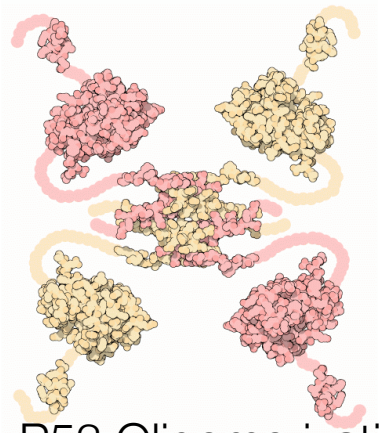
Introduction to Molecular Dynamics

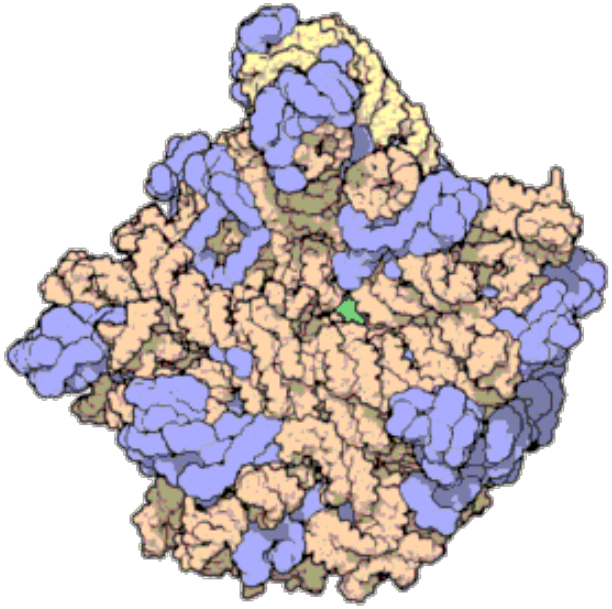
Vijay Pande

OpenMM Workshop, February 13, 2009

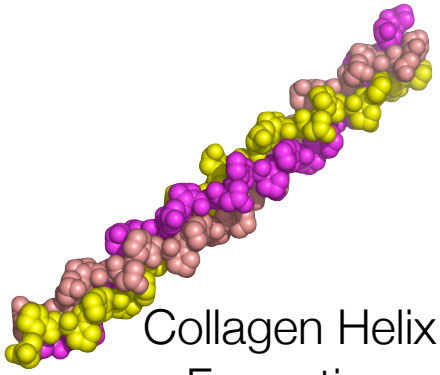
Crystallography gives structures, but …



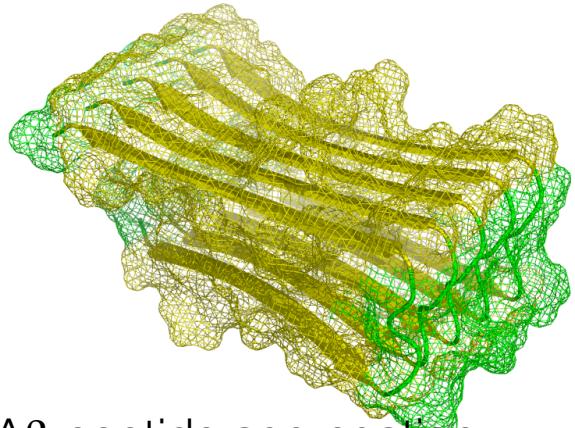
P53 Oligomerization
(50% of cancers)



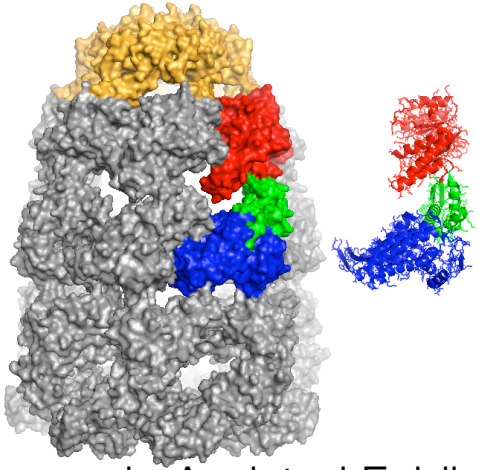
Ribosome:
(Last step of
Central Dogma,
Antibiotic resistance)



Collagen Helix
Formation
(*Osteogenesis Imperfecta*)



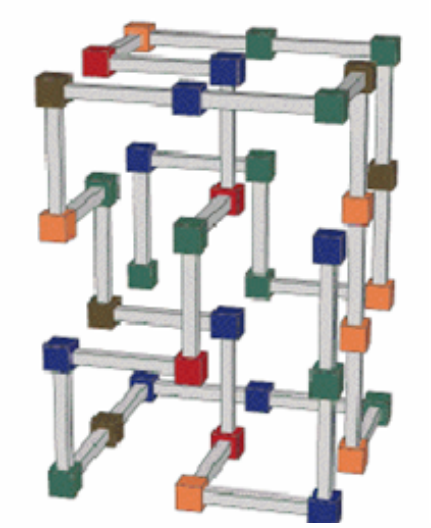
Aβ peptide aggregation
(Alzheimer's Disease)

