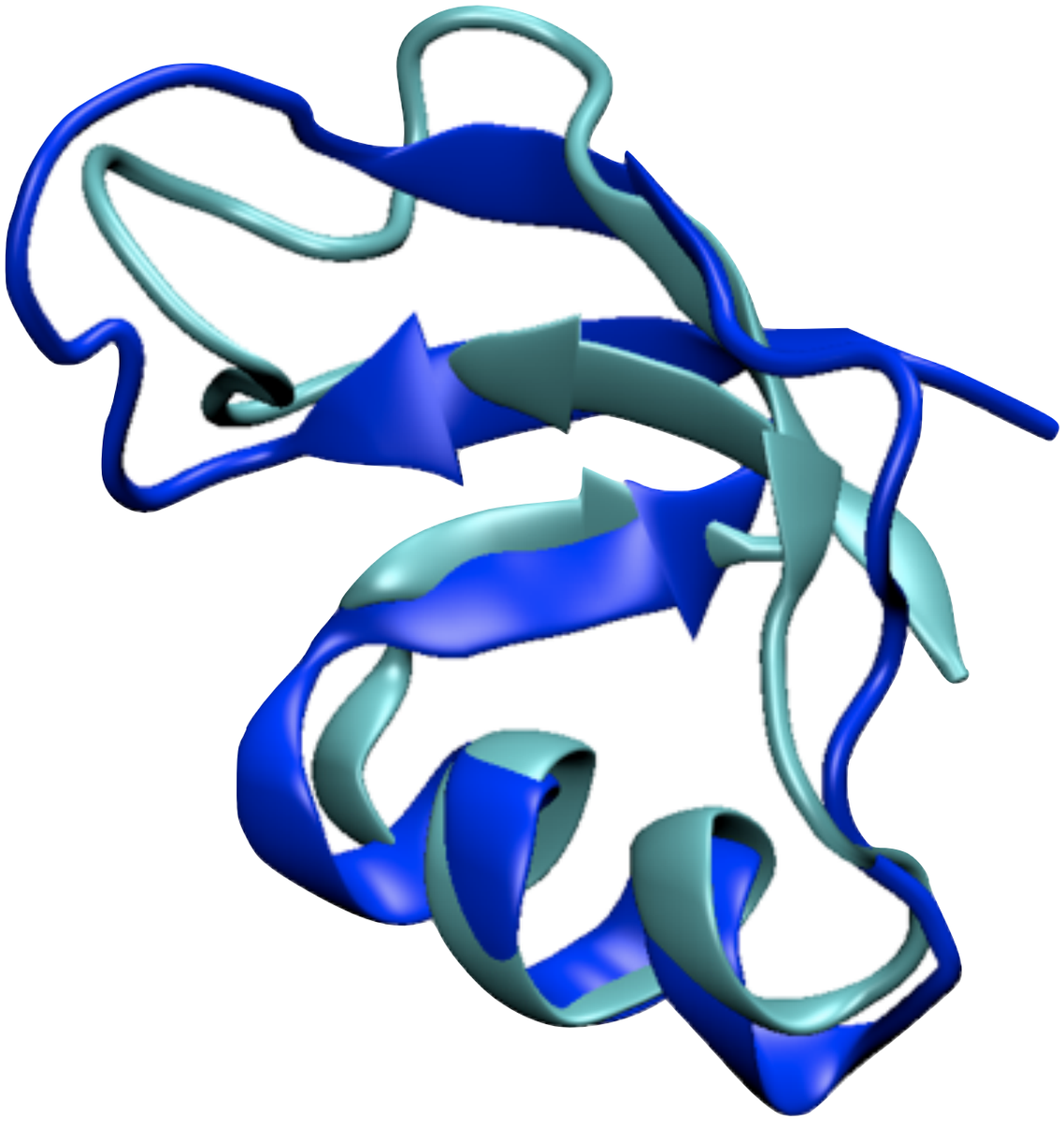


Application: folding on the millisecond timescale

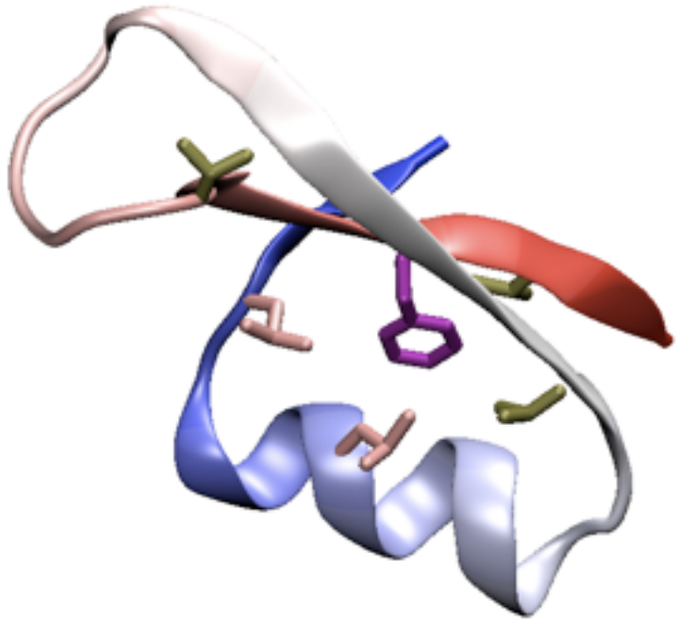


(Voelz, Bowman, Beauchamp, VSP)

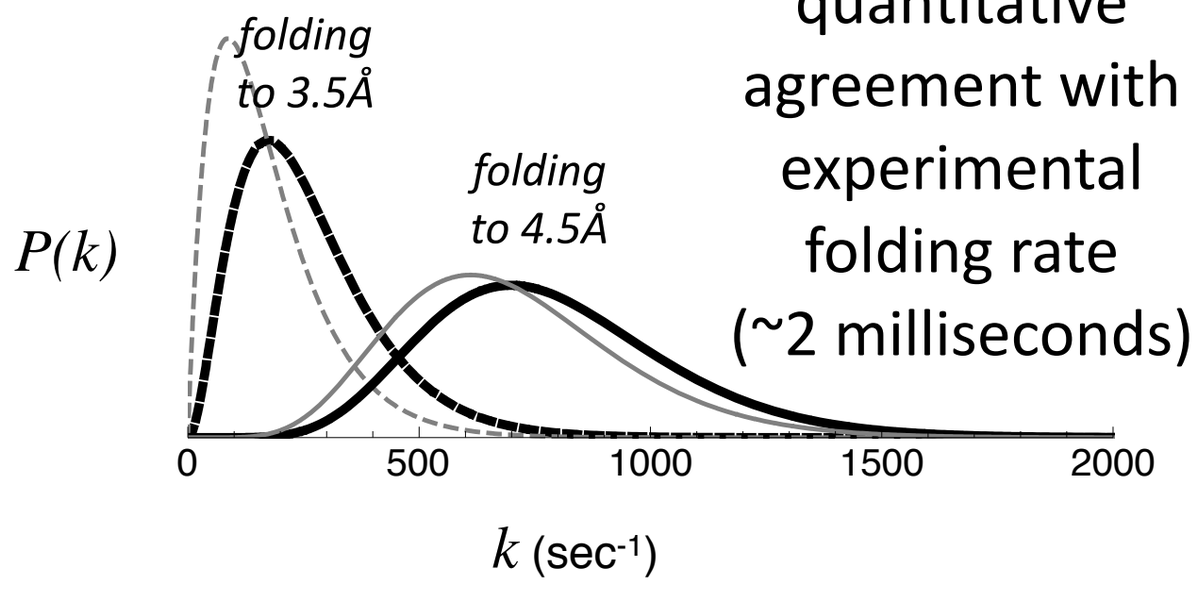
Successful millisecond protein folding: NTL9



crystal structure vs simulation
(3.17Å RMSD-Cα)



native-like packing, with PHE5 in the hydrophobic core



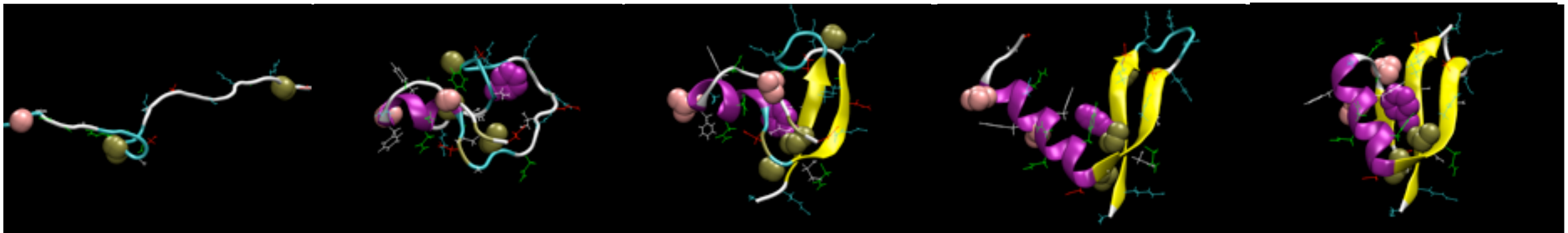
quantitative agreement with experimental folding rate (~2 milliseconds)

Pathway seen in the movie: *Series of **metastable states***

(Voelz, Bowman, Beauchamp, VSP)

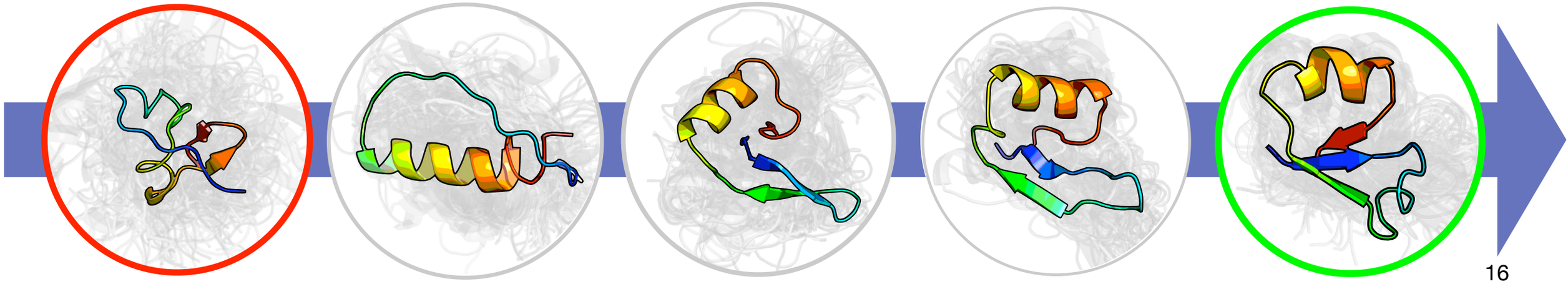
Voelz, Bowman, Beauchamp, Pande. *JACS* (2010)

snapshots from the movie:

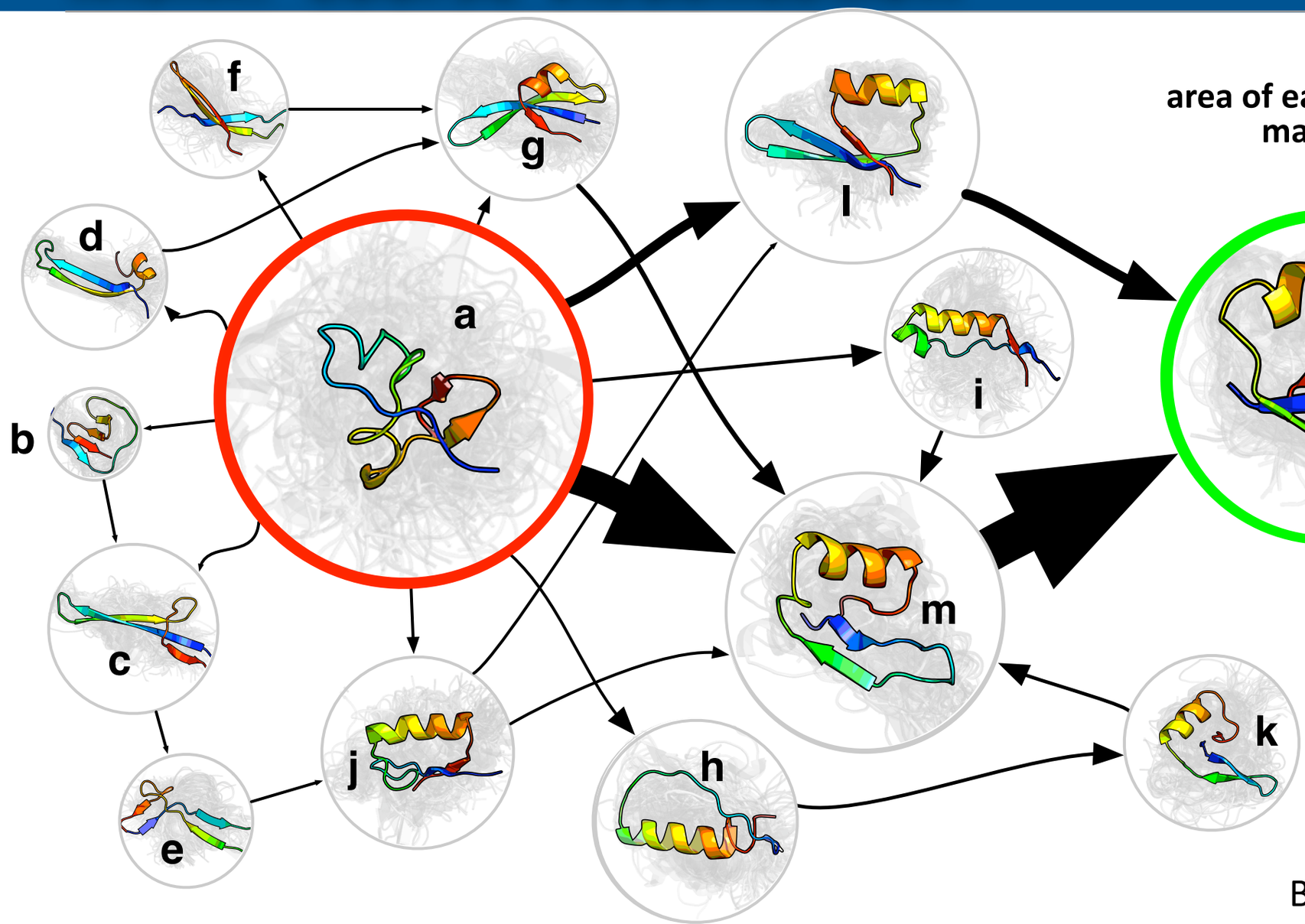


starts in unfolded state	helix forms early	collapse, then beta sheet forms	final part of beta ready to align	folded structure forms

correspond to states from our Markov State Model:



Repeating with many more trajectories yields an MSM: coarse visualization (Voelz, Bowman, Beauchamp, VSP)



area of each state is proportional to macrostate free energy

a→l→n and **a→m→n** comprise 10% of the total flux

width of each arrow is proportional to transition flux

Flux calculation method:

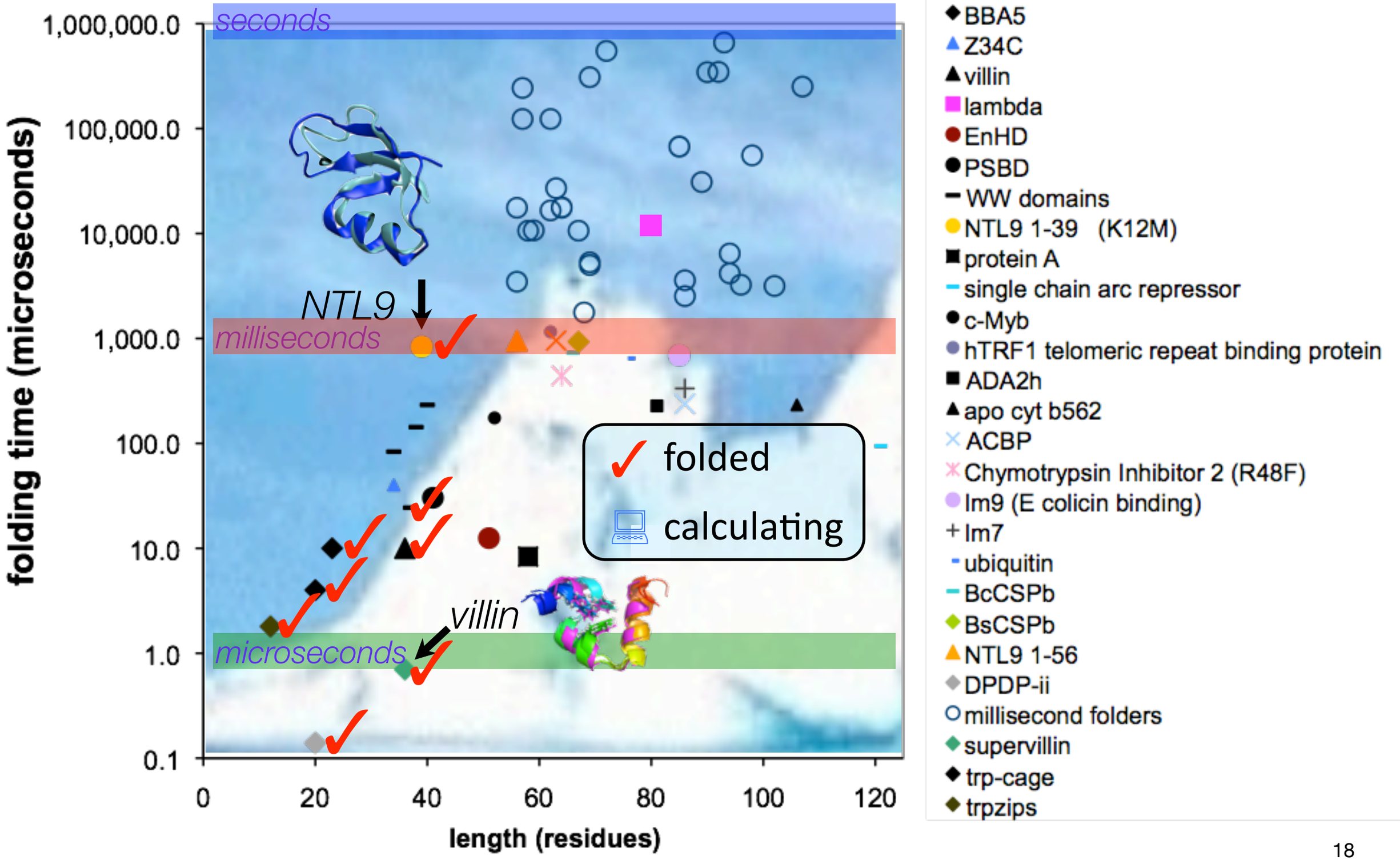
TPT: Vanden-Eijnden, *et al* (2006)

Berezhkovskii, Hummer, Szabo (2009)

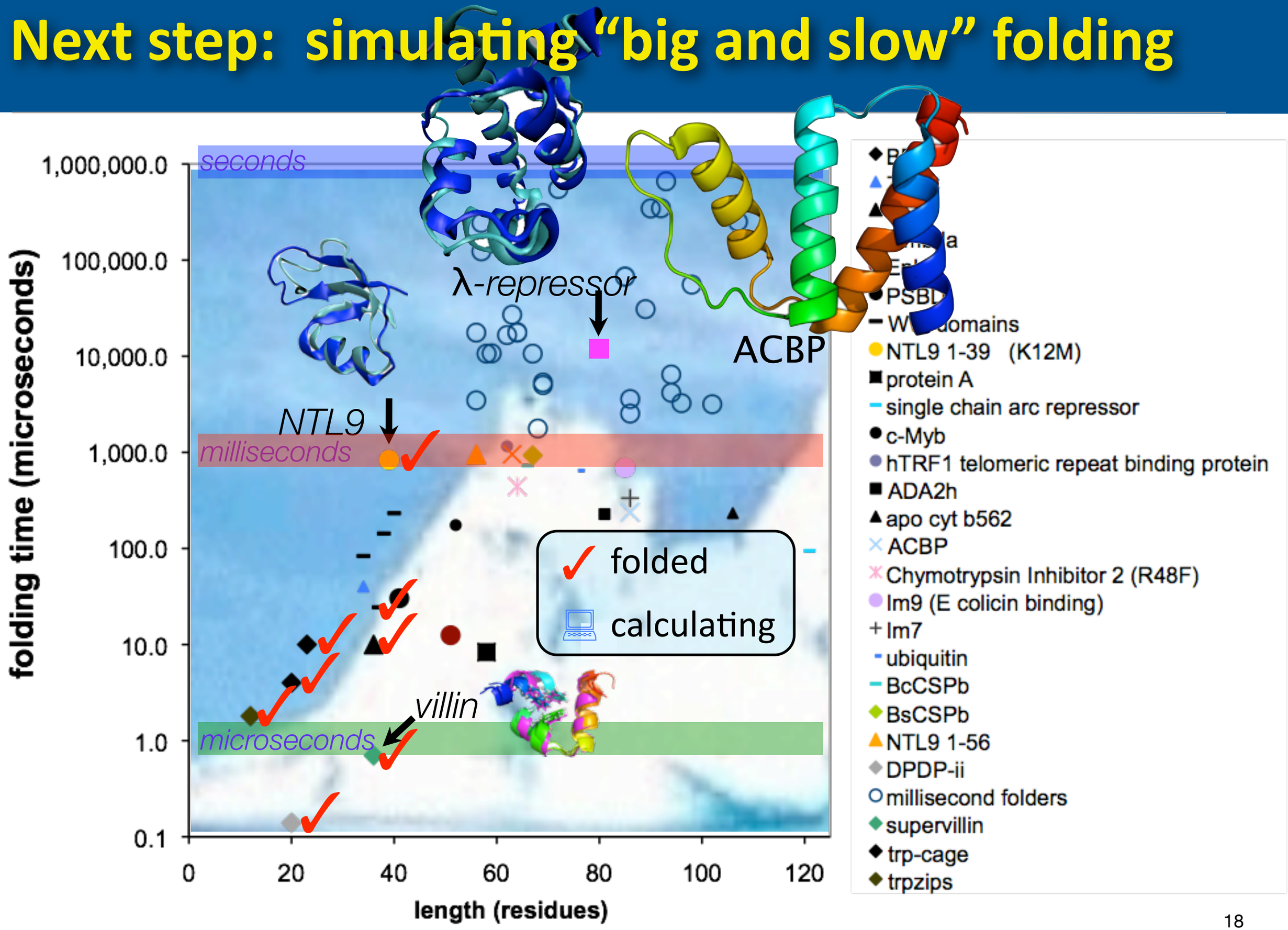
Top 10 folding pathways shows us:

- A great deal of pathway heterogeneity exists
- non-native structure plays a key role in many states
- metastability is often structurally localized (analogous to the foldon concept)

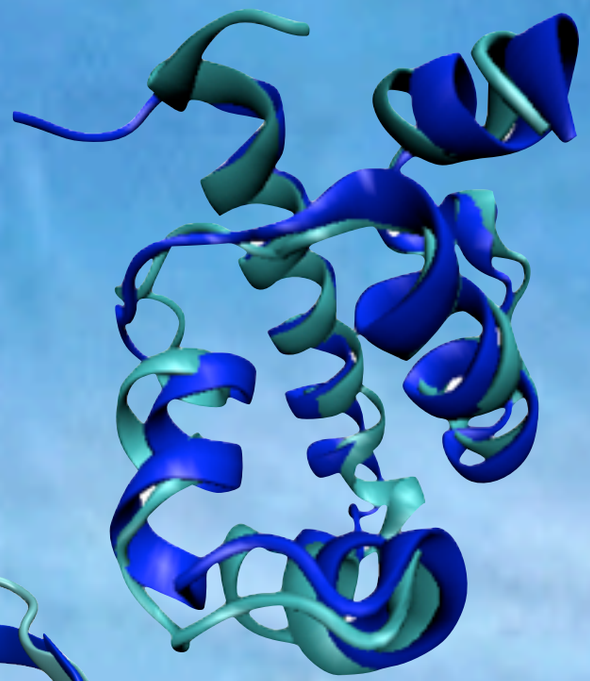
Next step: simulating "big and slow" folding



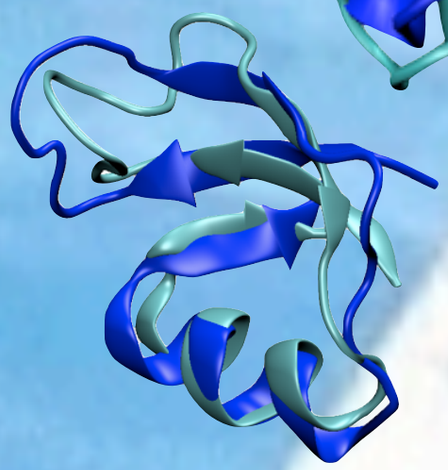
Next step: simulating "big and slow" folding



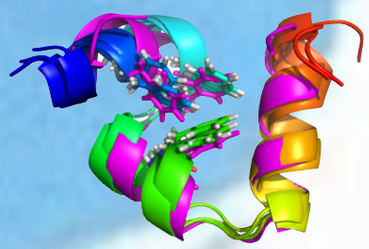
We are now getting close to the top of the mountain



Group 3/4:
Lambda repressor
(80 residues)

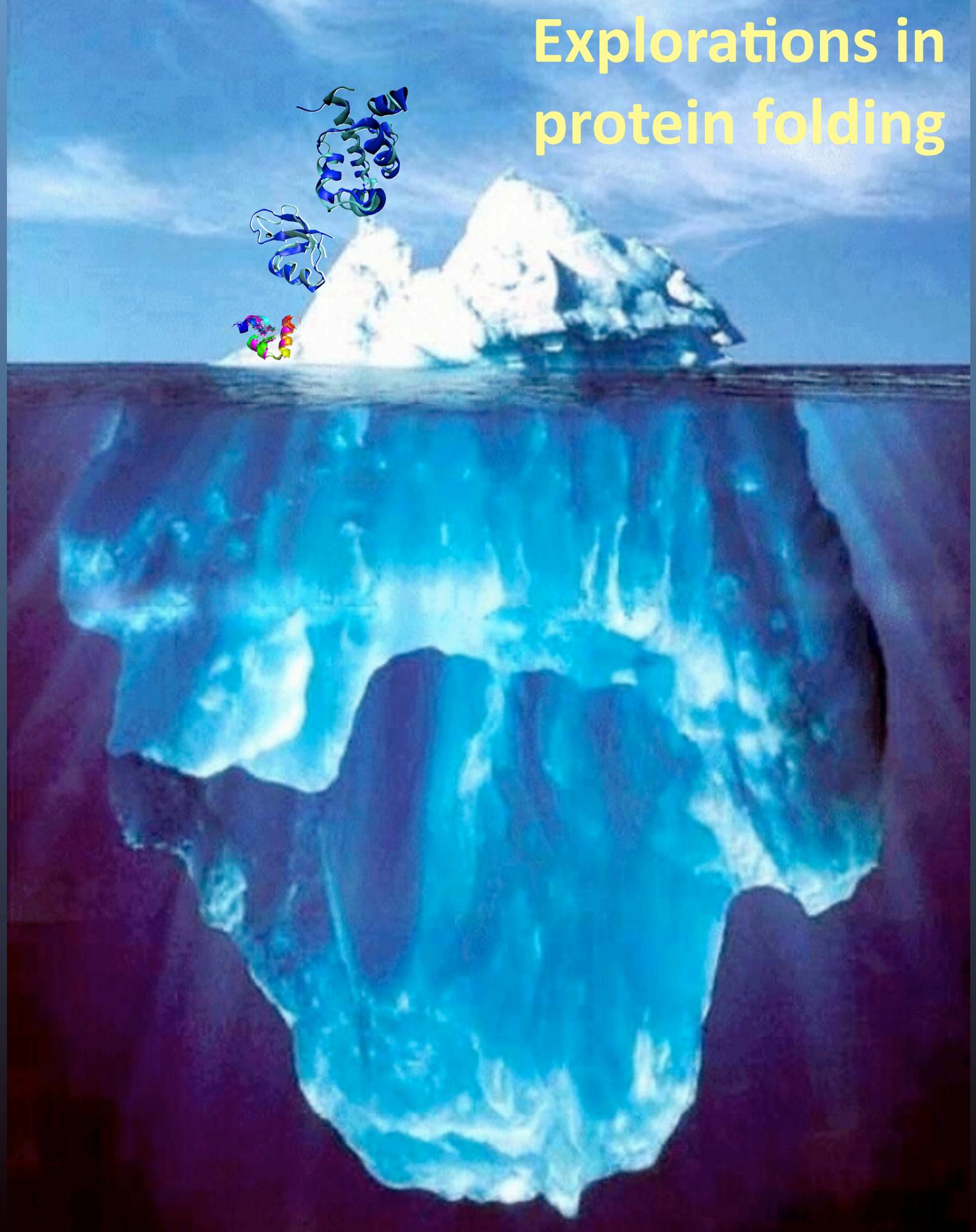


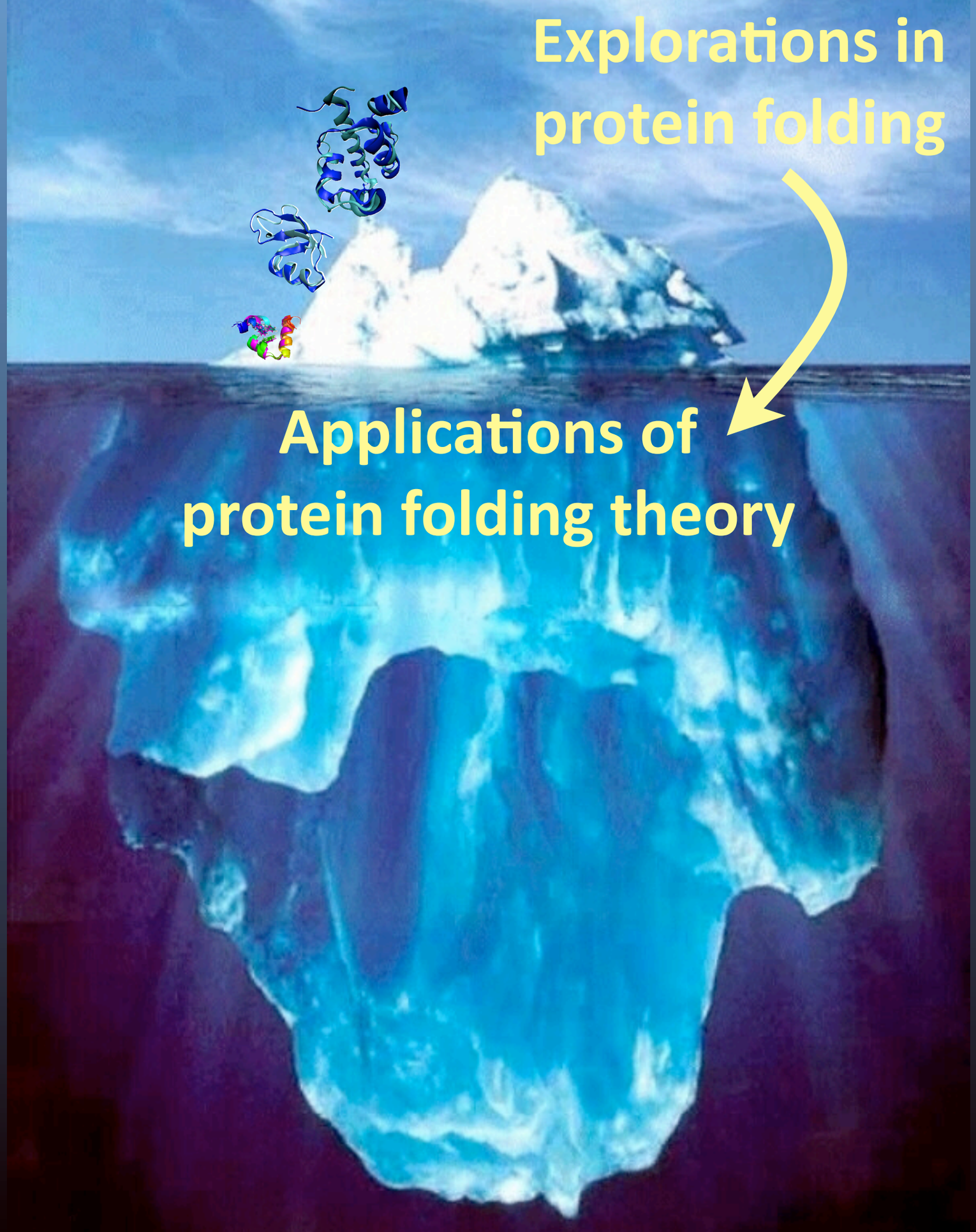
Group 2:
NTL9
(39 residues)



