Molecular simulation and structure prediction using CHARMM and the MMTSB Tool Set
Coarse-grained Models

Charles L. Brooks III
MMTSB/CTBP
2009 Summer Workshop

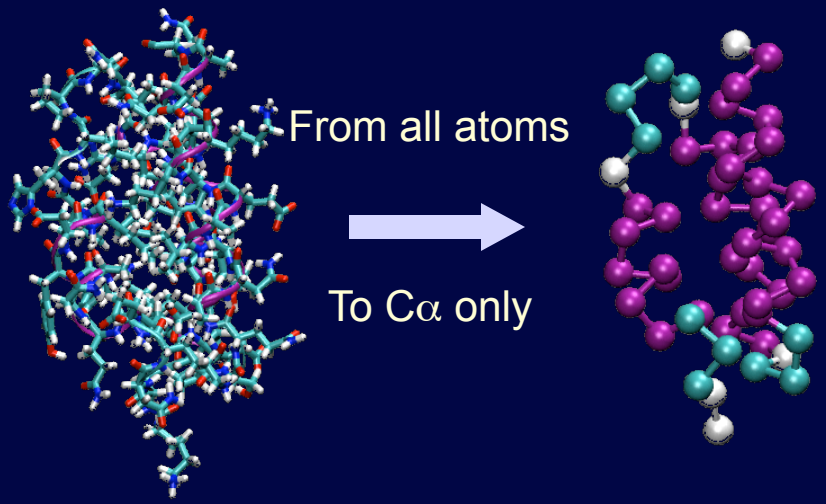
Developing coarse-grained models in CHARMM

- Case studies
  - Cα-based Go models
    - Encode native interactions via Cα-Cα "contacts", coarse-grain to level of one "bead" per aa
    - Useful as complement to protein folding mechanism studies
    - Helpful in understanding/informing single molecule "pulling" studies
    - MMTSB server available to provide "flavored" Go models for such studies
      - http://www.mmtsb.org/webservices/gomodel.html

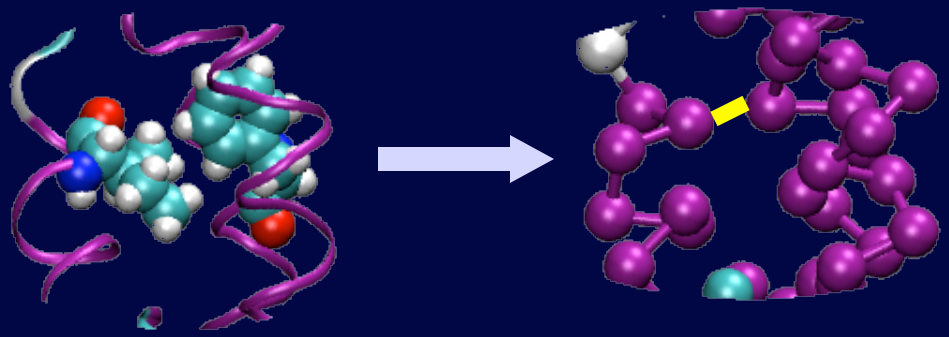
Developing coarse-grained models in CHARMM

- Case studies
  - Coarse-grained DNA models for sequence and salt effects on DNA melting
    - General coarse-graining of DNA to 3 “beads” per nucleotide (base, sugar, phosphate)
    - Developed by J. de Pablo and coworkers (Chem. Eng., U. Wisc.)
    - Helpful in understanding/informing thermodynamics of DNA melting

Essential Go Model Reductionism



All atom contacts
are replaced by



Cα-Cα contacts

Essential Go Model Reductionism

- Native contact interactions encoded as
  - 1/0 (traditional Go model)
  - ε_{ij}/0 (scaled by empirical energy scale - flavored Go model)
  - All other pairs are repulsive
- Chain connectivity given by bonds, angles and dihedrals
  - Bonds and angle terms described by harmonic restoring forces centered at psuedo bond and pseudo angle separations from known structure
  - Torsions are treated either as
    - Simple cosine term centered at observed torsion (templated)
    - Information-based cosine series depending on pair of aa

Essential Go Model Reductionism

- Relevant references for Go-type models
  - Conventional Go models
    - JE Shea, YD Nochomovitz, Z Guo and CL Brooks, III. Exploring the space of protein folding Hamiltonians: The balance of forces in a minimalist β-barrel model. *J Chem Phys*, 1998, 109, 2895-903
    - JE Shea, JN Onuchic and CL Brooks, III. Exploring the origins of topological frustration: design of a minimally frustrated model of fragment B of protein A. *PNAS*, 1999, 96, 12512-7
    - C Clementi, H Nymeyer, JN Onuchic. Topological and energetic factors: What determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J Mol Biol,* 2000, 298, 937-53
    - N Koga and S Takada. Roles of native topology and chain-length scaling in protein folding: a simulation study with a Go-like model. *J Mol Biol* 2001, 313, 171-80
    - MS Cheung, AE Garcia, and JN Onuchic. Protein folding mediated by solvation: Water expulsion and formation of the hydrophobic core occur after the structural collapse. *PNAS,* 2002, 99, 685-90

Essential Go Model Reductionism

- Relevant references for Go-type models
  - Flavored Go models
    - J Karanicolas and CL Brooks, III. The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci*, 2002, 11, 2351-61
    - J Karanicolas and CL Brooks, III Improved Go-like Models Demonstrate the Robustness of Protein Folding Mechanisms Towards Non-native Interactions. *J Mol Biol*, 2003, 334, 309-25

Representing Go models in CHARMM

- Specifying topology and parameters

Sequence/mass information

```
read rtf card
* Topology for Go model of 1bdc
*
  20   1
MASS 1   G1      101.000000
MASS 2   G2      71.000000
MASS 3   G3      114.000000
MASS 4   G4      114.000000
MASS 5   G5      128.000000
MASS 6   G6      147.000000
MASS 7   G7      114.000000
MASS 8   G8      128.000000
MASS 9   G9      128.000000
MASS 10  G10     128.000000
MASS 11  G11     128.000000
MASS 12  G12     114.000000
MASS 13  G13     71.000000
.
.
.
MASS 57  G57     71.000000
MASS 58  G58     97.000000
MASS 59  G59     128.000000
MASS 60  G60     71.000000
```

Residue information

```
DECL +CA

AUTOGENERATE ANGLES
DIHEDRAL

RESI G1        0.0
GROU
Atom  CA  G1      0.0
Bond CA +CA
```

Representing Go models in CHARMM

- Specifying topology and parameters

Bonds and angles

```
read param card
* Parameters for Go model of 1bdc
*

BOND
G1    G2      378.000000  3.795046
G2    G3      378.000000  3.808982
G3    G4      378.000000  3.800045
G4    G5      378.000000  3.791182

ANGLE
G1    G2    G3      75.600000  108.672972
G2    G3    G4      75.600000  112.756549
G3    G4    G5      75.600000  124.755262
G4    G5    G6      75.600000  110.565786
```