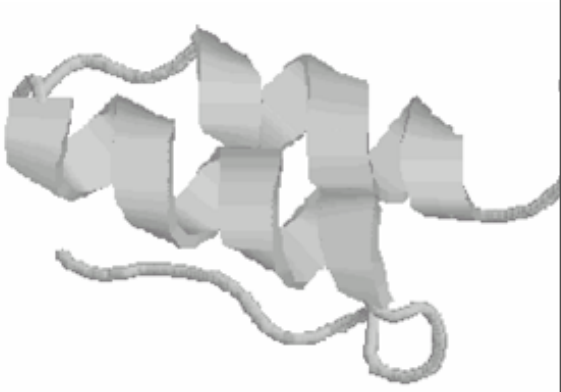


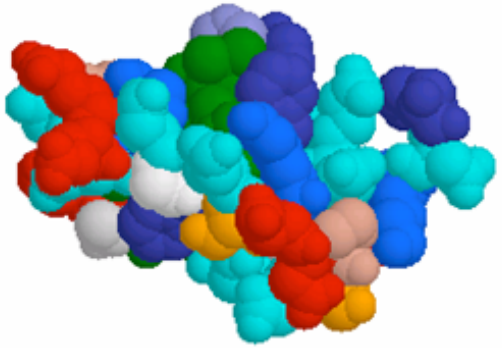
Chaperonin Assisted Folding
(relevant to cancer: HSP90 inhibitors)

Outline

- Philosophy
- Challenges
- Models
- Sampling
- Examples







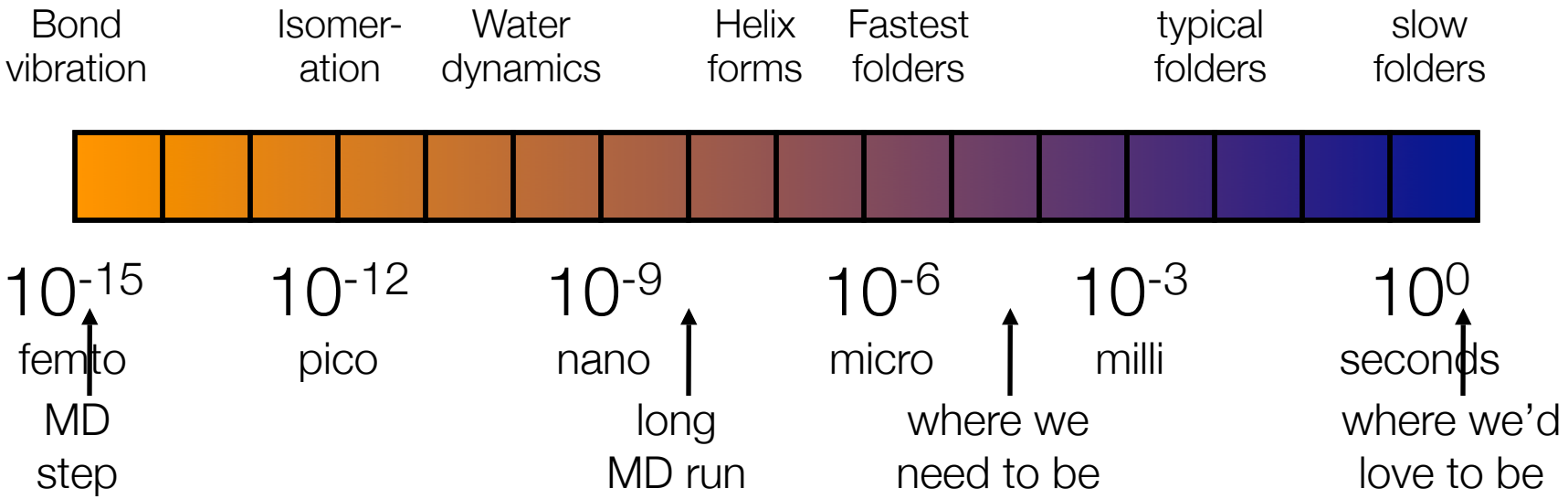
Challenges of Molecular Simulation

Models *vs* **Sampling**

Are our models sufficiently accurate to answer the questions we're asking?

Have we reached the appropriate equilibrium conditions?

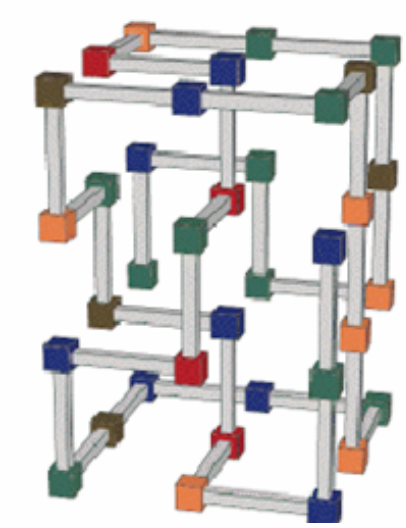
Timescales to sample

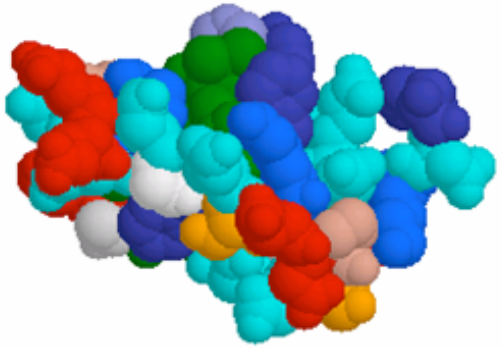


- **16 order of magnitude range**
  - Femtosecond timesteps
  - Need to simulate micro to milliseconds

Outline

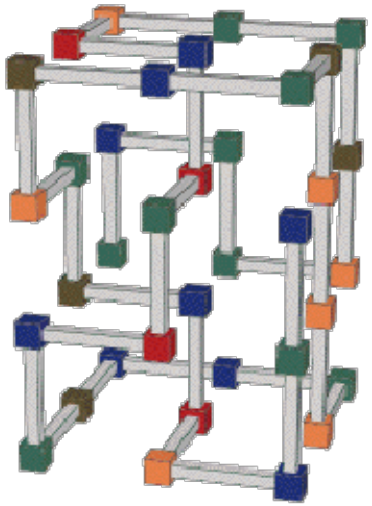
- Philosophy
- Challenges
- **Models**
- Sampling
- Examples