Group 1:
Villin headpiece
(36 residues)

Explorations in protein folding

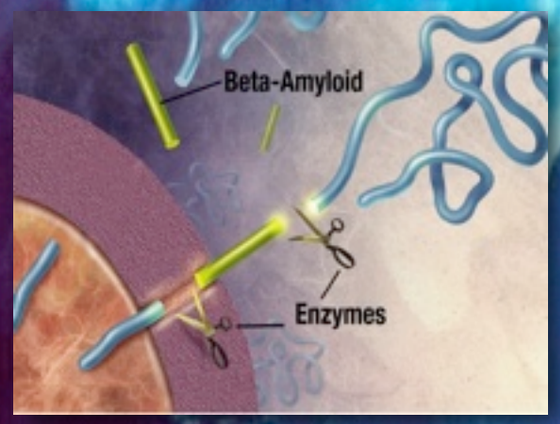



Explorations in
protein folding
Applications of
protein folding theory

Explorations in protein folding

Applications of protein folding theory

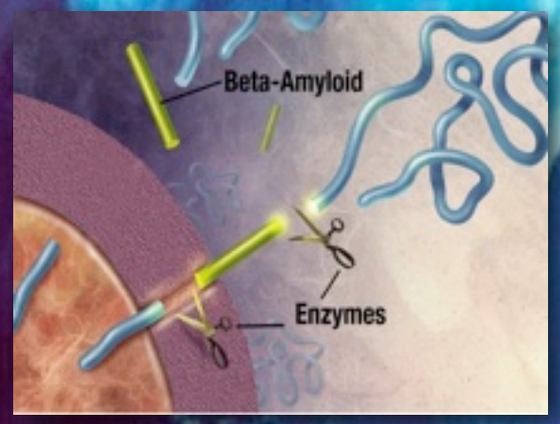
protein misfolding



Explorations in protein folding

Applications of protein folding theory

protein misfolding

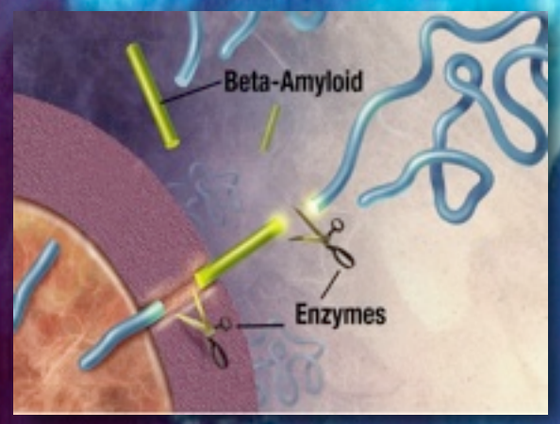


dynamics & function

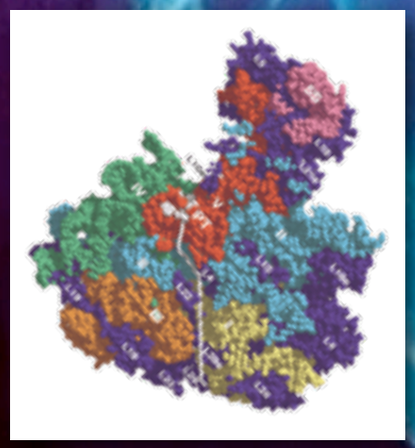
Explorations in protein folding

Applications of protein folding theory

protein misfolding

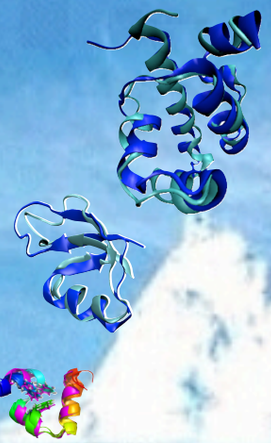


folding of nascent chains



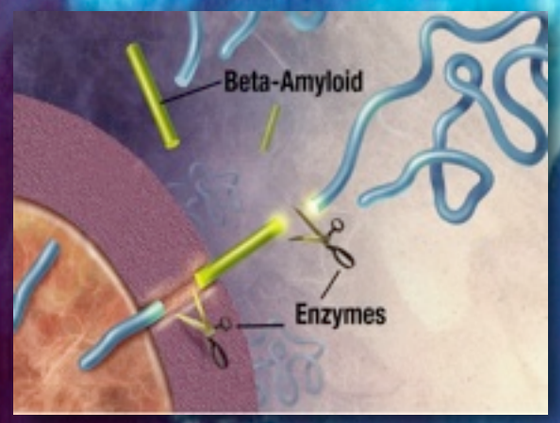
dynamics & function

Explorations in protein folding



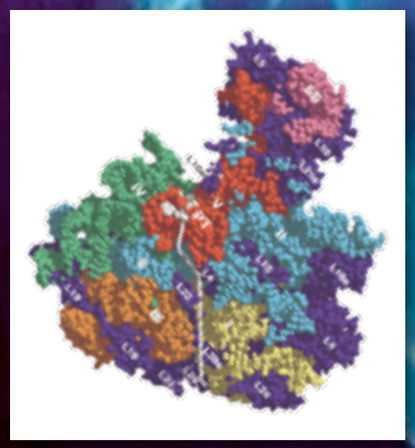
Applications of protein folding theory

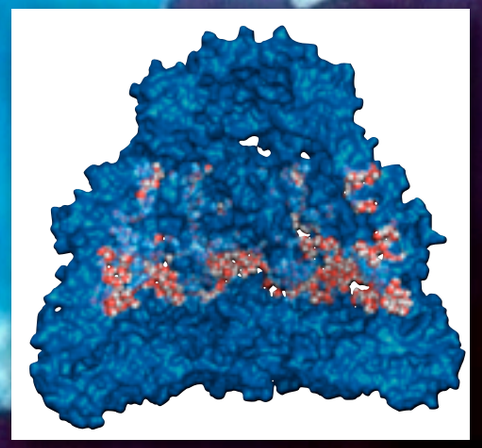
protein misfolding



dynamics & function

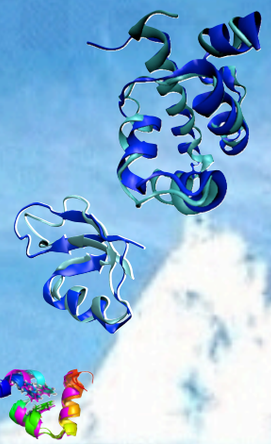