Torsions

```
DIHEDRAL
G1   G2   G3   G4    0.382494  1  284.943180
G1   G2   G3   G4    1.026981  2  266.456266
G1   G2   G3   G4    0.017622  3  114.131745
G1   G2   G3   G4    0.195028  4  107.766228
G2   G3   G4   G5    0.434771  1  296.199841
G2   G3   G4   G5    0.524659  2  253.486984
G2   G3   G4   G5    0.108980  3  25.409709
G2   G3   G4   G5    0.056961  4  96.428204

NONBONDED NBXMOD 3 ATOM CDIEL SHIFT VATOM -
  VDISTANCE VSWITCH -
  CUTNB 399.0 CTOFNB 398.5 CTONNB 395.5 EPS 1.0 WMIN 1.5

G1      0.0  -0.000132  4.037732
G2      0.0  -0.000132  5.474578
G3      0.0  -0.000132  6.595057
```

Non-specific non-bonded repulsion

Representing Go models in CHARMM

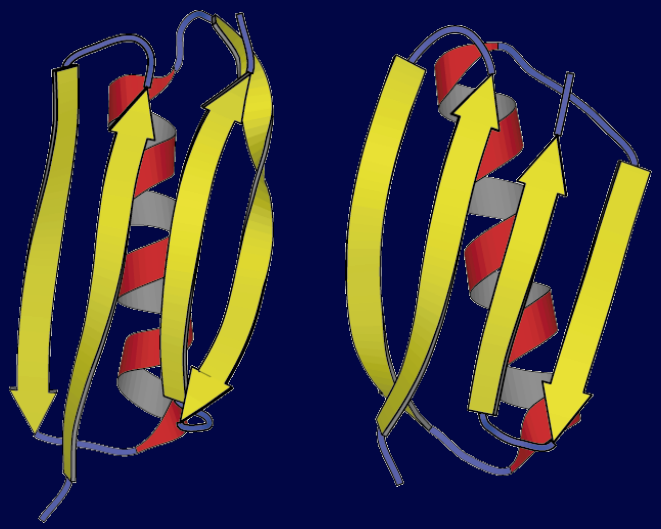
- Specifying topology and parameters
  - Residue pair specific (native contact) non-bonded parameters

```
NBFIX
G1     G4      -0.043567   6.871368
G1     G7      -0.043567   8.971603
G2     G39      -0.047043   14.179823
G2     G40      -0.046579   15.310104
G3     G6      -0.080644   9.319967
G3     G40      -0.037773   12.423546
```

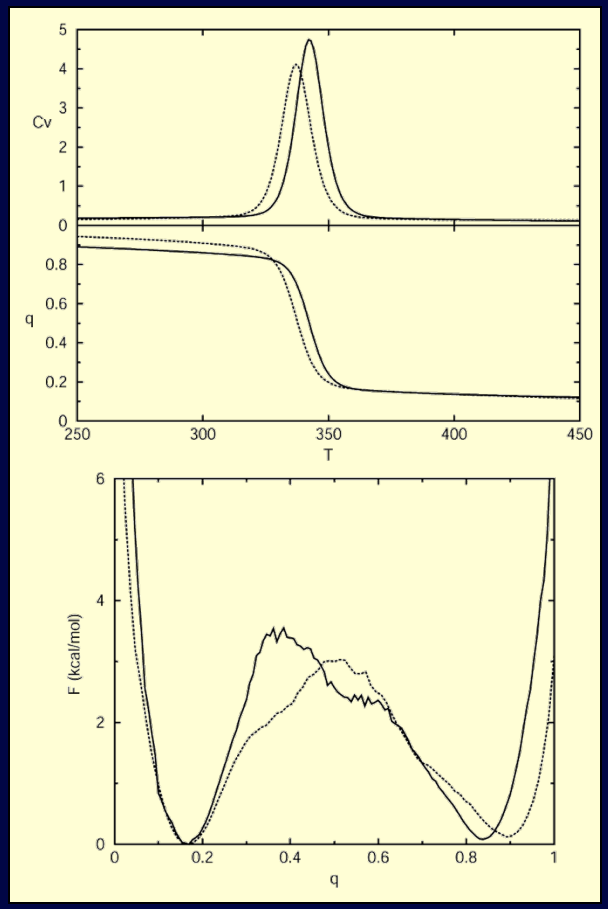
Can we understand different mechanisms of folding in similar topologies?

Sequence specific Go-like models yield two-state like folding for both proteins

LB1 GB1



Similar heat capacities, cooperativity and folding free energy surfaces (versus q, fraction of native contacts)



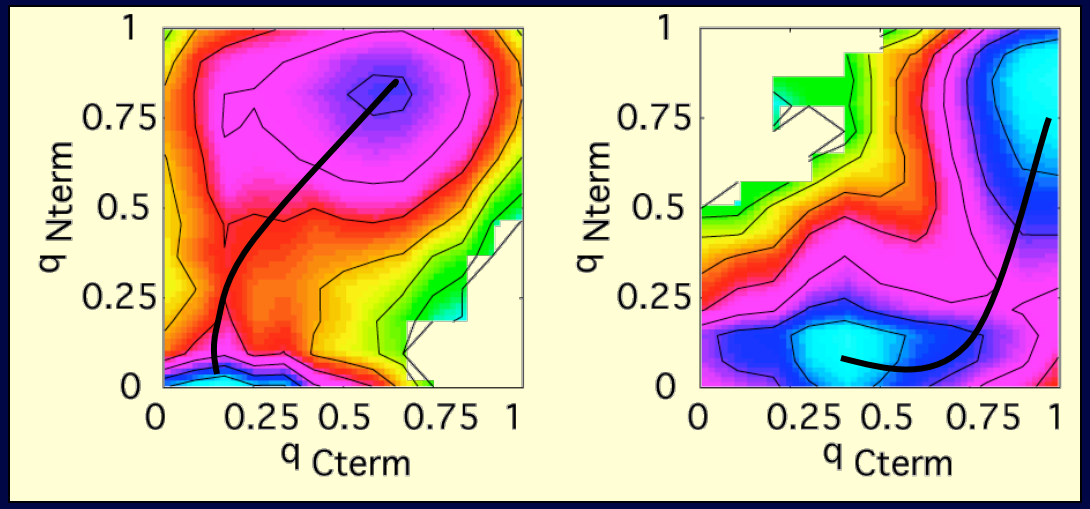
Segment B1 of peptostreptococcal protein L (LB1) and segment B1 of streptococcal protein G (GB1) have very similar topologies but different folding mechanisms