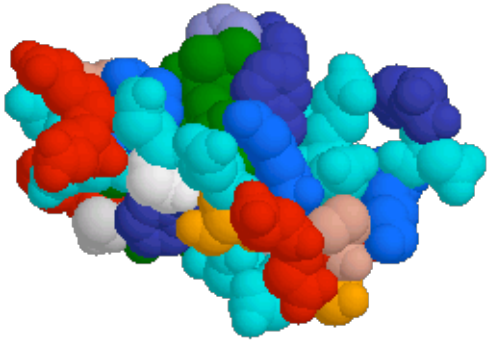




Range of possible models

Great sampling — **Accurate model**





Lattice models: simple & generic

CPU minute

Off-lattice models: simple models of particular proteins

CPU hour

All-atom models: very detailed, typically intractable

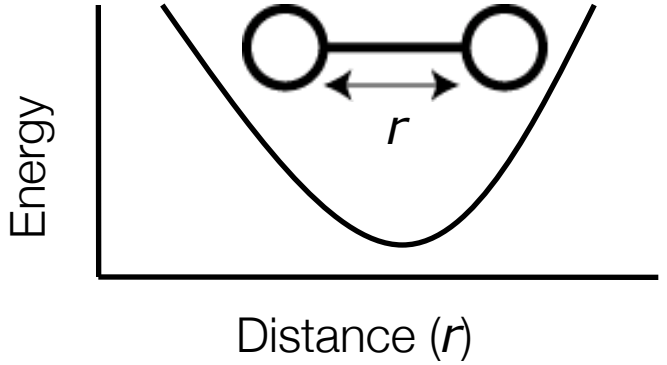
1000 CPU years

Building an atomistic model

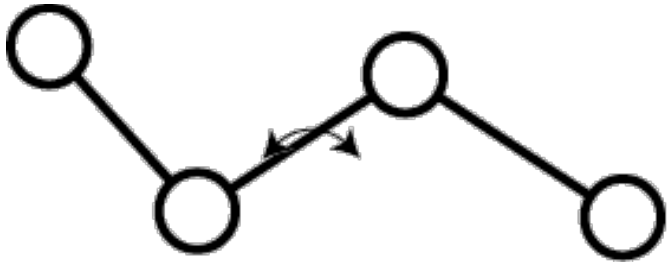
- **What are the important atom-atom forces in biomolecules?**
- **Can we approximate them with classical models**
  - QM would slow the calc down by 1000x
  - A classical approximation should work well in many cases (eg no bond breaking)
- **Can we find the parameters needed in some methodical way**
  - no bias
  - automated procedure

short range interactions

- **Bonds connect atoms**
  - vibrate with a given frequency
  - known bond length
  - approximate energy w/2nd order term
  - connect them by springs



- **Sterics**
  - angles & dihedrals
  - control how atoms bend & move locally

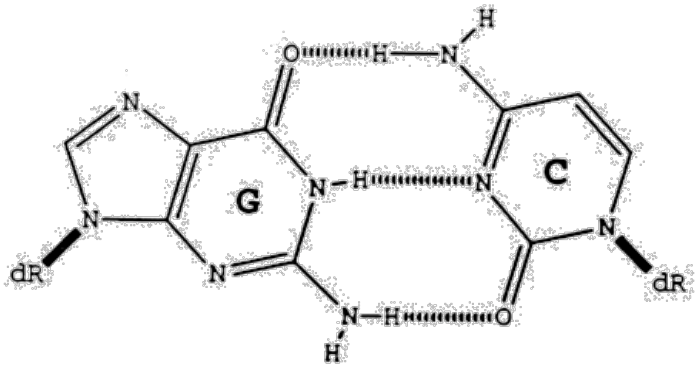


- **van der Waals**
  - dipole-dipole interaction: $-(\sigma/r)^6$
  - Hard core repulsion: modeled as $(\sigma/r)^{12}$
  - Leads to Lenard-Jones: $V_{LJ}(r) = \varepsilon [(\sigma/r)^{12} - (\sigma/r)^6]$

Charge-charge interactions

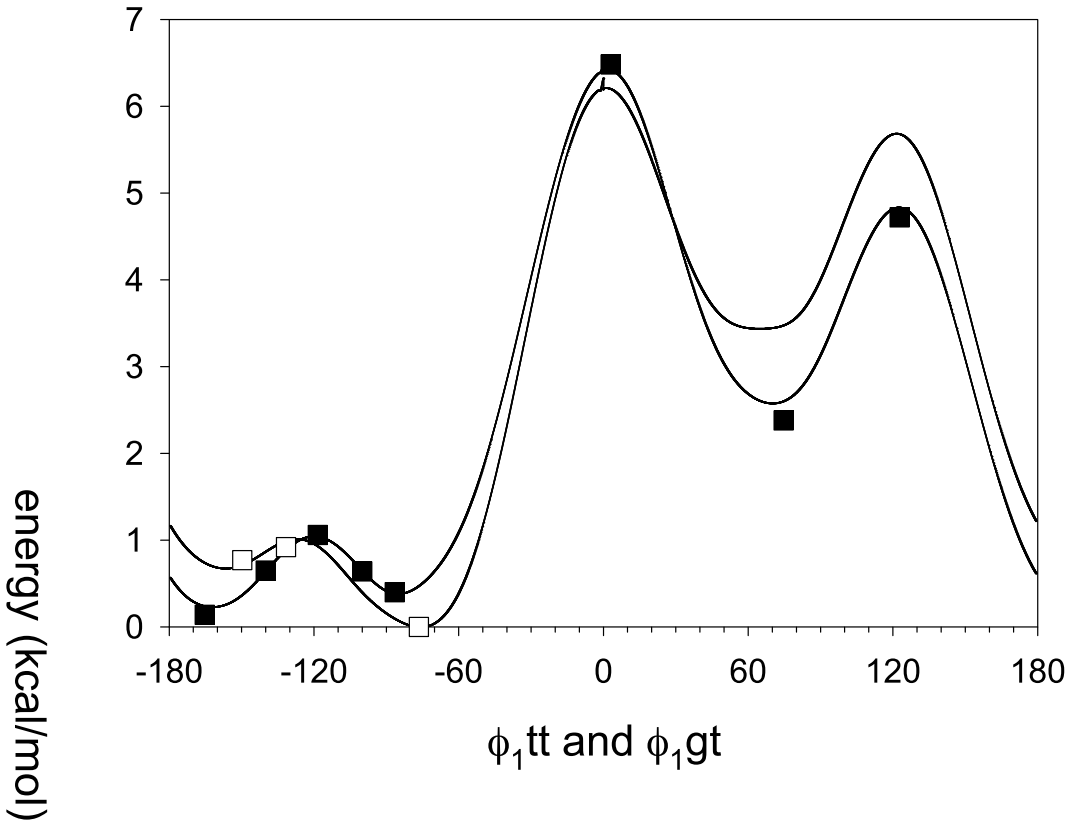
- **Charge-charge interactions: Coulomb's law**