folding of nascent chains





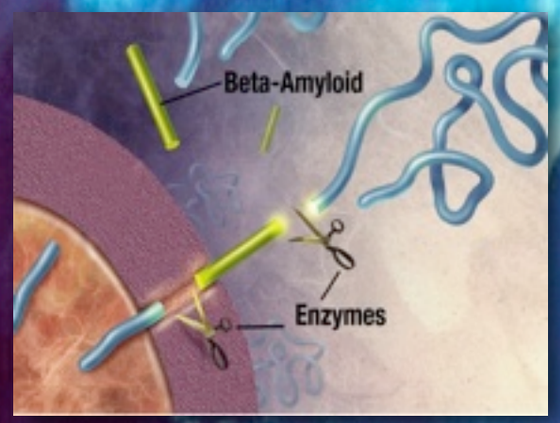
chaperonin function

Explorations in protein folding

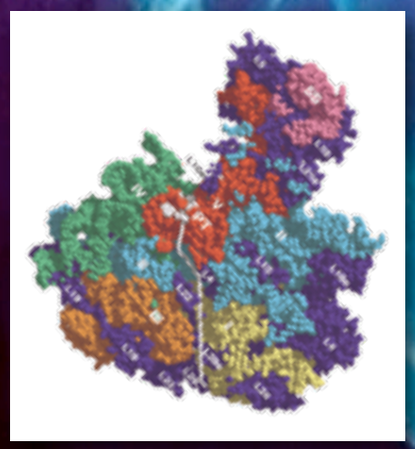


Applications of protein folding theory

protein misfolding

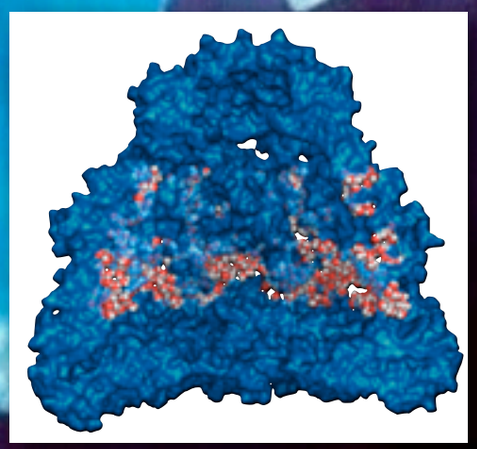


folding of nascent chains

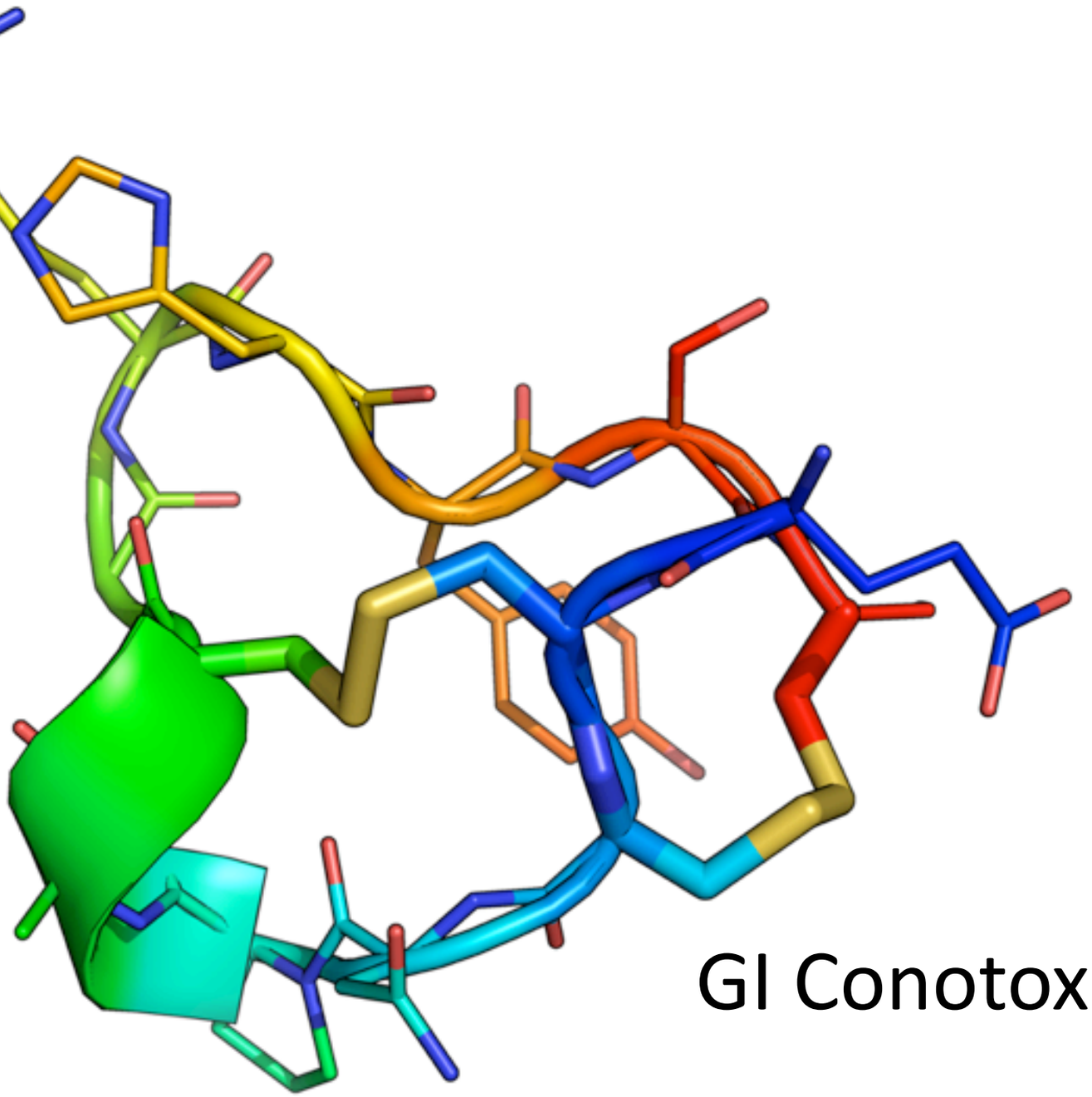


dynamics & function

chaperonin function

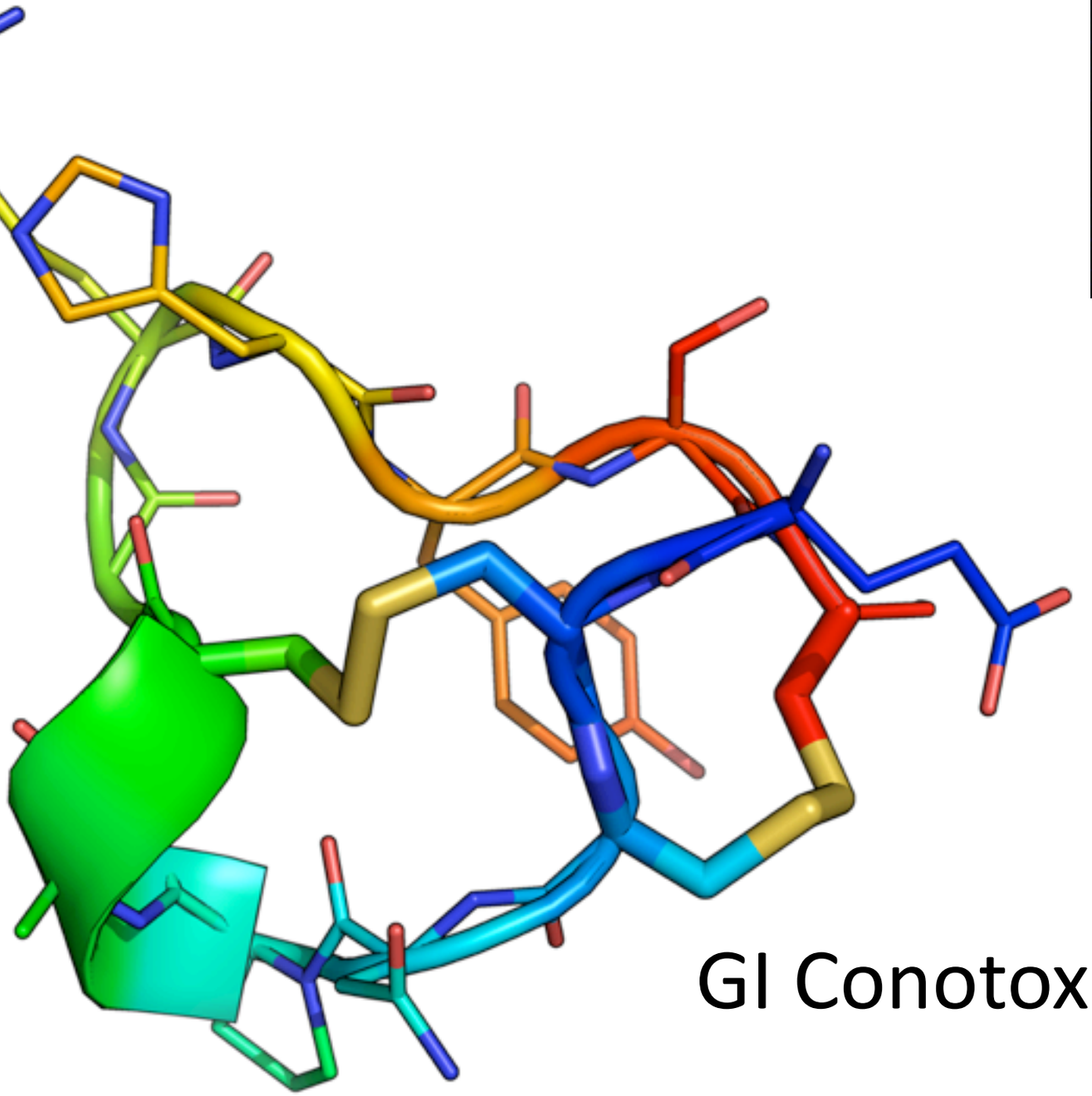


The Hope of Dynamically Derived Theraputics



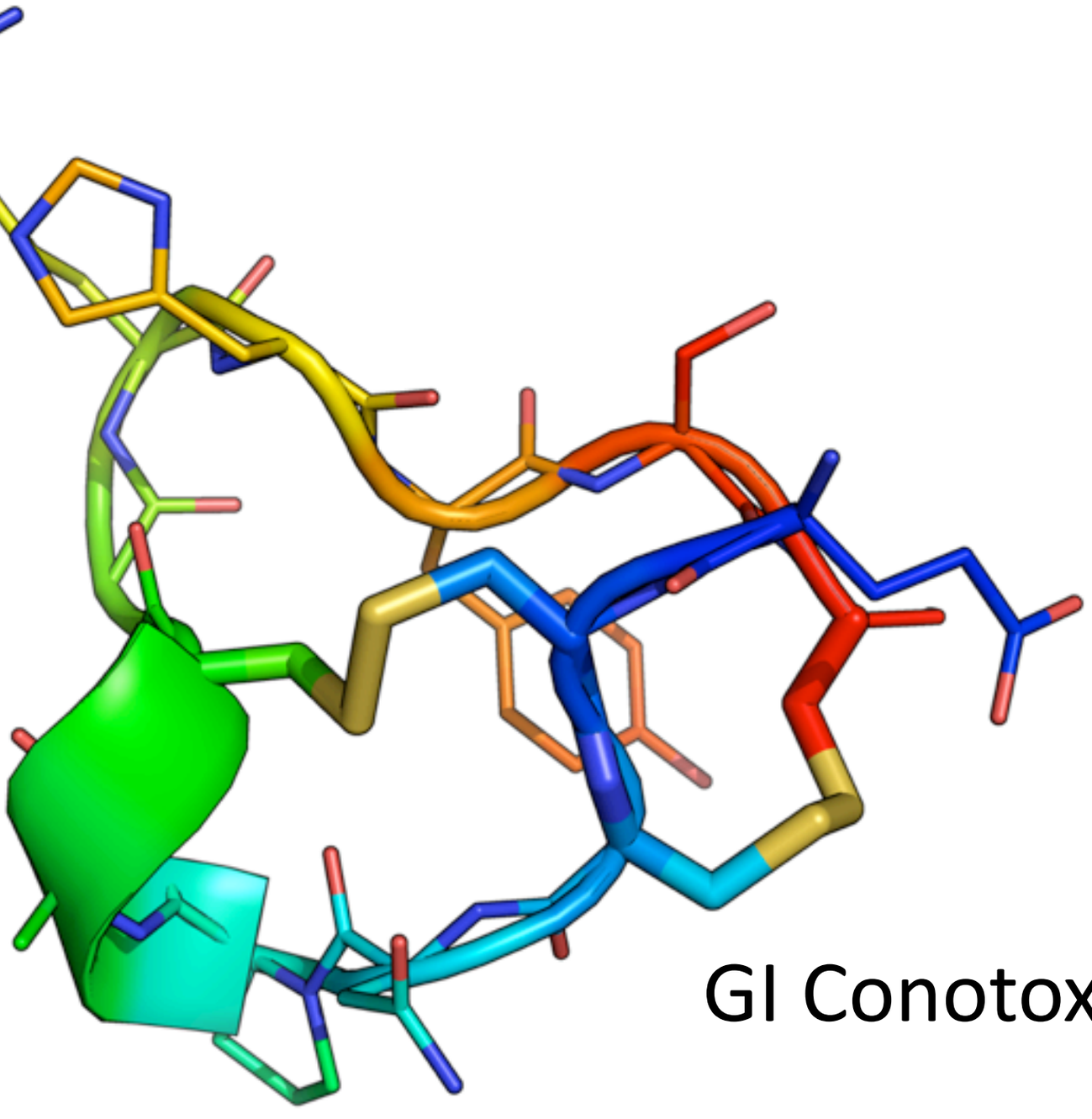
GI Conotoxin

The Hope of Dynamically Derived Theraputics



GI Conotoxin

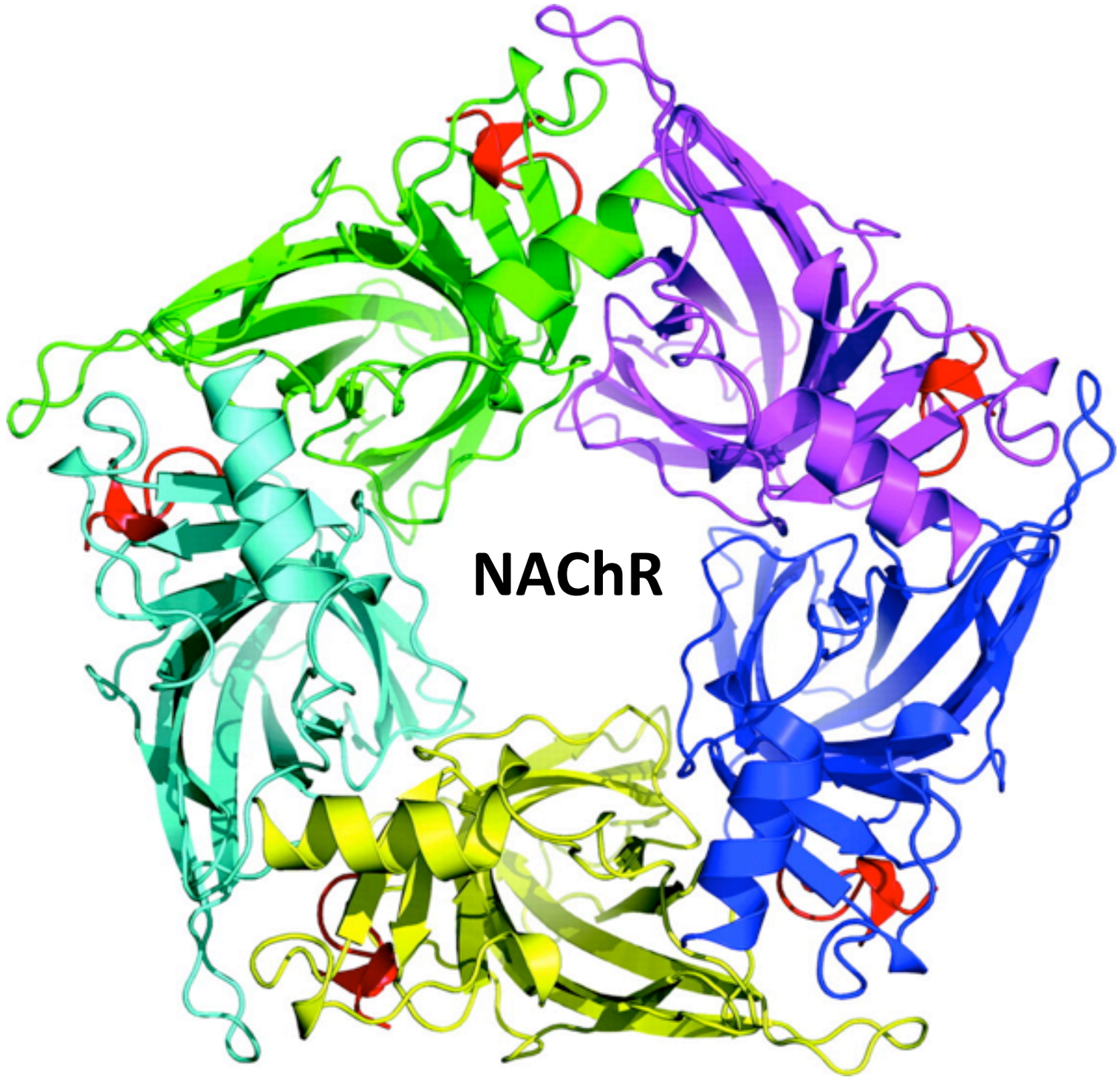
The Hope of Dynamically Derived Theraputics



GI Conotoxin

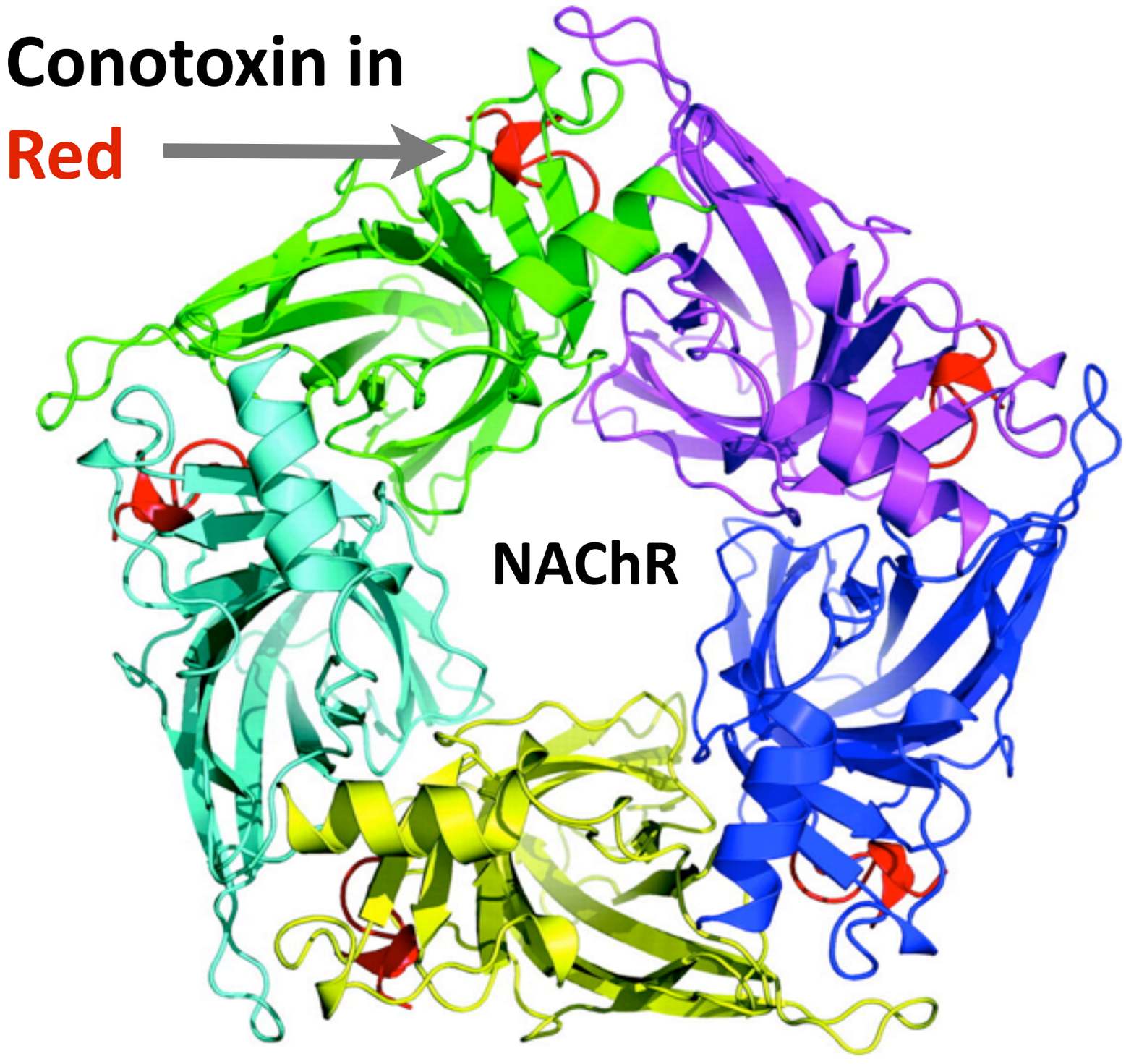
Ziconotide (Prialt)
100-1000x as potent
as morphine

GI Offers a Number of Appealing Advantages



Ulens et. al. PNAS March 7, 2006 vol. 103 no. 10 3615-3620

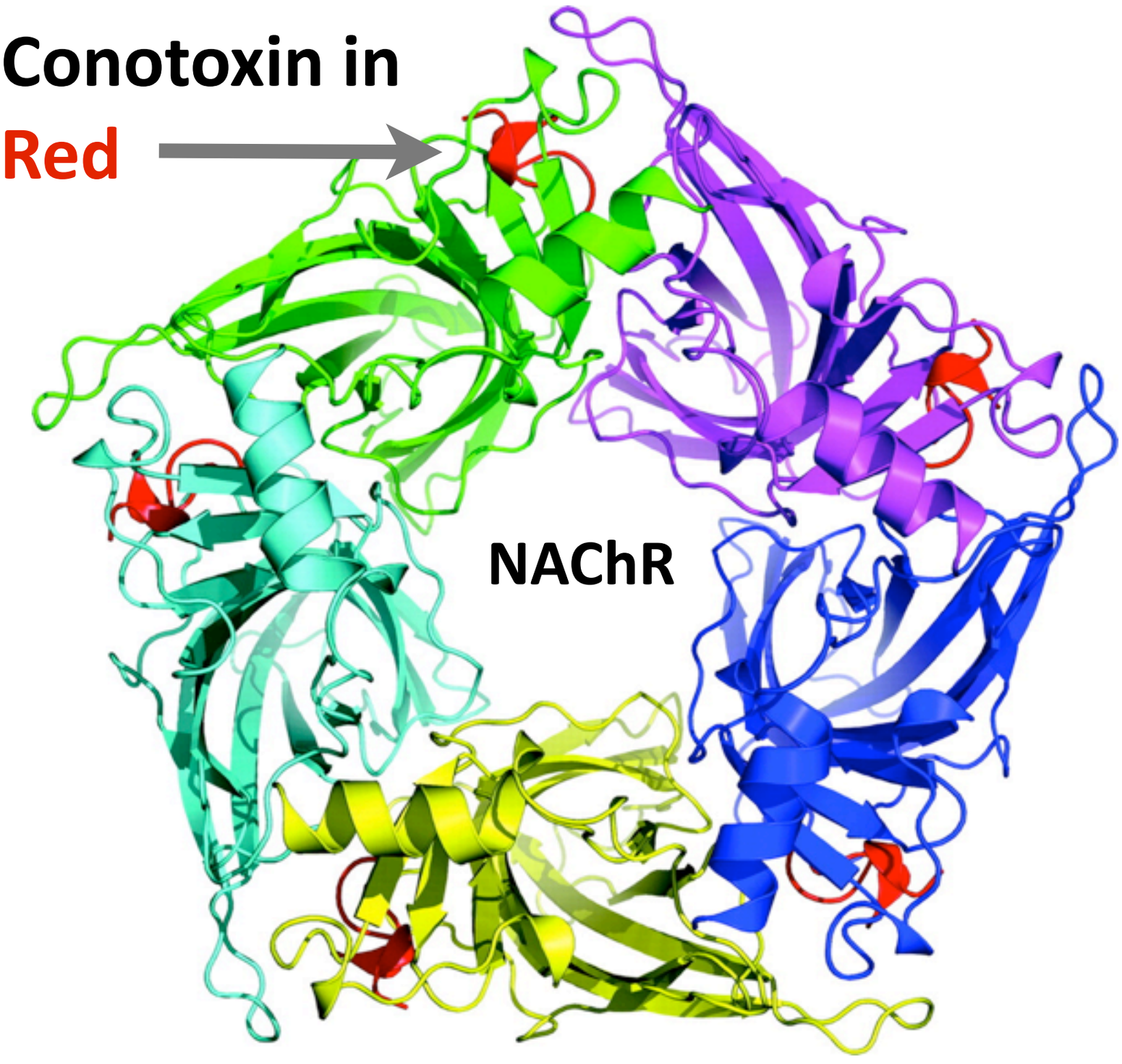
GI Offers a Number of Appealing Advantages