Karanicolas & Brooks, *Prot. Sci.*, 2002

Different sequences, analogous topology, yield different folding mechanism

Consistent with experimental findings

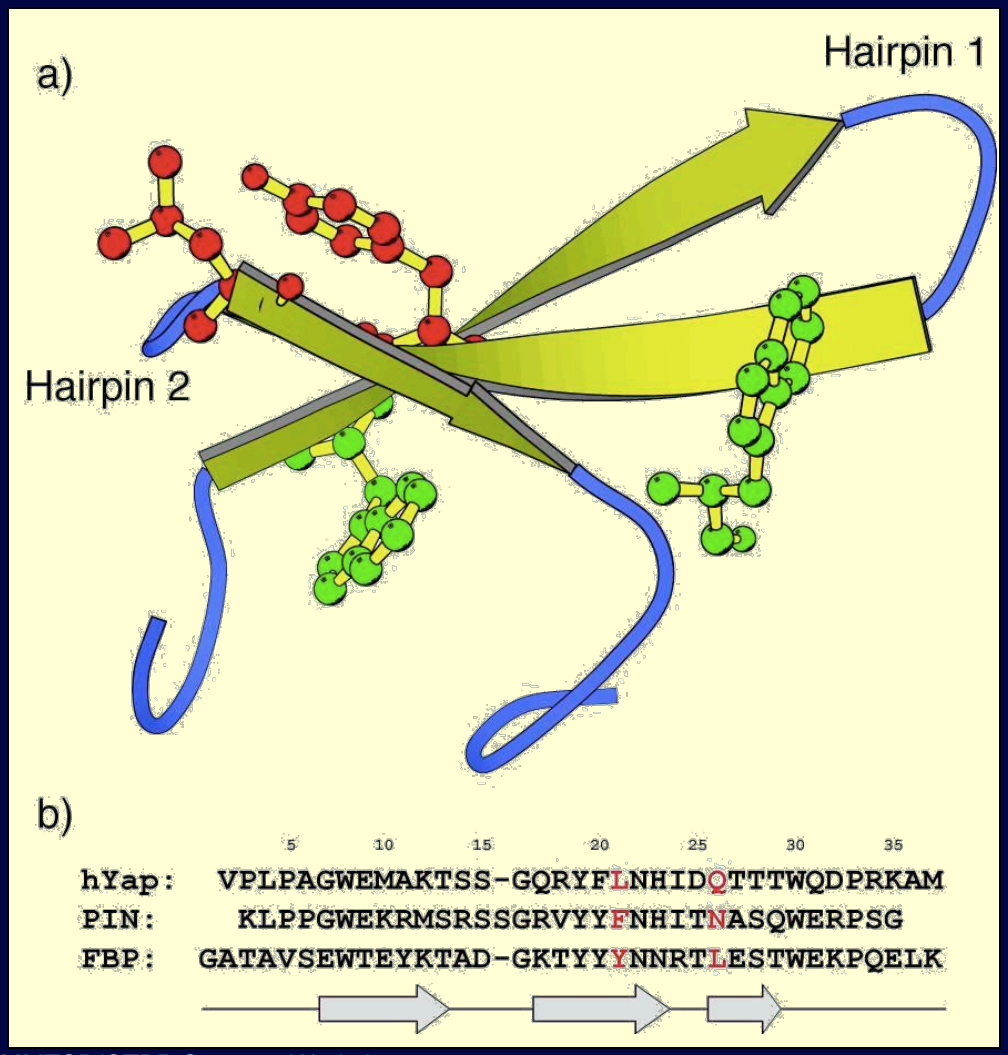
LB1 GB1



- For LB1 the N-terminal hairpin precedes folding of C-terminal hairpin
- In GB1 (as already seen from all-atom calculations) C-terminal hairpin forms earlier

Karanicolas & Brooks, *Prot. Sci.,* 2002

Kinetics and mechanism of WW domains using Go-like models



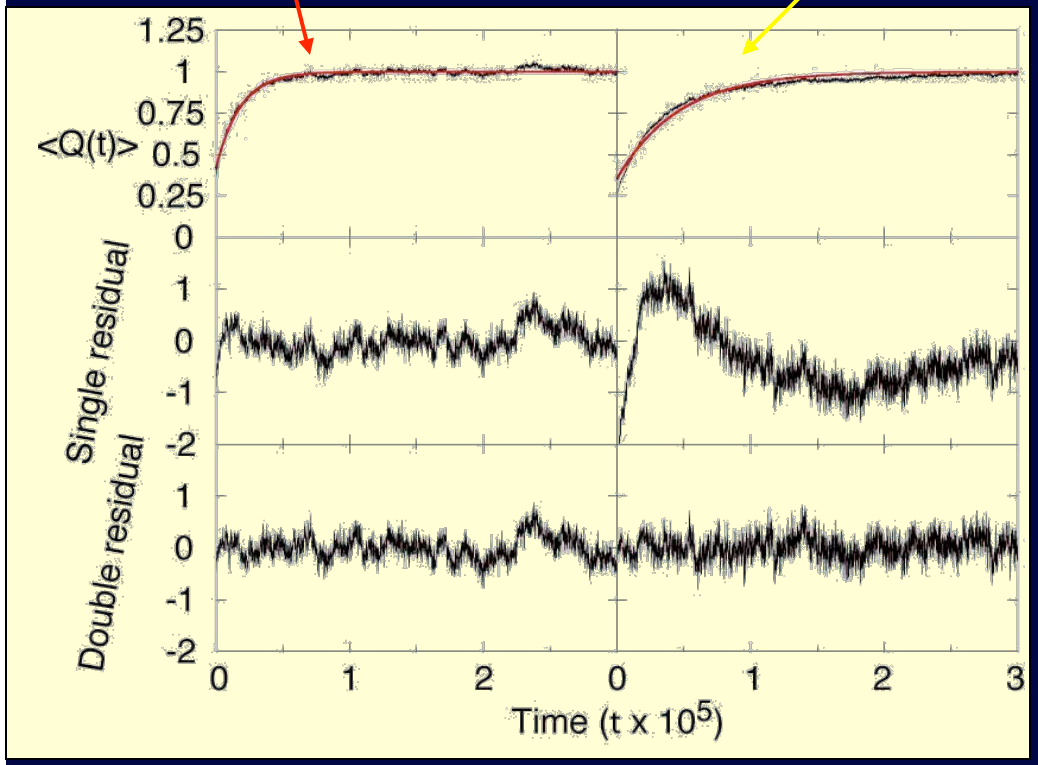
- WW domains are simple β-sheet "proteins" that show a sequence dependent switch between 2-state and 3-state folding kinetics
- What is the folding mechanism?
- What is the origin of the switch?