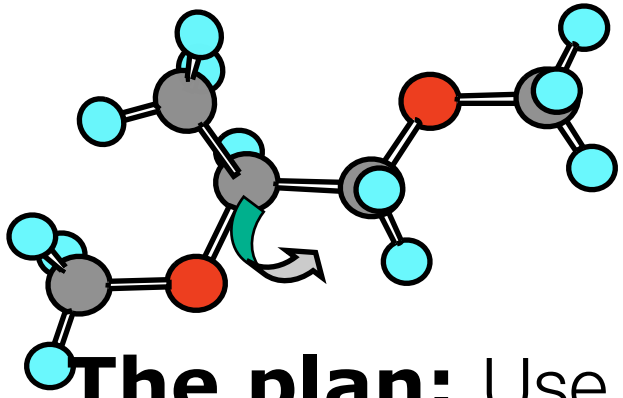
$$U_{el} = \frac{1}{2}\sum_{i \neq j} \frac{q_i q_j}{r_{ij}}$$



- **Physically driven by electrostatics of sorts**
  - NH will be positive
  - CO will be negative
  - hence, attraction
- **In models**
  - handled by partial charges on N,H,C,O
  - partial charges now derived from quantum mechanics
- **Directionality?**
  - partial charges yield a dipole interactions, hence directionality
  - Previously, specific angular functions have been used

How do we get parameters?





The plan: Use quantum mechanics to calculate parameters and then fit to classical potentials

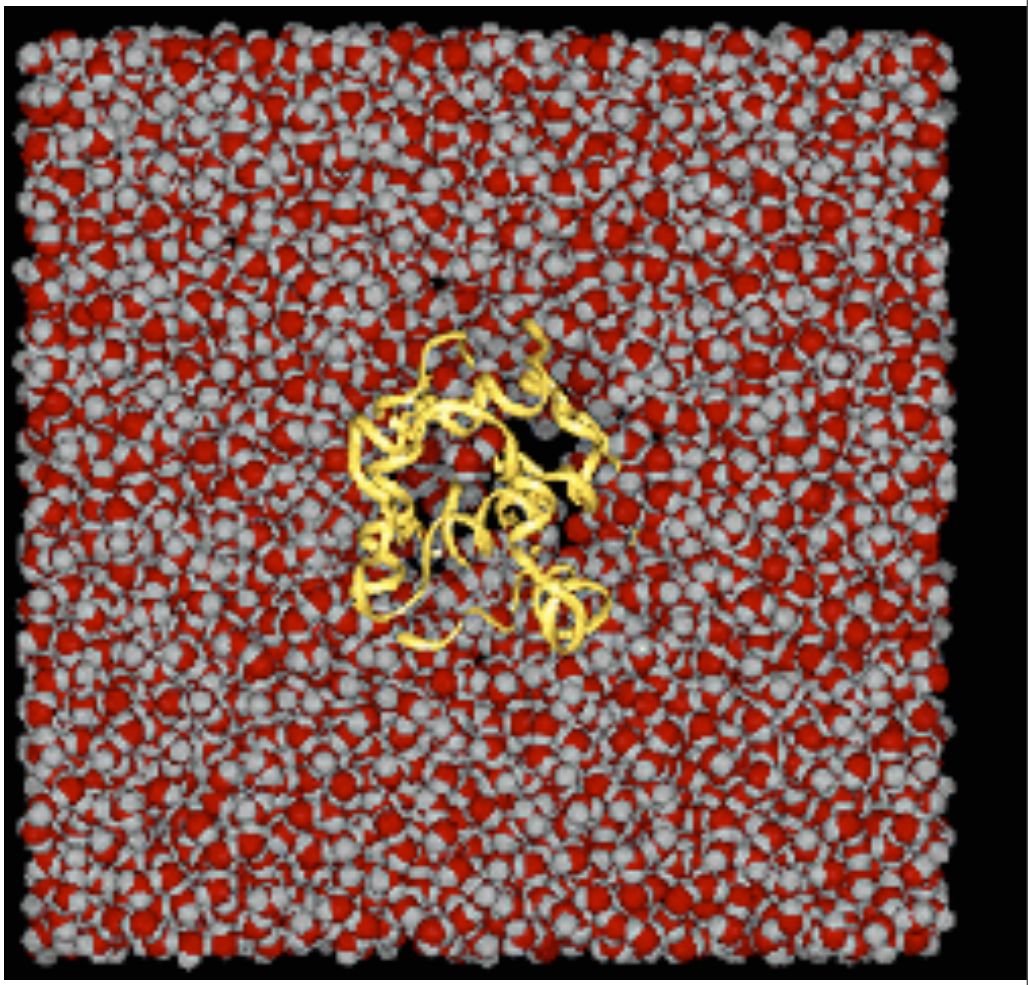
Large number of force fields to choose

- **AMBER**
  - ff94
  - ff96
  - ff99
  - ff99sb: modifications to improve torsions
  - ff03: latest, intended to be balanced
- **OPLS**
  - OPLS-ua (unified atom)
  - OPLS-aa: classic all atom force field
  - OPLS-aa/L: new torsions
- **CHARMM**
  - CHARMM19 (unifed atom)
  - CHARMM27 (latest)
  - CMAP (new torsions)
- **Other**
  - GROMOS (van G.)
  - GROMACS
  - Encad (Levitt)
- **Polarizable force fields**
  - AMOEBA

What about water?