Ulens et. al. PNAS March 7, 2006 vol. 103 no. 10 3615-3620

GI Offers a Number of Appealing Advantages

1. **GI is Small:** Good ADME (adsorption, distribution, metabolism, excretion)

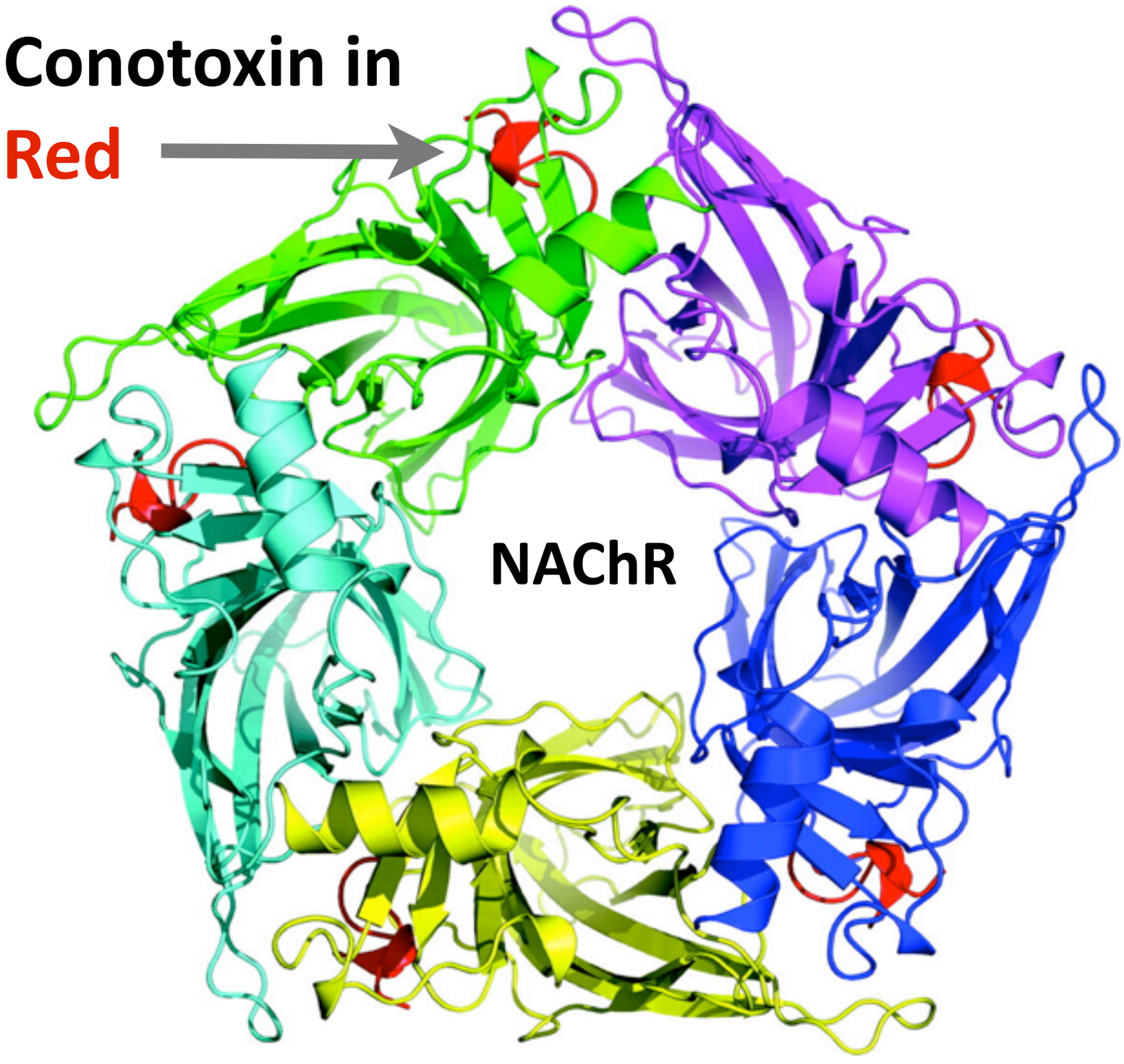


Ulens et. al. PNAS March 7, 2006 vol. 103 no. 10 3615-3620

GI Offers a Number of Appealing Advantages

1. **GI is Small:** Good ADME (adsorption, distribution, metabolism, excretion)

2. **Easily Synthesized:** Peptide synthesizers can create grams robotically



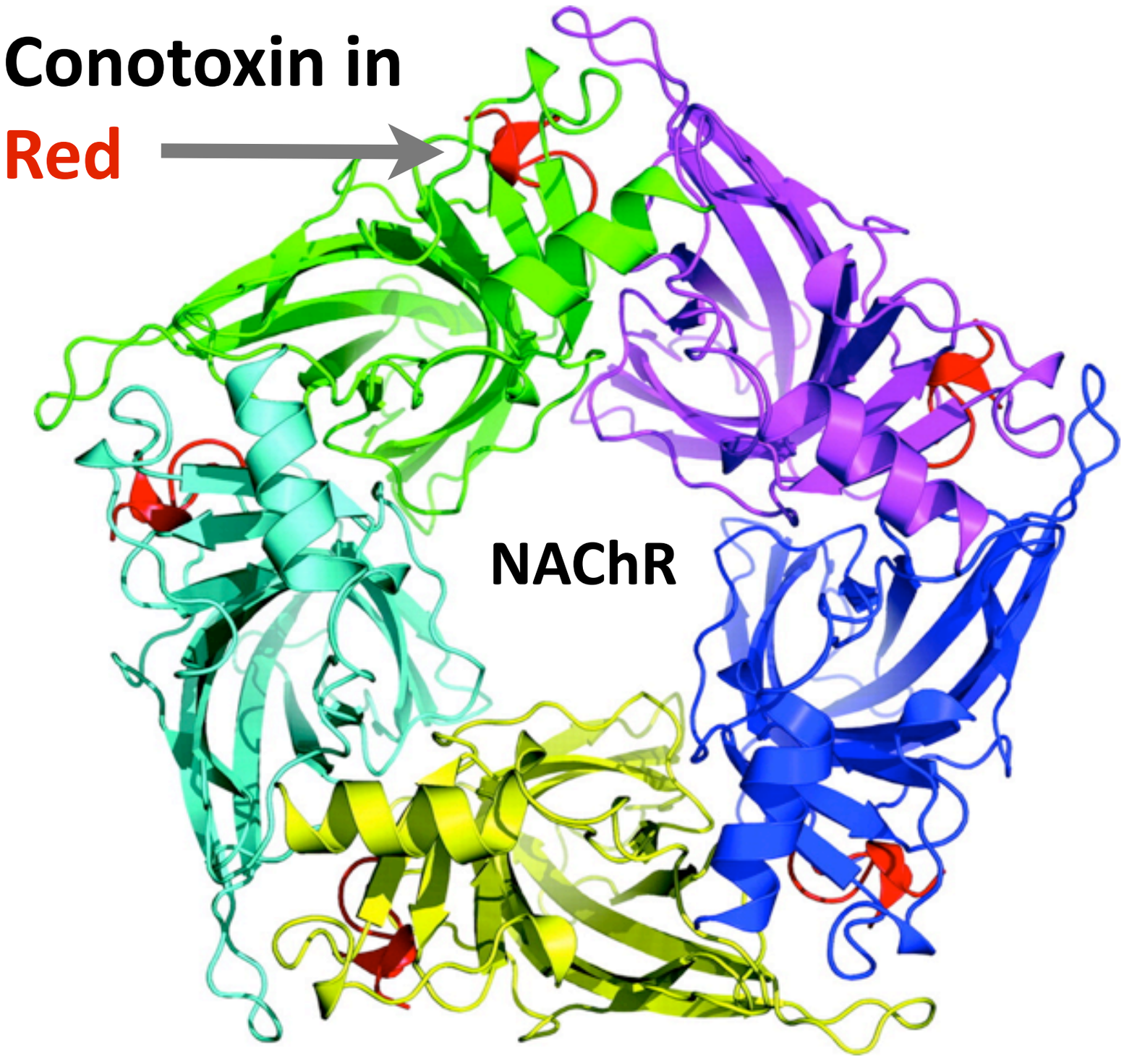
Ulens et. al. PNAS March 7, 2006 vol. 103 no. 10 3615-3620

GI Offers a Number of Appealing Advantages

1. **GI is Small:** Good ADME (adsorption, distribution, metabolism, excretion)

2. **Easily Synthesized:** Peptide synthesizers can create grams robotically

3. **Easily Modified:** Mutagenesis is a well-studied process that is easy to realize computationally and in the lab



Ulens et. al. PNAS March 7, 2006 vol. 103 no. 10 3615-3620

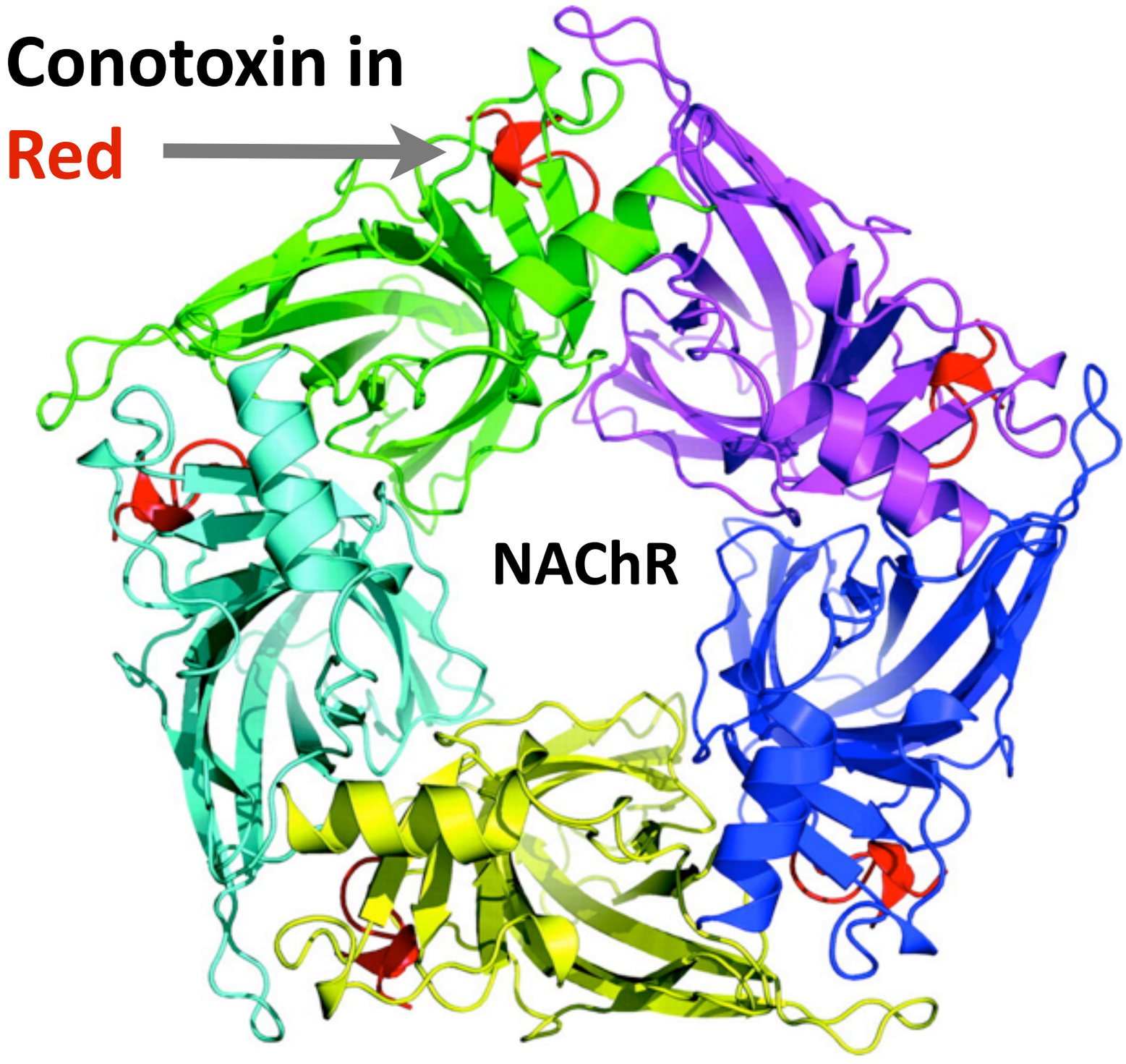
GI Offers a Number of Appealing Advantages

1. **GI is Small:** Good ADME (adsorption, distribution, metabolism, excretion)

2. **Easily Synthesized:** Peptide synthesizers can create grams robotically

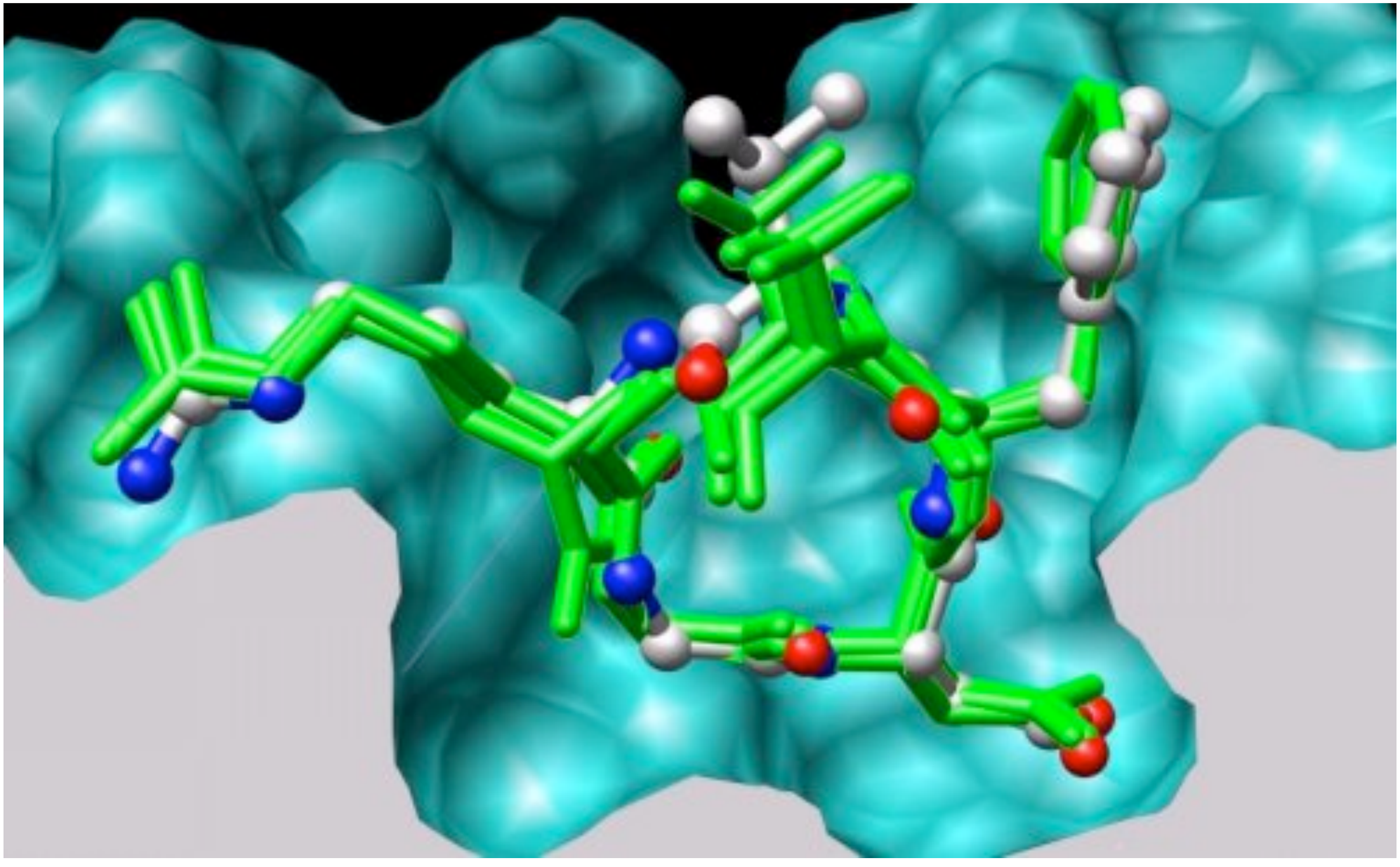
3. **Easily Modified:** Mutagenesis is a well-studied process that is easy to realize computationally and in the lab