Karanicolas & Brooks, *PNAS,* 2003

Folding kinetics reproduce experimental observations

PIN WW domain shows 2-state kinetics

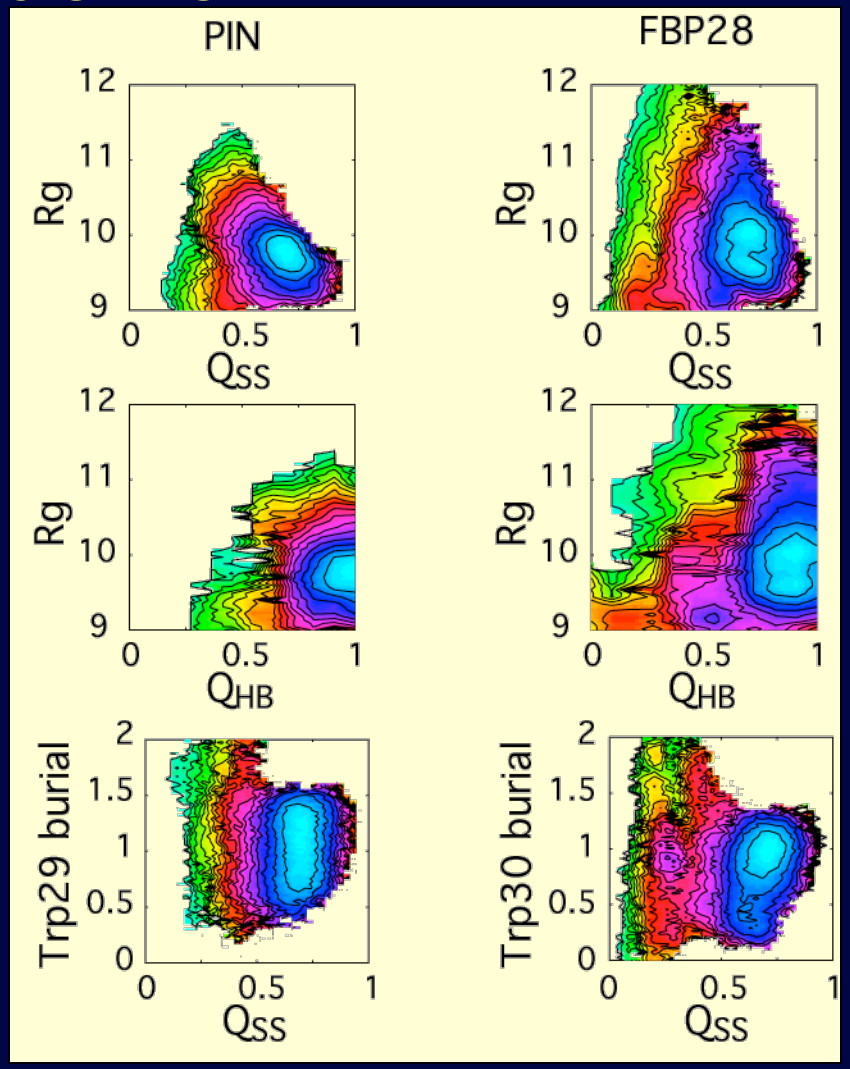
FBP WW domain follows 3-state kinetics



- FBP shows loop 2 folding dominates folding kinetics
- Parallel pathways for formation of loop 1 and loop 2
- Registration of loop 2 is rate determining in FBP

Karanicolas & Brooks, *PNAS,* 2003

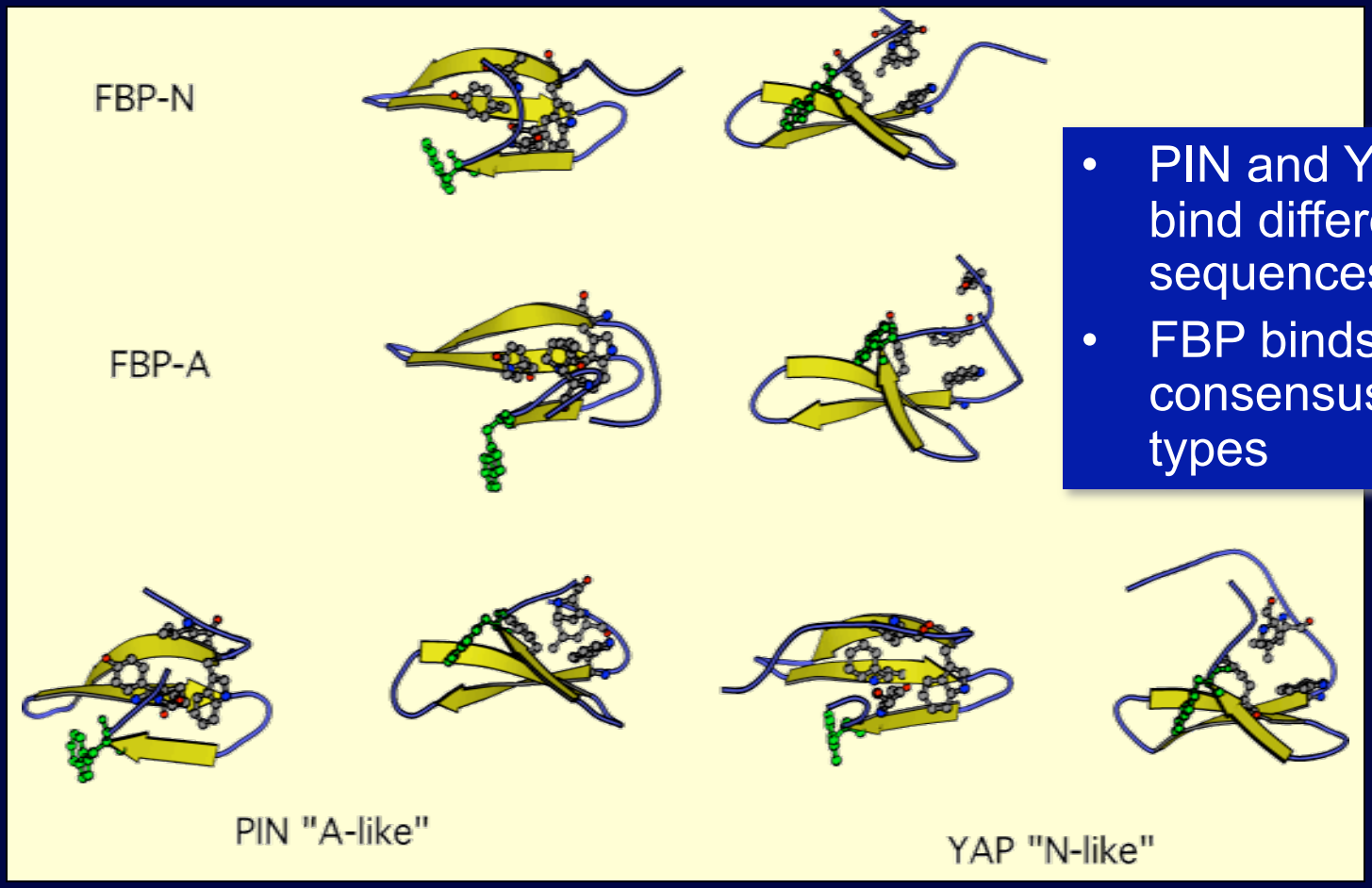
Free energy landscapes indicate presence of intermediate in FBP WW domain