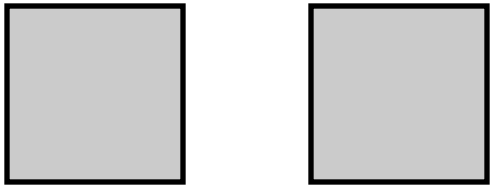
Solvation models

- **Water is very important**
  - Creates the hydrophobic effect
  - Hydrogen bonding to water
- **Explicit water**
  - Water modeled atomistically
  - TIP3P, SPC, etc
- **Implicit water**
  - Water modeled mathematically
  - PBSA, GBSA

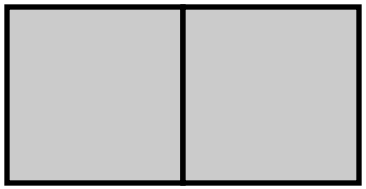


Hydrophobic effect

- **The iceberg model: a simple model**
  - Water forms HB network around hydrophobic solutes
  - This reduces the solvent entropy
  - When two hydrophobic solutes are brought together
    - this reduces the exposed surface area
    - reduces the number of “bound” water
    - increases the entropy, decreasing ΔG



***solute apart**:*
exposed surface area

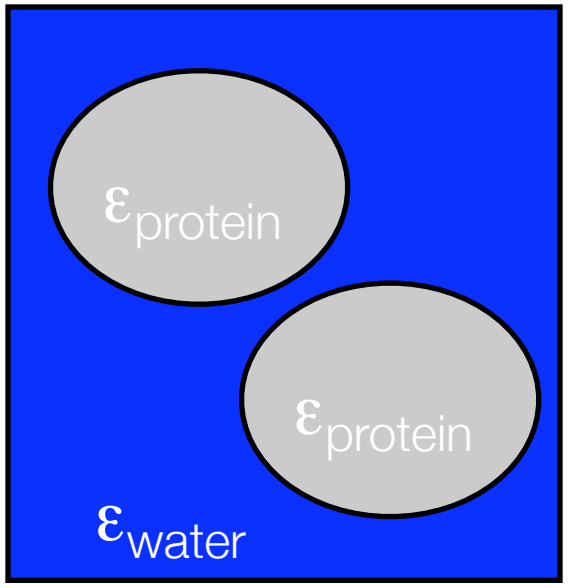


***solute together**:*
less exposed surface area

- **Important for biomolecules**
  - hydrophobic cores of proteins
  - lipid membrane interior vs exterior

Dielectric properties

- **Why dielectric?**
  - proteins have lots of charges
  - charges induce polarization in dielectric media
  - water and the protein can act as a dielectric medium



- **Importance**
  - a great deal of the solvation free energy can come from dielectric properties
  - especially for charged amino acids

$$\varepsilon_{water} \sim 80$$

$$\varepsilon_{protein} \sim 4$$

$$\varepsilon_{vacuum} = 1$$

Implicit solvation model: PB/SA

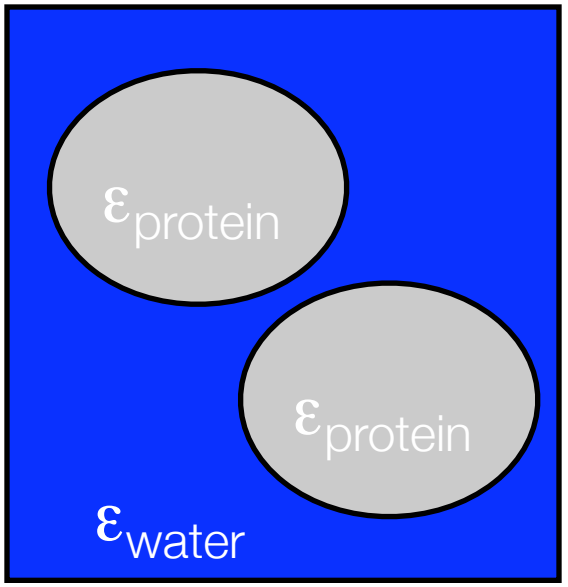
- **Dielectric (Poisson-Boltzmann)**
  - model protein-water system as 2-dielectrics
  - For dielectric $\varepsilon(\mathbf{x})$, electrostatic potential $\phi(\mathbf{x})$, and charge density $\rho(\mathbf{x})$, we get

$$\nabla\varepsilon(\mathbf{x})\,\nabla\phi(\mathbf{x}) = -4\pi\,\rho(\mathbf{x})$$