4. **There are 100s of similar "disulfide-rich" proteins**



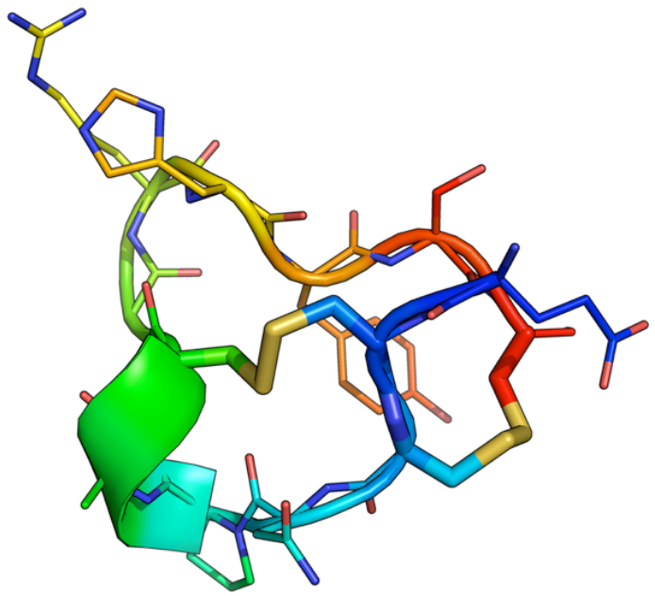
Ulens et. al. PNAS March 7, 2006 vol. 103 no. 10 3615-3620