- Free energy landscapes calculated with detailed atomic models show intermediate "shoulder" in FBP WW domain
- Presence of meta-stable state consistent with Go model kinetics

Karanicolas & Brooks, *PNAS,* 2004

Multi-phase folding is a hallmark of functional substates - folding and function cooperate



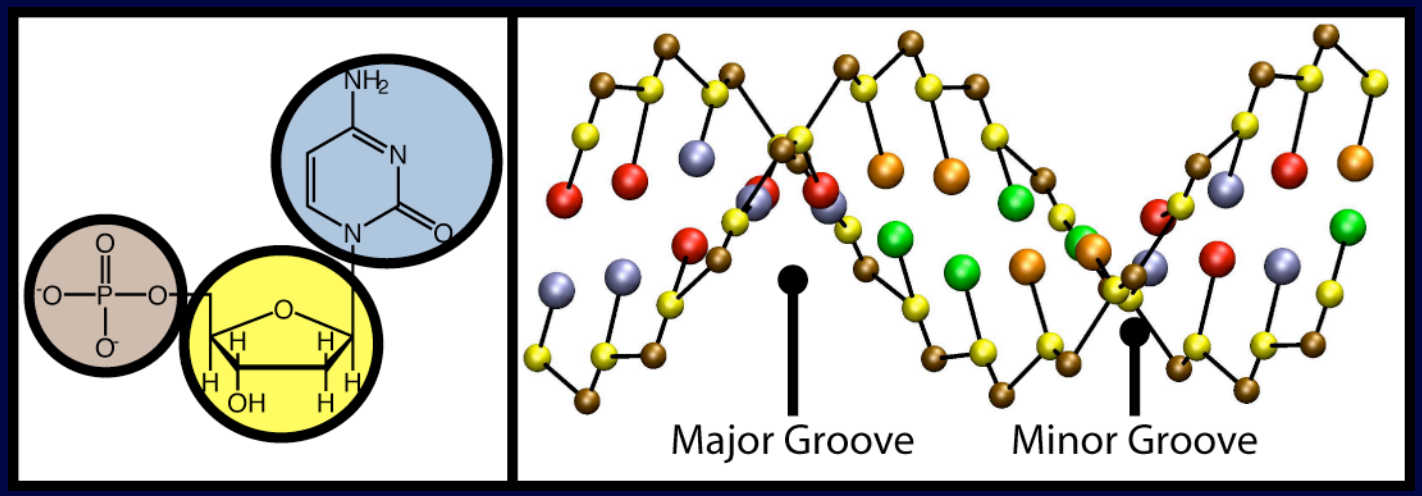
- PIN and YAP domains bind different consensus sequences
- FBP binds two consensus sequence types

Karanicolas & Brooks, *PNAS,* 2004

Coarse-grained DNA model

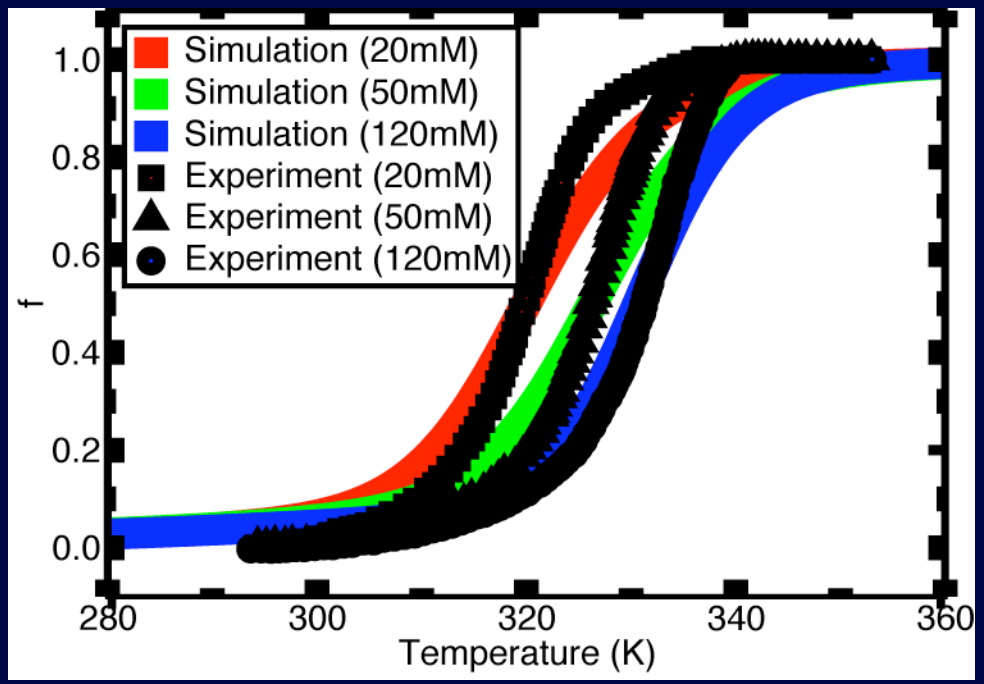
- J de Pablo and coworkers
  - DNA reduced to three beads per nucleotide
  - Bond, angle and torsion potentials as in MM force fields
  - Non-nonded specific for specific base-stacking and pairing
  - Electrostatics via screened coulomb law