- **What about counter ions?**
  - Two types of charges $\rho = \rho_{fixed} + \rho_{mobile}$
    - fixed (on the protein)
    - mobile (counter ions)
  - We say the counter ions immediately equilibrate $\rho_{mobile} = \exp(-c\phi/kT)$
  - We get the Poisson-Boltzmann equation

$$\nabla\varepsilon\,\nabla\phi = -4\pi[\rho_{fixed} + \exp(-c\phi/kT)]$$

- **Generalized Born (GB) is an approximation to PB**



$$\varepsilon_{water} \sim 80$$

$$\varepsilon_{protein} \sim 4$$

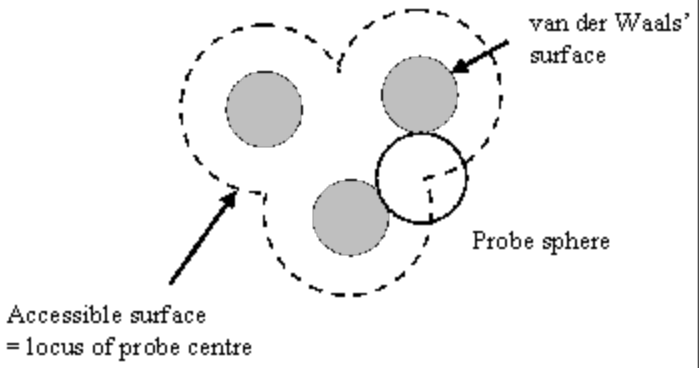
$$\varepsilon_{vacuum} = 1$$

Implicit solvation model: PB/SA

- **Hydrophobicity (surface area)**
  - We make the approximation that hydrophobicity is related to buried surface area
  - The more buried area, the better

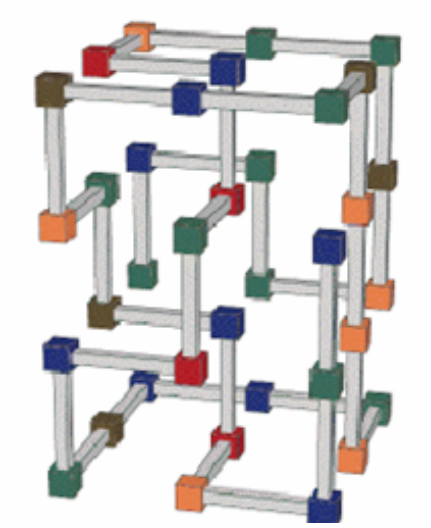
- **Surface area terms as an effective energy**
  - add $H_{SA} = \Sigma_i \sigma_i A_i$ to energy
    - A_i is the surface area
    - σ_i is the coefficient, related to hydrophobicity scale
  - in the end, we need A_i to correlate with solvation free energies more than a geometrical calculation of area

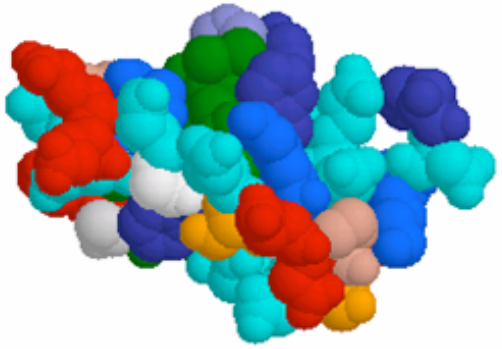
- **How to parameterize PB/SA?**
  - Compare to solvation free energies of small molecules