Docking

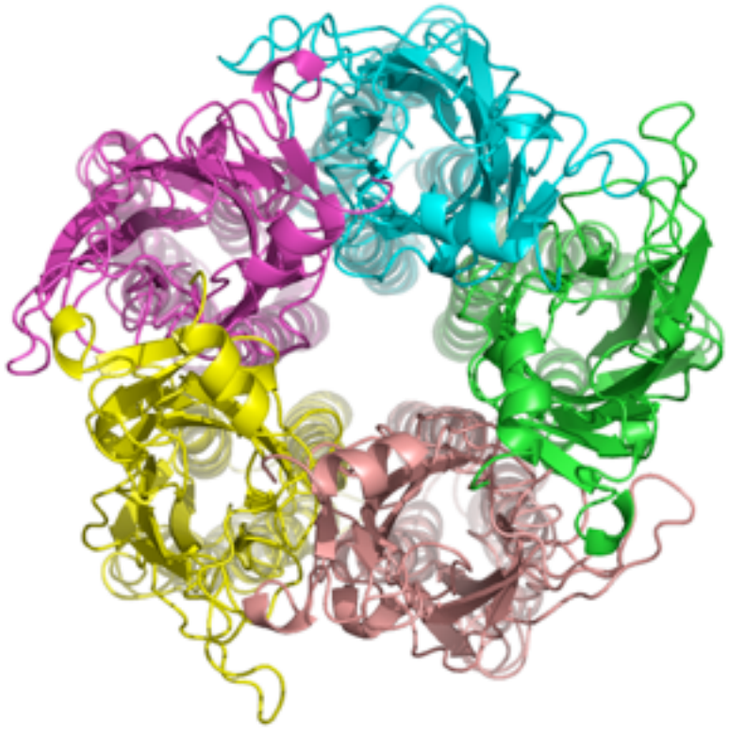


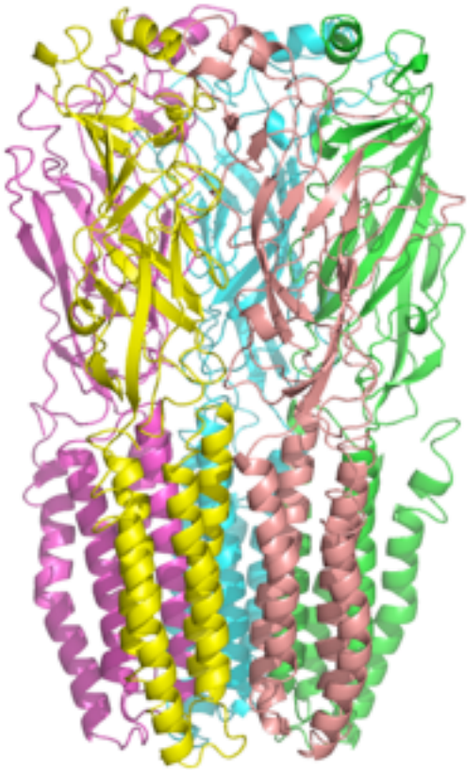
MSM Methods May Be Employed

Ligand



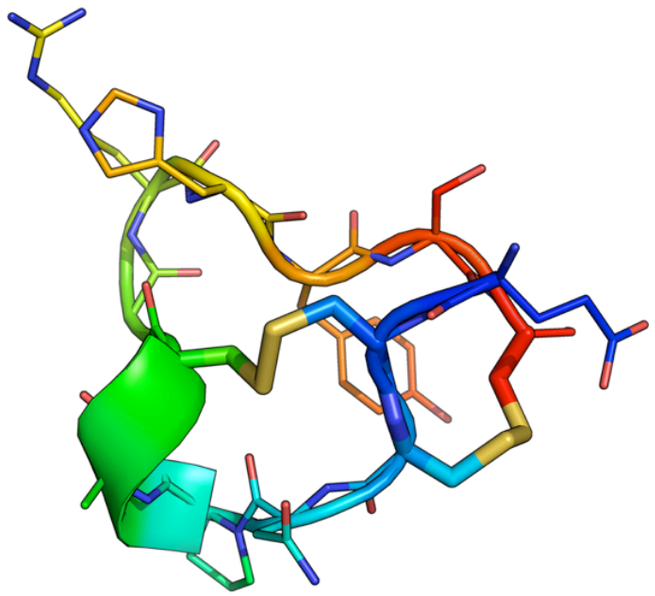
Target

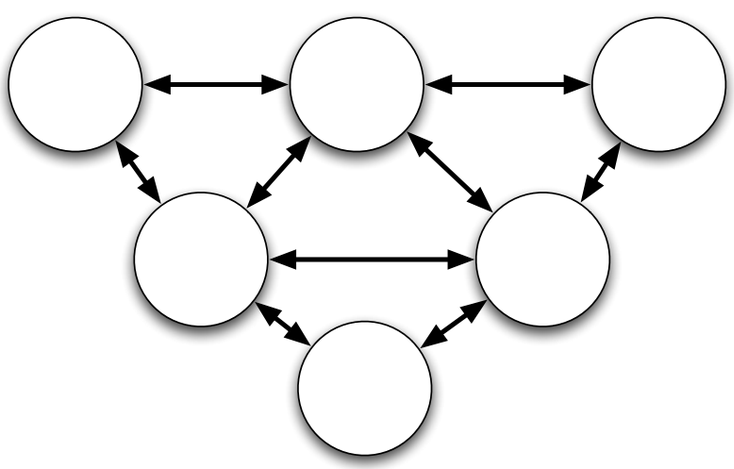




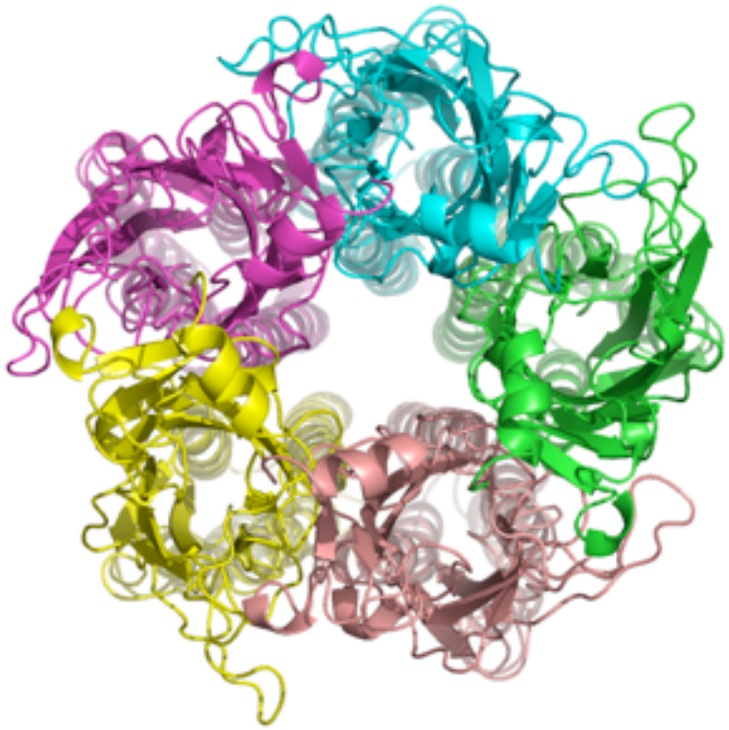
MSM Methods May Be Employed

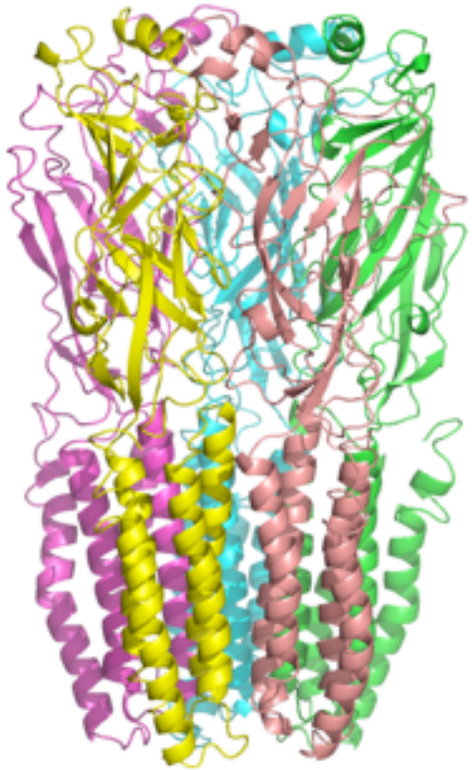
Ligand

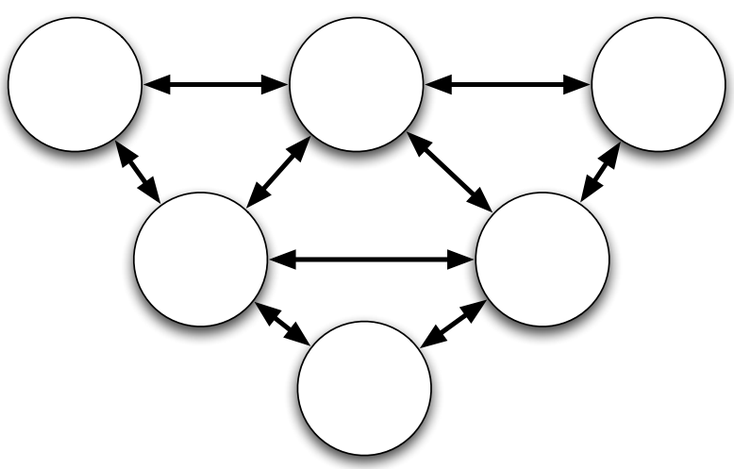




Target

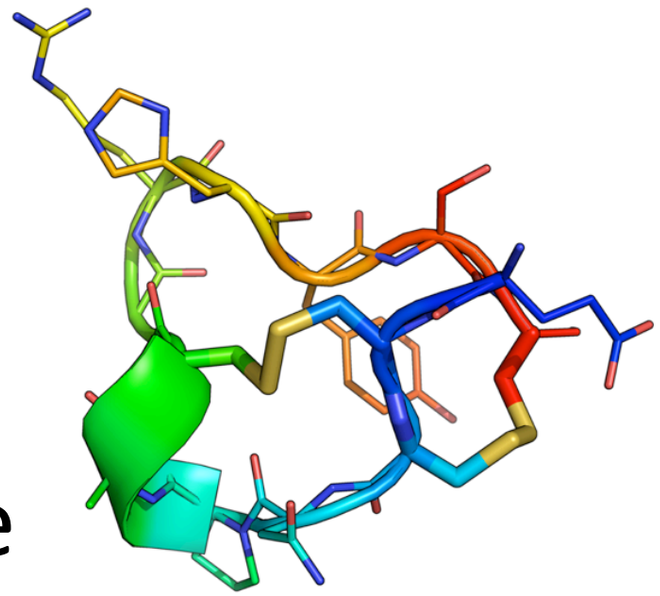




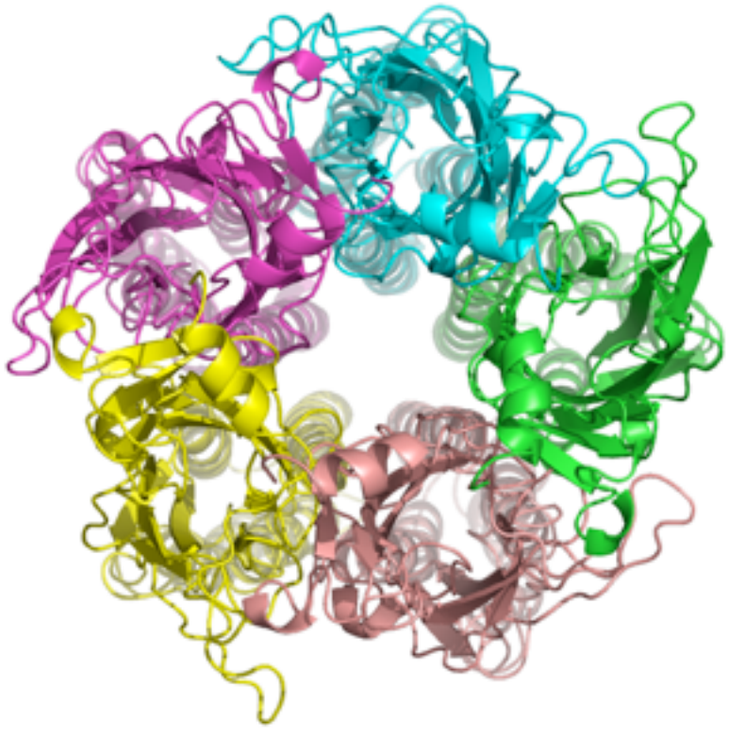


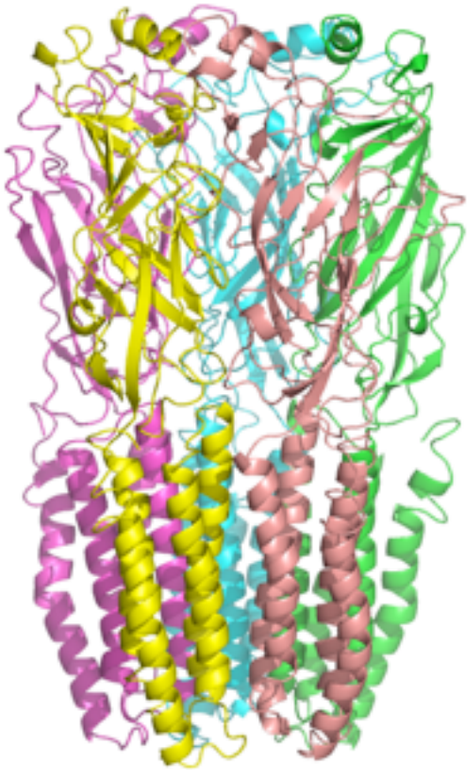
MSM Methods May Be Employed

Ligand

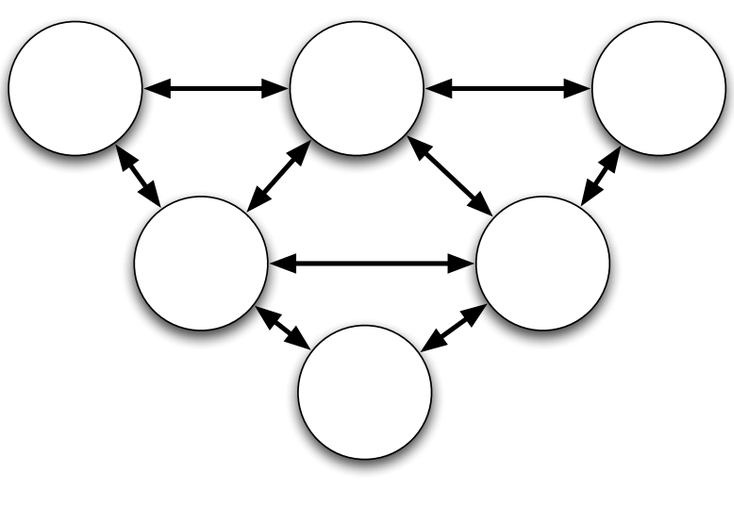


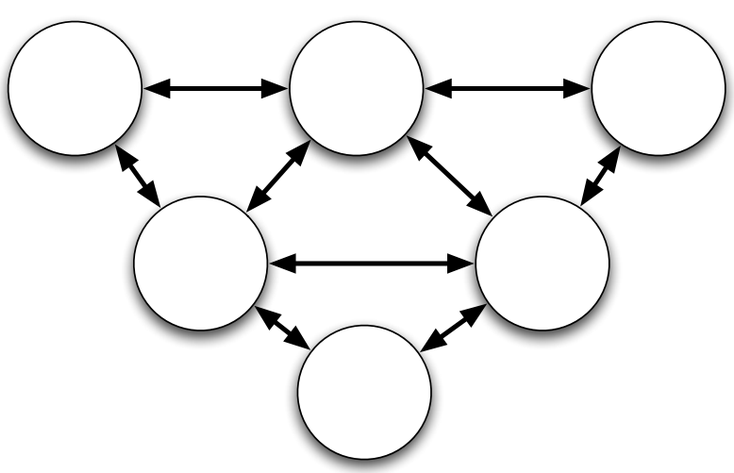
Target





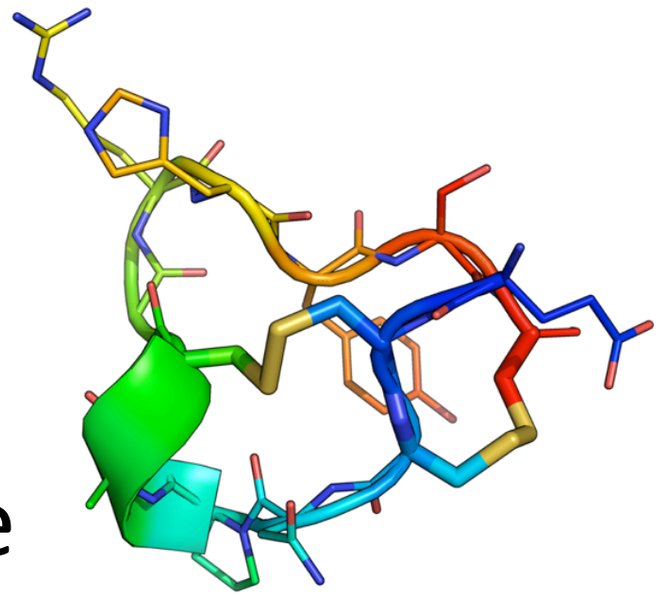
Free Energy



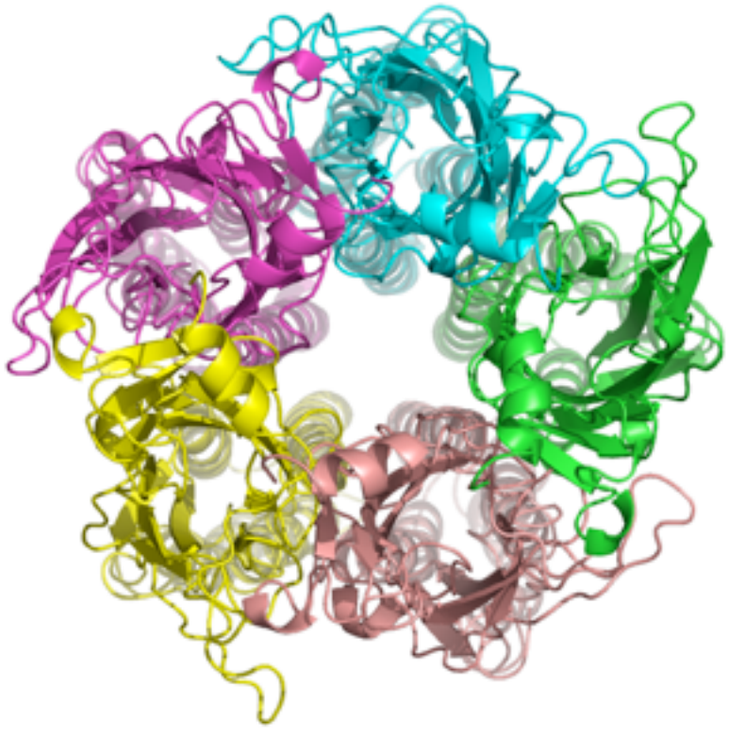


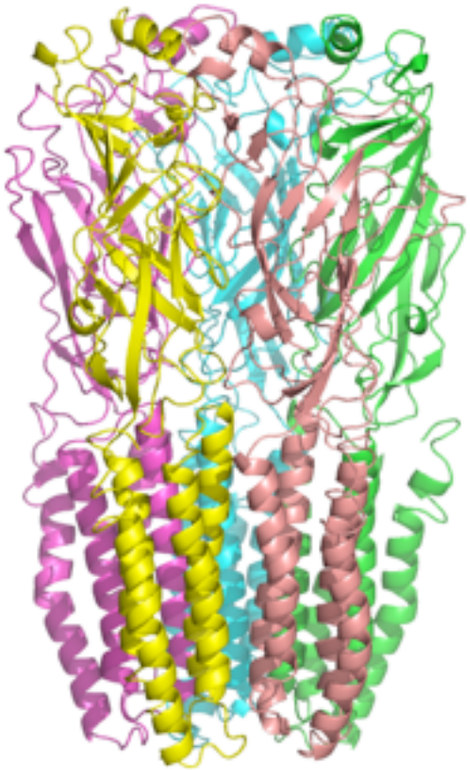
MSM Methods May Be Employed

Ligand



Target



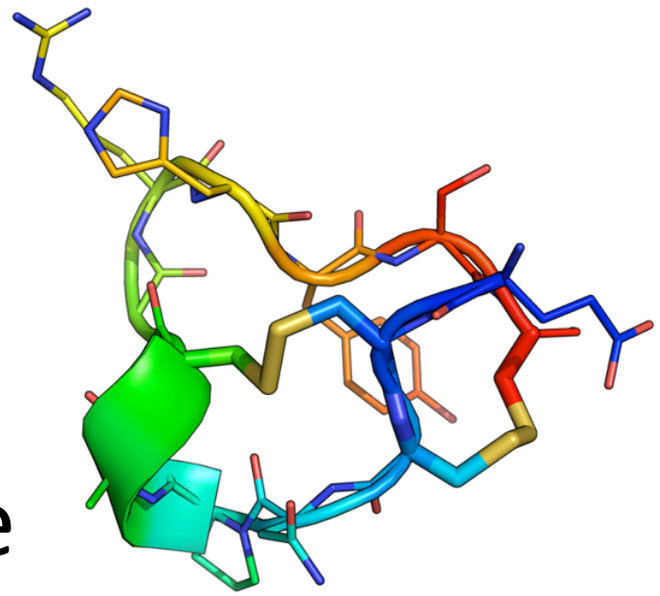


Free Energy

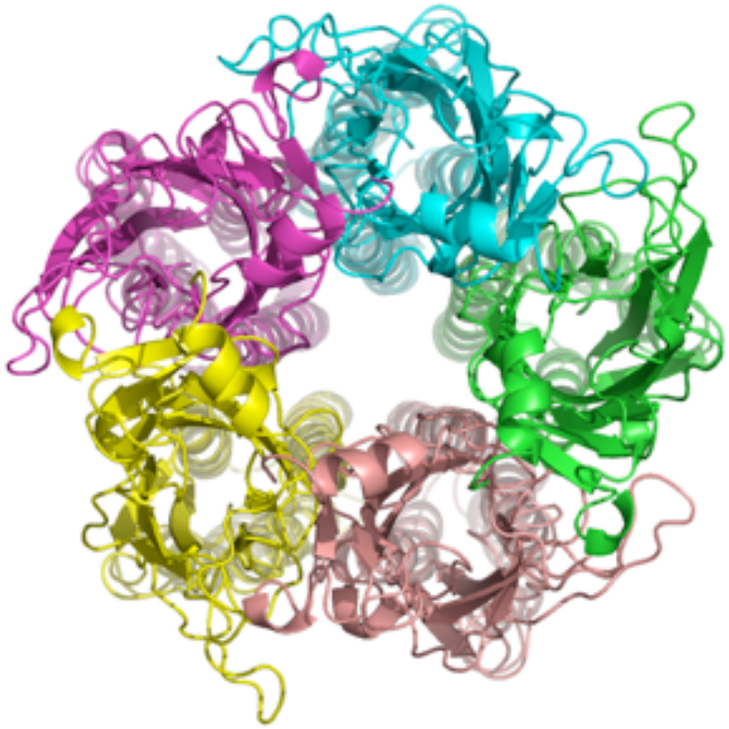
traditional

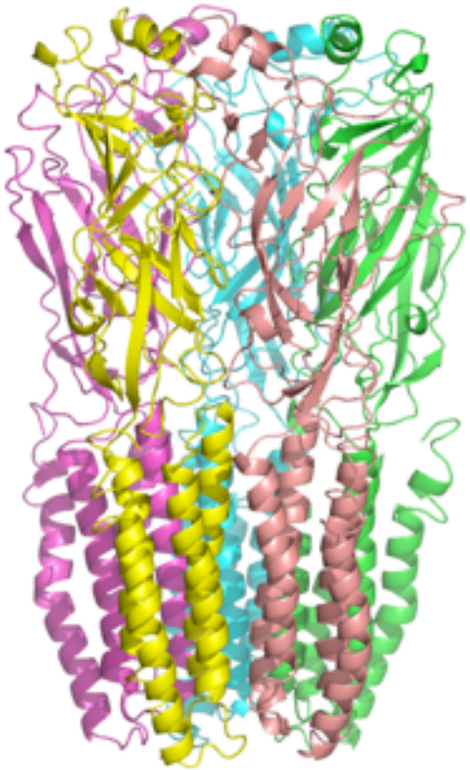
MSM Methods May Be Employed

Ligand

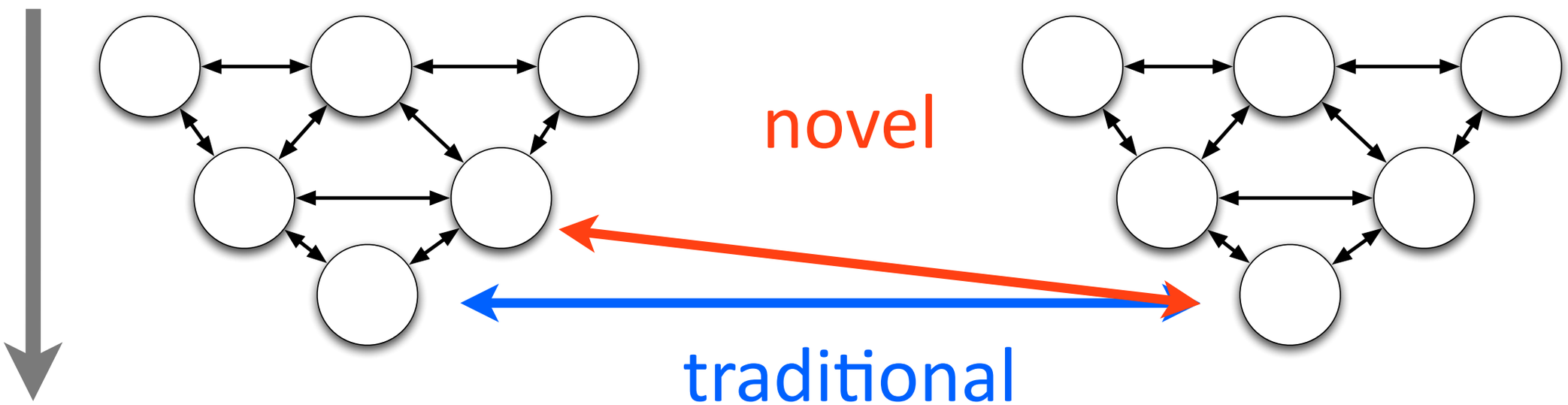


Target



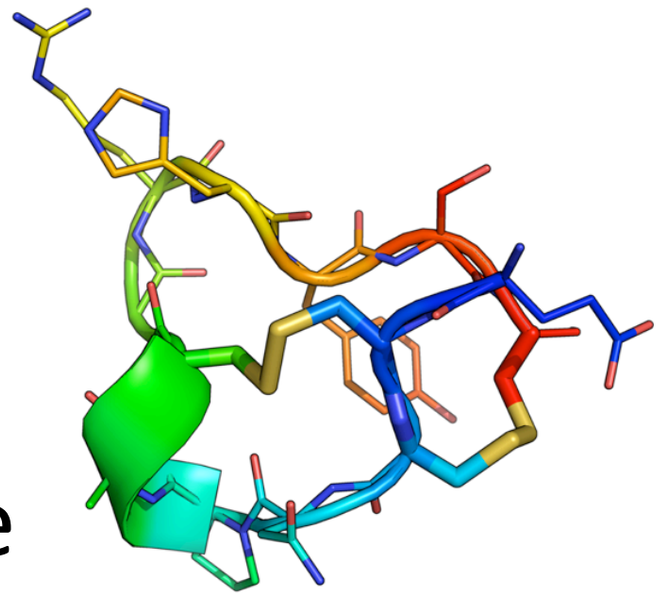


Free Energy