$$u_{elec}(r) = \frac{e^{-\kappa r}}{\varepsilon r}$$



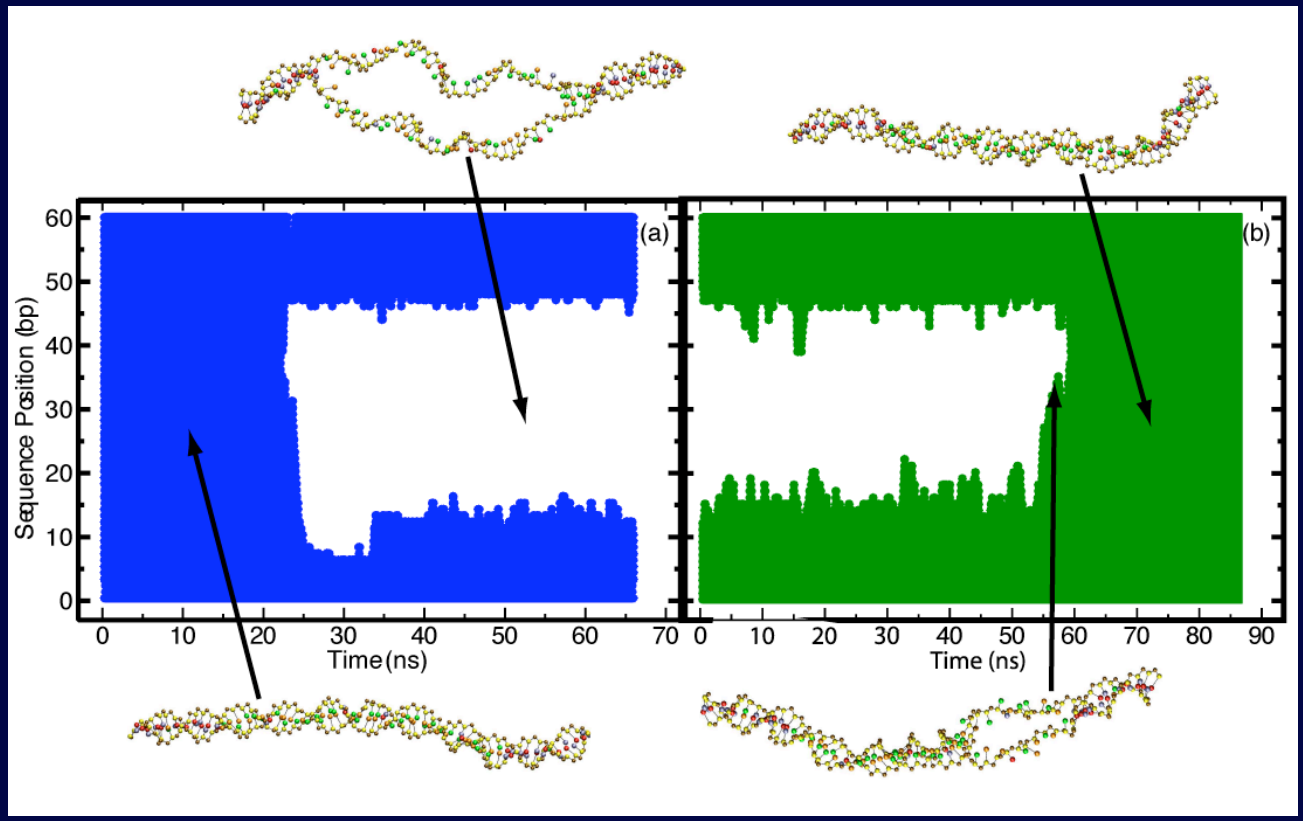
Coarse-grained DNA model

- J de Pablo and coworkers
  - Model reproduces salt-dependent DNA melting

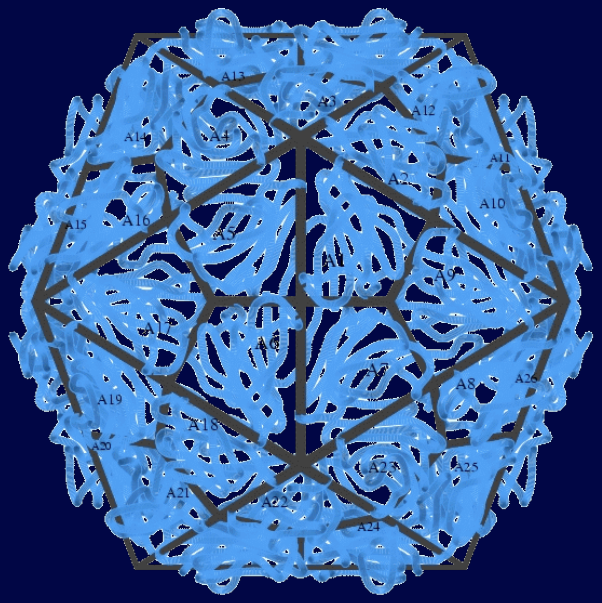


Coarse-grained DNA model

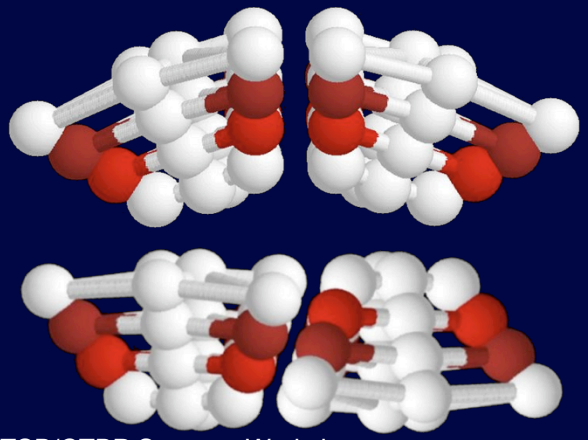
- J de Pablo and coworkers
  - Model reproduces salt-dependent DNA melting

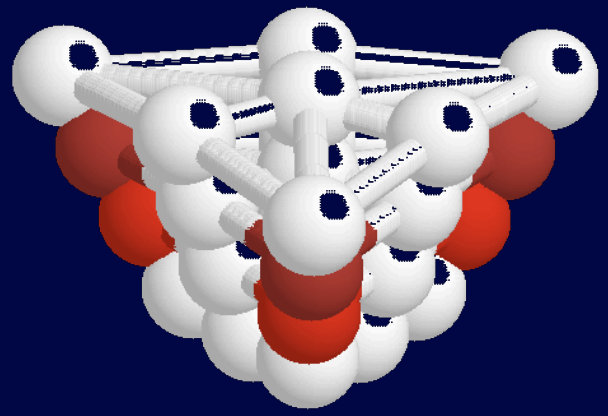


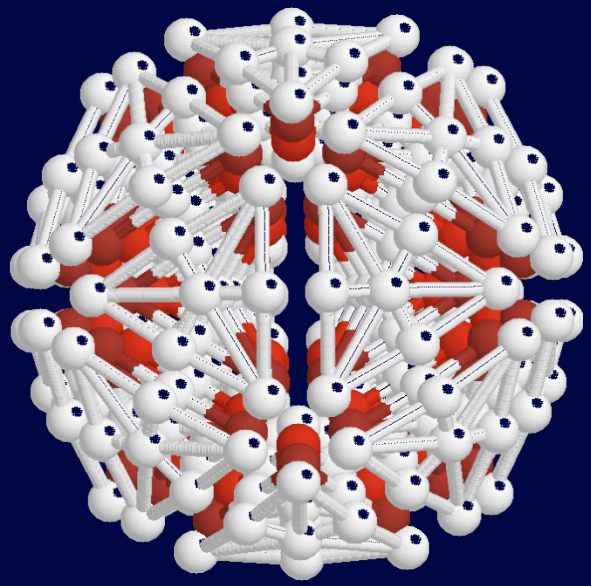
Coarse-grained model for virus assembly



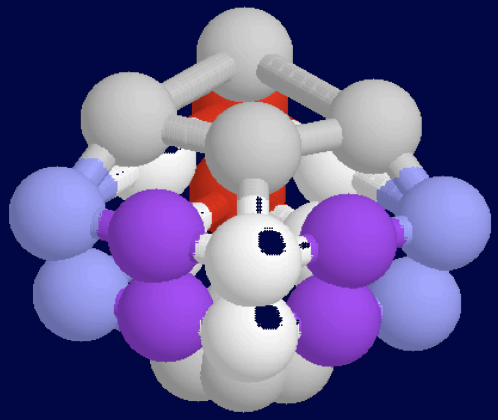
Native and non-native associations possible