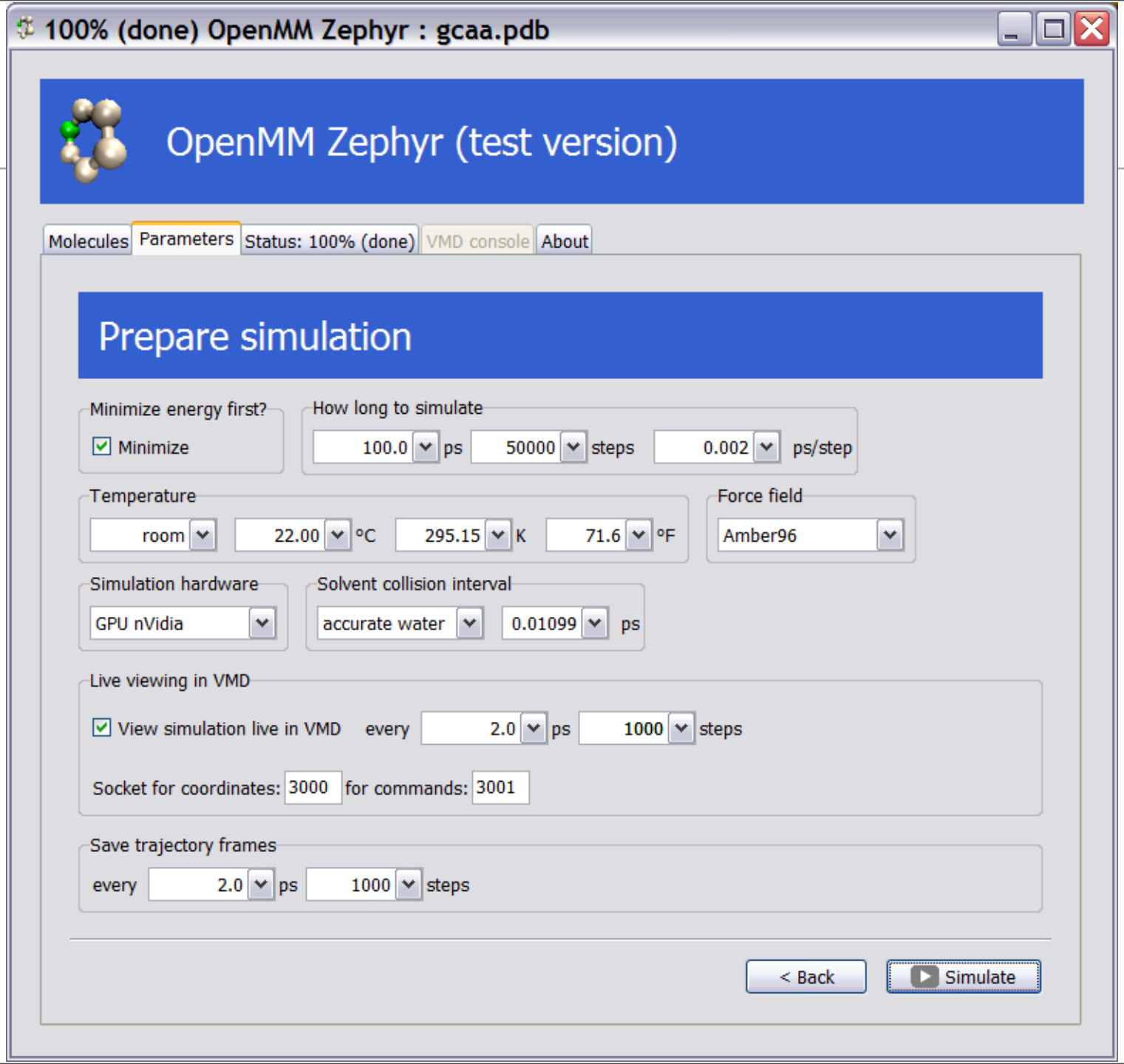


Outline

- Philosophy
- Challenges
- Models
- **Sampling**
- Examples






100% (done) OpenMM Zephyr : gcaa.pdb
OpenMM Zephyr (test version)
Molecules
Parameters
Status: 100% (done)
VMD console
About
Prepare simulation
Minimize energy first?
Minimize
How long to simulate
100.0 ps
50000 steps
0.002 ps/step
Temperature
room
22.00 °C
295.15 K
71.6 °F
Force field
Amber96
Simulation hardware
GPU nVidia
Solvent collision interval
accurate water
0.01099 ps
Live viewing in VMD
View simulation live in VMD
every
2.0 ps
1000 steps
Socket for coordinates: 3000
for commands: 3001
Save trajectory frames
every
2.0 ps
1000 steps
< Back
Simulate

Kinetics: How to simulate Molecular Dynamics

- **Integrate equations of motion**
  - $\mathbf{F} = m\,\mathbf{a} = \mathbf{F}_{ext} - \gamma\,\mathbf{v}$
    $= \mathbf{F}_{ext} - \gamma\, d\mathbf{x}/dt$
    $d\mathbf{x} = \mathbf{F}_{ext}\, dt/\gamma$
  - Choose dt to match timescale (typically dt ~ 1 femtosecond)
- **Reproducing "true" dynamics**
  - simulating the motion of all the atoms
  - Useful for kinetics
  - Given sufficient sampling, MD yields correct thermodynamics (states are Boltzmann weighted)

PUT BILLIARDS PLOT

Integrating Newton's equations

- **Leapfrog verlet**
  - Velocities

$$\begin{aligned} \boldsymbol{v}(t+\frac{\Delta t}{2}) &= \boldsymbol{v}(t-\frac{\Delta t}{2})+\frac{\boldsymbol{F}(t)}{m}\Delta t \\ \boldsymbol{r}(t+\Delta t) &= \boldsymbol{r}(t)+\boldsymbol{v}(t+\frac{\Delta t}{2})\Delta t \end{aligned}$$

  - Positions

$$\boldsymbol{r}(t+\Delta t) = 2\boldsymbol{r}(t)-\boldsymbol{r}(t-\Delta t)+\frac{\boldsymbol{F}(t)}{m}\Delta t^2+O(\Delta t^4)$$

- **Langevin dynamics**
  - add a random force

$$m_i\frac{\mathrm{d}^2\boldsymbol{r}_i}{\mathrm{d}t^2} = -m_i\xi_i\frac{\mathrm{d}\boldsymbol{r}_i}{\mathrm{d}t}+\boldsymbol{F}_i(\boldsymbol{r})+\mathring{\boldsymbol{r}}_i$$

  - random force obeys certain properties based on the temperature (eg its variance is $2m_i\xi_i k_B T$)

Significance of viscosity: γ

- **Physical interpretation**
  - $1/\gamma$ = timescale for velocity decorrelation
    typical value for water: 90/ps

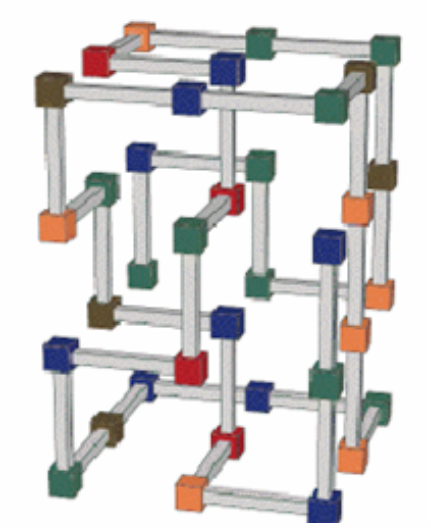
- **Thermodynamics**
  - thermodynamics independent of viscosity
  - for thermodynamics calculations, one can freely use low viscosity if one thinks it will help

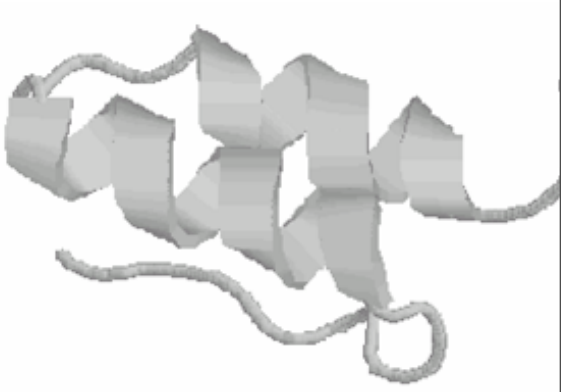
- **Kinetics**
  - Kinetics does depend on viscosity
  - Two regimes:
    - linear regime (10/ps and greater)
    - sqrt regime (10/ps and lower)