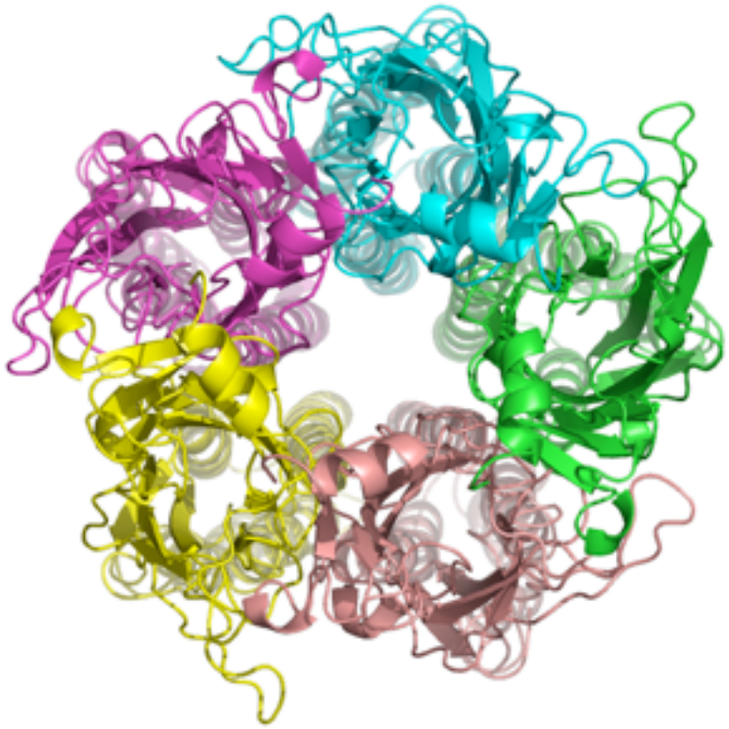


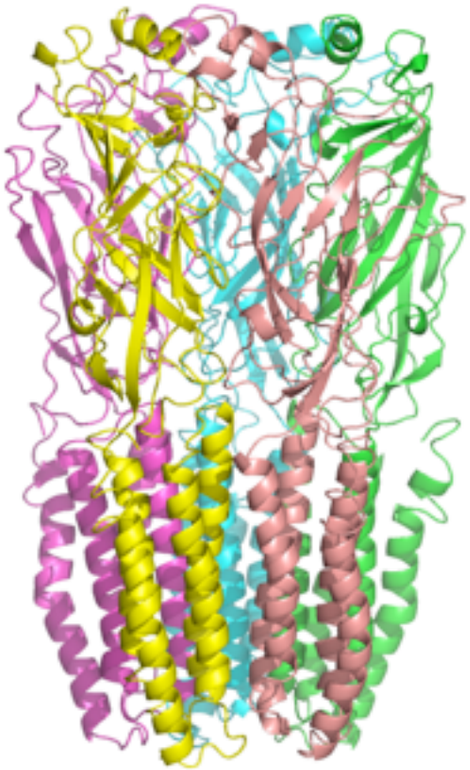
MSM Methods May Be Employed

Ligand

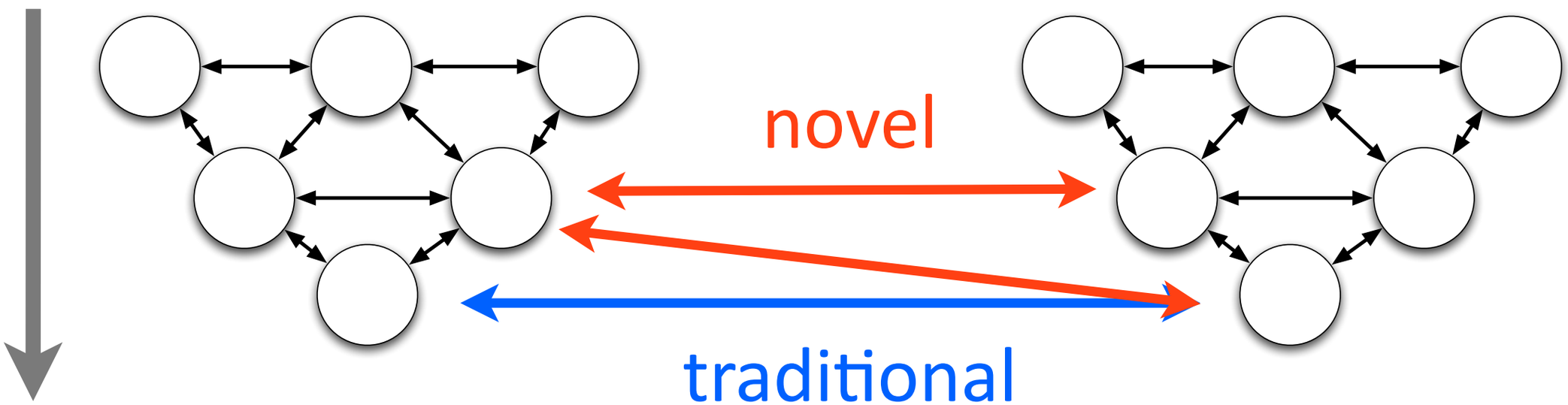


Target

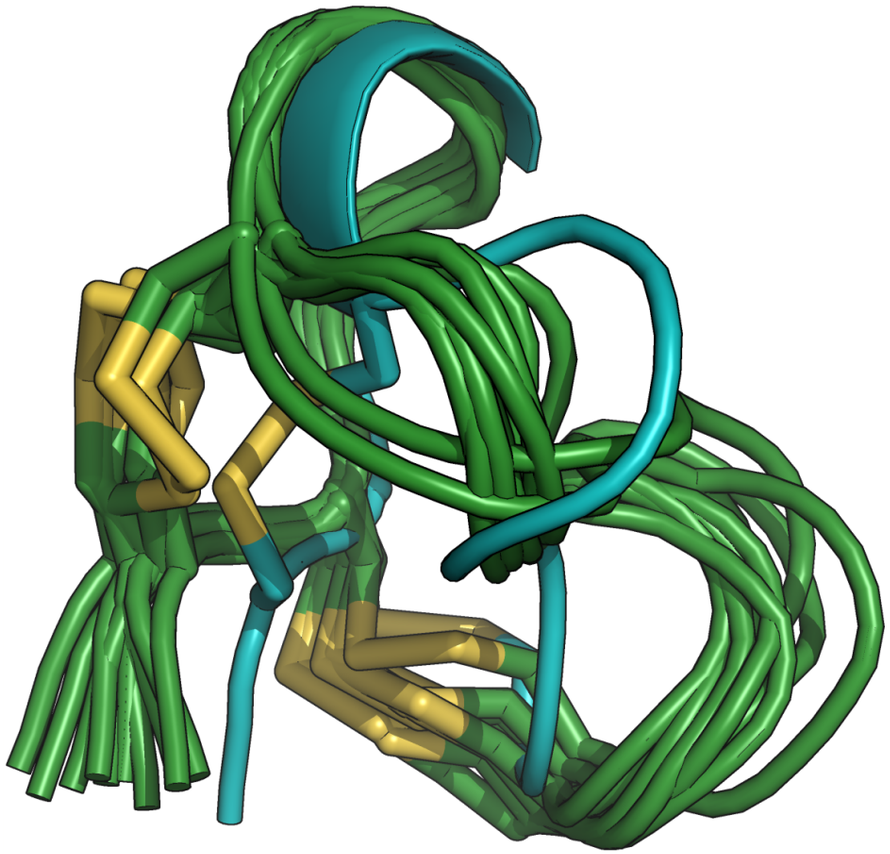


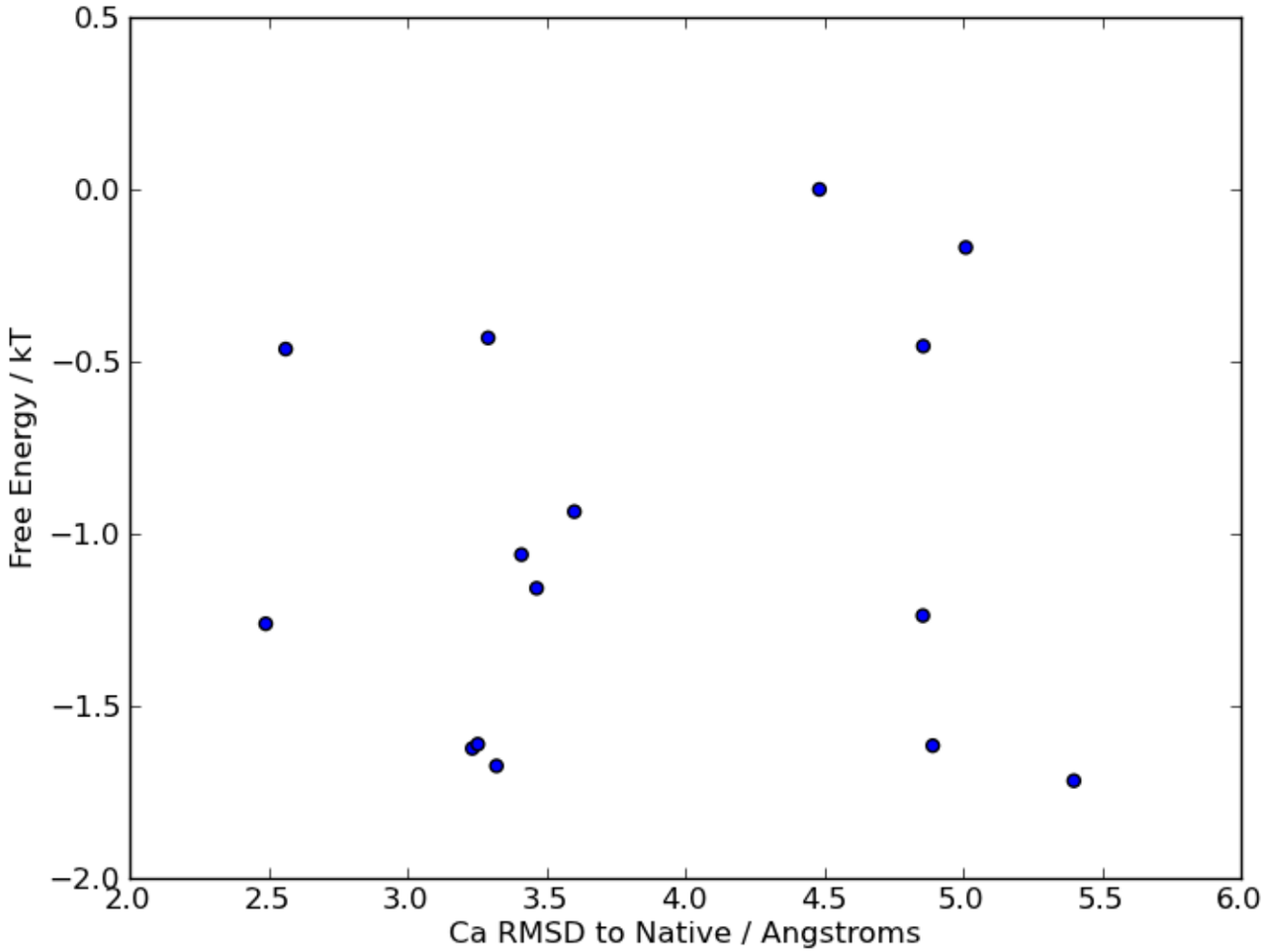


Free Energy



Some preliminary results...



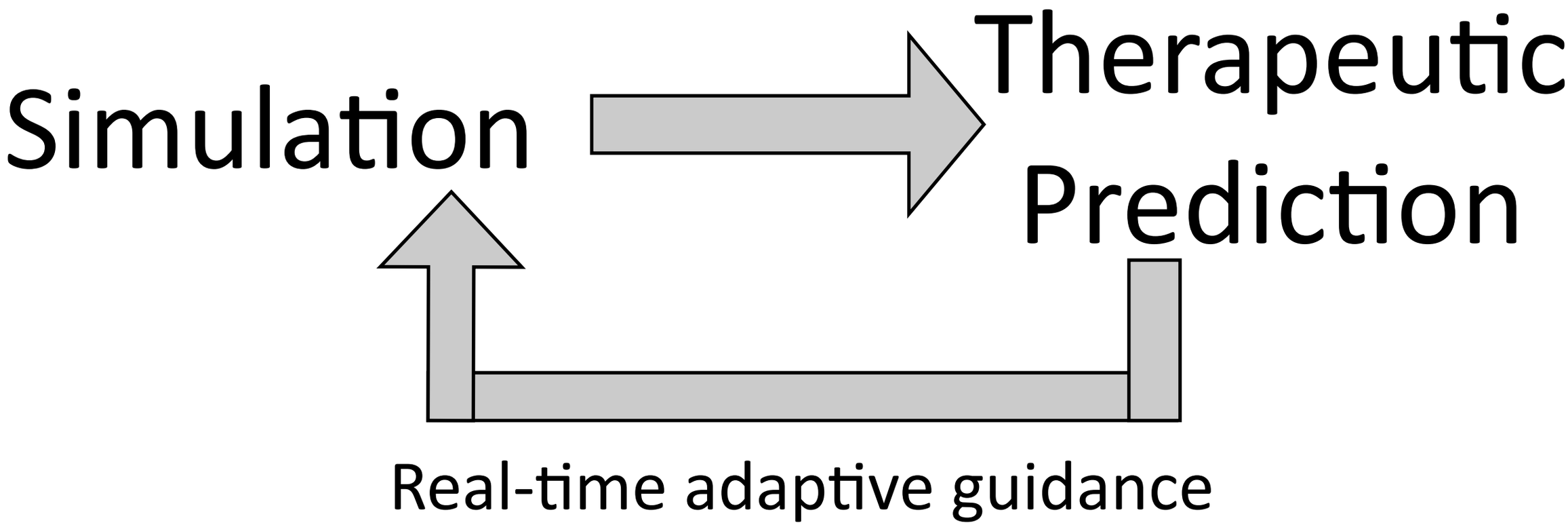


We still have some work to do for general prediction...

Why, then, a supercomputer?

Simulation → Therapeutic Prediction

Why, then, a supercomputer?



protein folding

simulations: K. Beauchamp, G. Bowman, D. Ensign, P. Kasson, V. Volez (Pande Lab, Stanford)

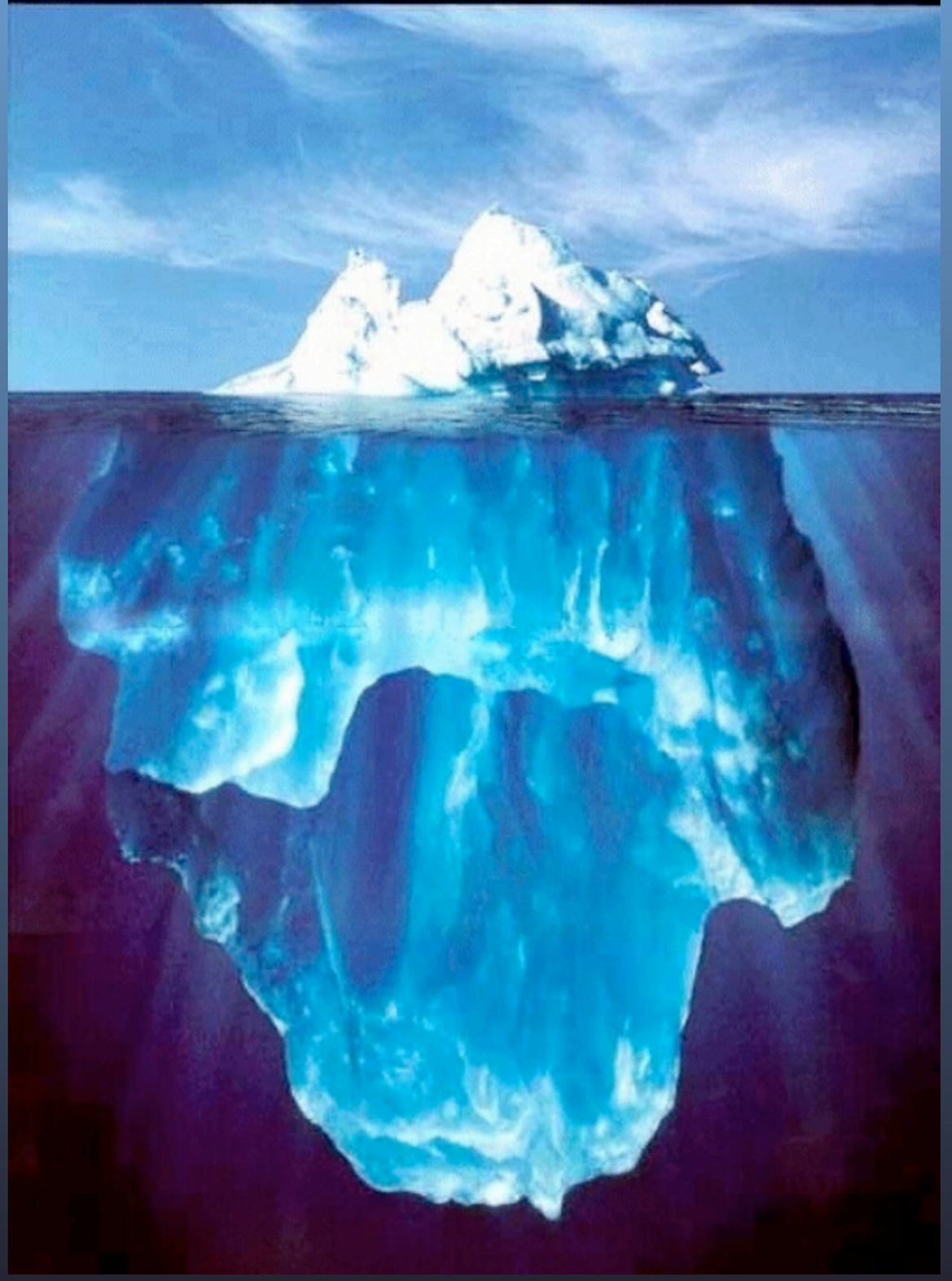
experiments: V. Singh, L. Lapidus (Lapidus Lab, Michigan State); S. Weis (UCLA)

MSM theory

S. Baccalado, K. Beauchamp, G. Bowman, J. Chodera [now @UCB], X. Huang [now @HK], S. Park [now @ANL], N. Singhal [now @Chicago] (Pande Lab, Stanford); Bill Swope & Jed Pitera (IBM); Ken Dill (UCSF)

funding

NIH Nanomedicine Protein Folding Center, NIH Roadmap Center Simbios, NIH NIGMS, NSF Chemistry, NSF Molecular Biophysics



tjlane@stanford.edu
http://pande.stanford.edu
http://folding.stanford.edu