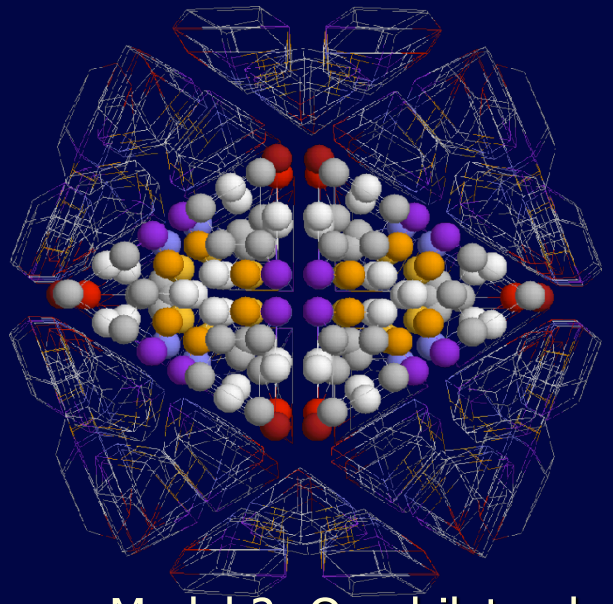






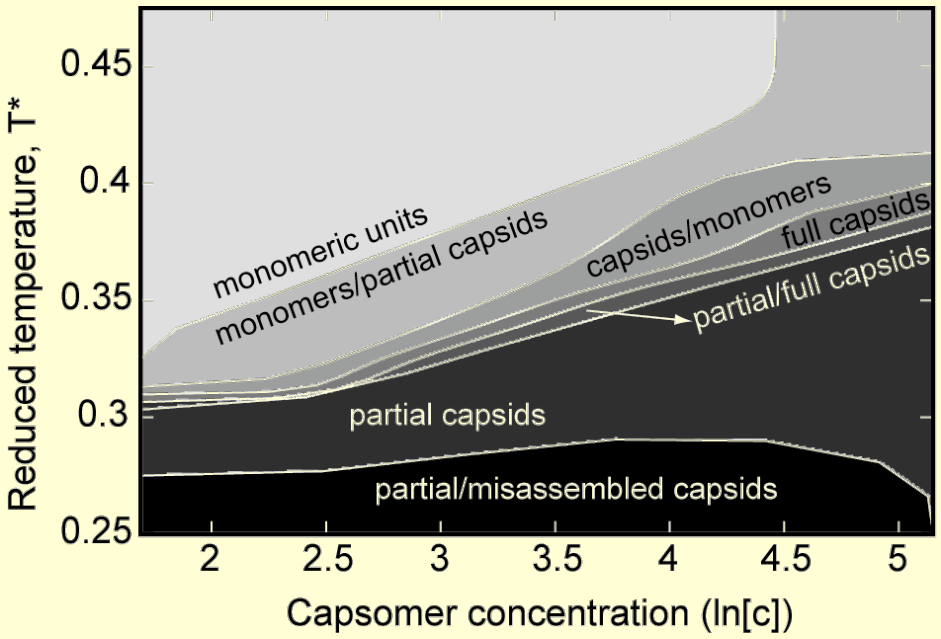
Model 1: Triangular Capsomers, $T=1_{20}$

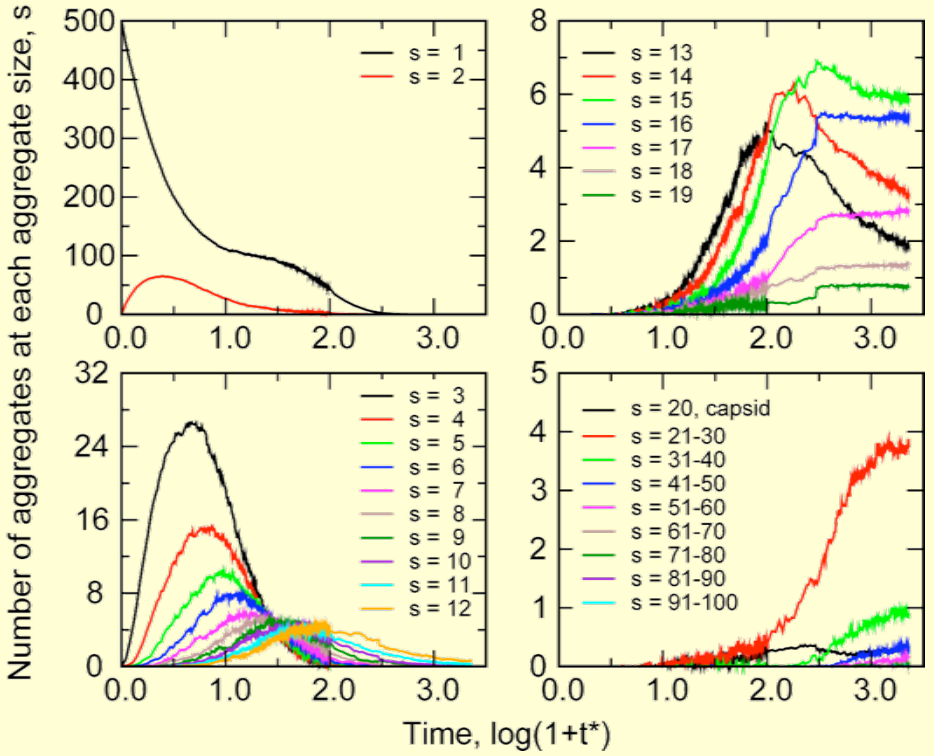




Model 2: Quadrilateral Units, $T=1_{60}$

Probing viral assembly kinetics and thermodynamics





HD Nguyen, VS Reddy and CL Brooks, III, JACS, 2008.

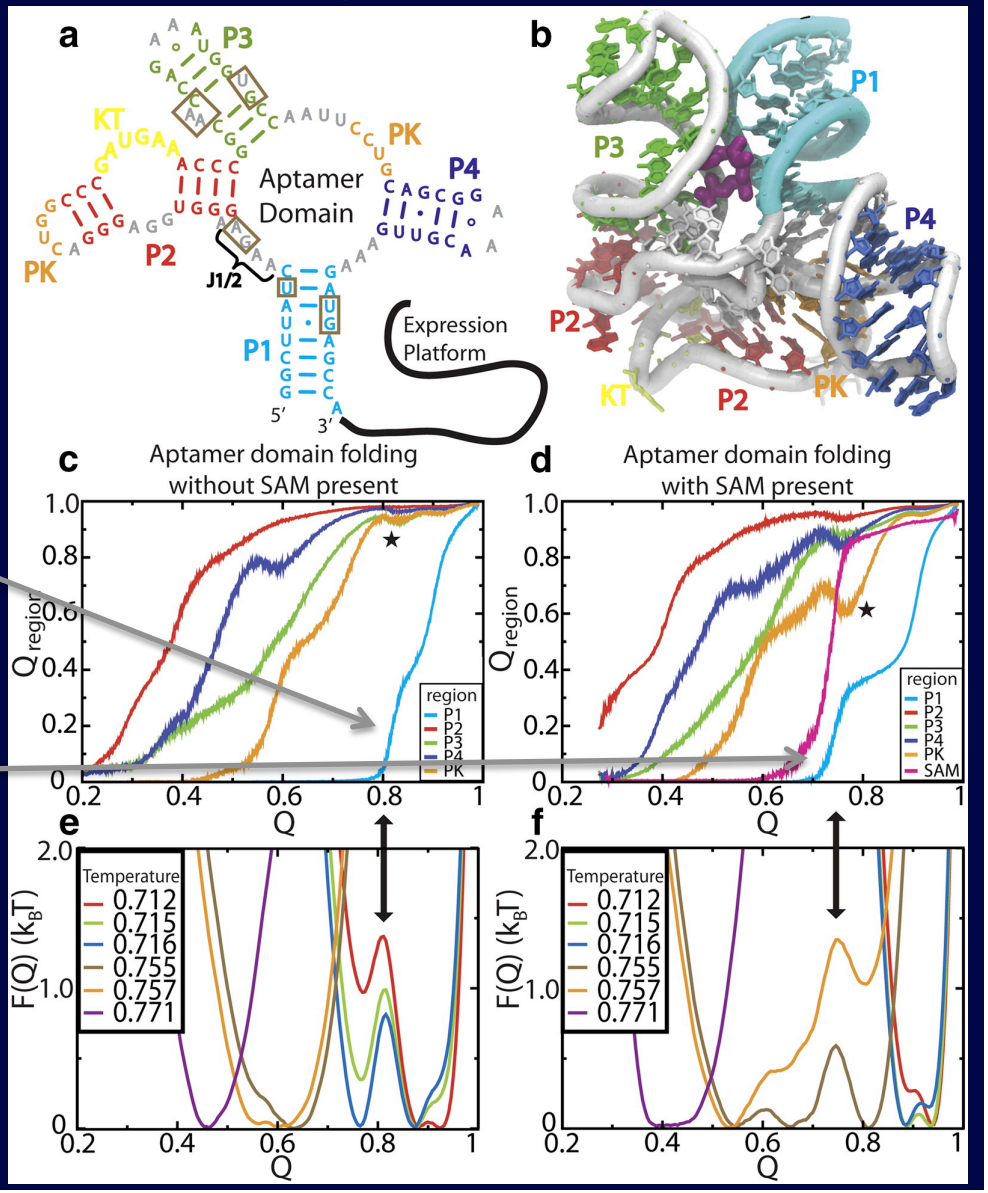
SAM-1 folding thermodynamics and kinetics in competition with ligand binding

- All atom Go-models of SAM-1 riboswitches
  - SAM riboswitches bind S-adenosyl methionine (SAM) to regulate methionine and SAM biosynthesis and transport
  - Go-model server at http://sbm.ucsd.edu/

Whitford et al, Biophys. J., 2009

SAM-1 folding thermodynamics and kinetics in competition with ligand binding

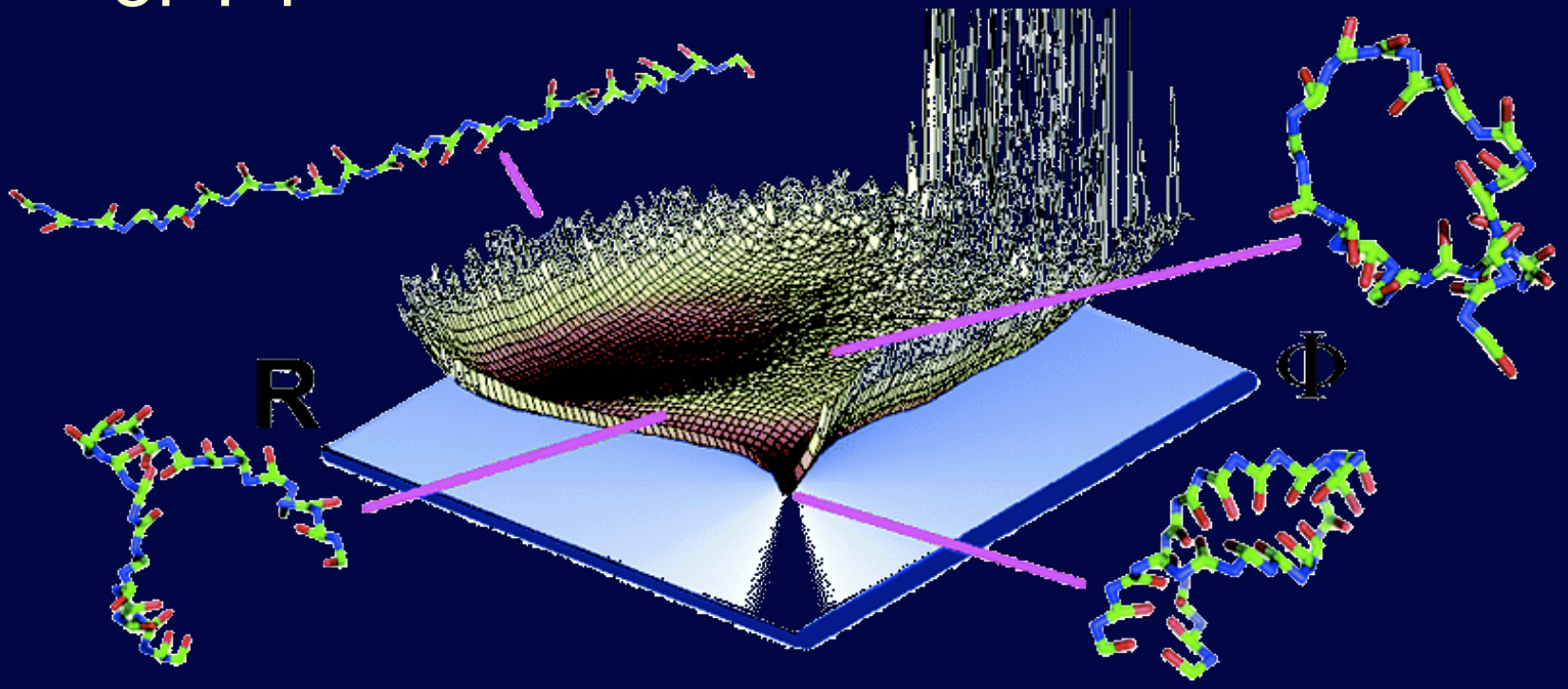
- All atom Go-models of SAM-1 riboswitches
- P1 helix formation is rate limiting
- Presence of substrate SAM assists P1 helix folding in aptamer domain



Whitford et al, Biophys. J., 2009

Force induced folding/unfolding of an RNA hairpin probes rugged energy landscape

- Three-site structure based models used to explore hairpin folding as a function of T-f



Hyeon & Thirmualai, JACS, 2008