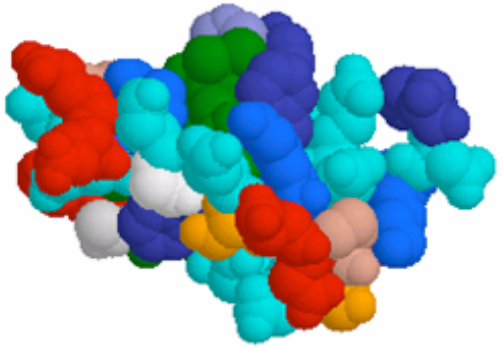
PUT BOJAN RATE vs GAMMA PLOT

Outline

- Philosophy
- Challenges
- Models
- Sampling
- **Examples**



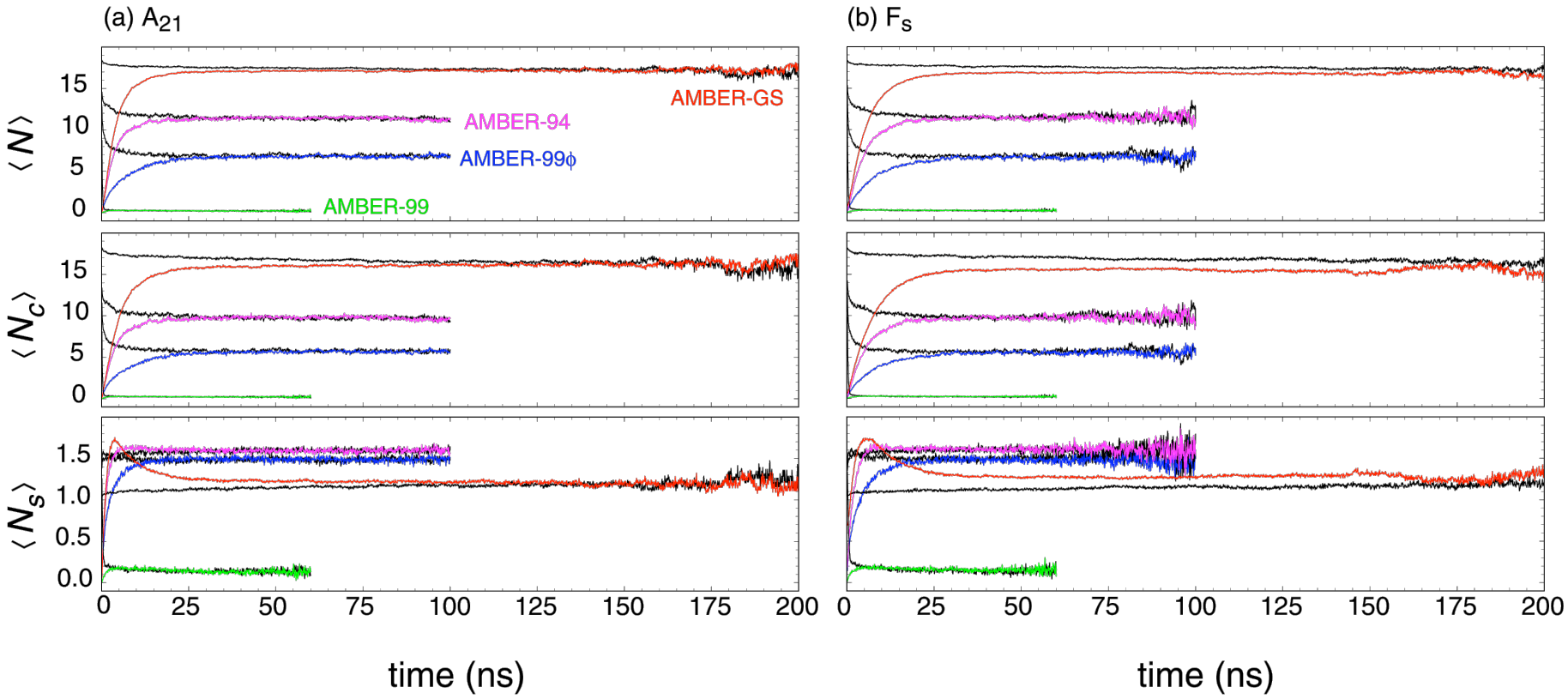




How good are these models?

Test 1: Helix-coil transition

- **many and long MD sims**
  - Thousands of runs for >100 ns, *each*
  - two sets (started folded, started unfolded) for each force field and peptide (A_{21} and Fs)
  - Rates are not strongly dependent on ff, but structure is



Quantitative agreement with experiment

Comparison of 305 K equilibrium ensemble simulation results to experimental values: Kinetics (rates), structure, and thermodynamics (Lifson-Roig parameters)

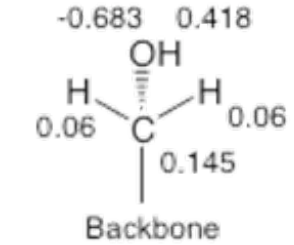
Metric	AMBER-94		AMBER-GS		AMBER-99		AMBER-99ϕ		Exp. (F_s)	Ref.
	A_{21}	F_s	A_{21}	F_s	A_{21}	F_s	A_{21}	F_s		
$v^{(a)}$	0.35	0.36	0.68	0.70	0.06	0.06	0.26	0.26	0.036	61
$w^{(a)}$	1.66	1.67	3.70	3.70	0.70	0.70	1.27	1.26	~1.3	61
$\langle \%\, 3_{10} \rangle_{eq}$	6.4	6.4	0.15	0.04	16.0	16.5	17.8	17.3	~16%	59; 60
$k_{C\alpha H}$ (ns^{-1})	0.15	0.11	0.12	0.11	0.00	0.00	0.06	0.05	0.06	22
$\langle \tau_{coil}$ (ns)$\rangle$	0.21	0.24	0.32	0.38	0.81	0.89	0.26	0.28	0.3	23
$\langle Rg$ (Å)$\rangle_{eq}$	9.32	9.40	9.56	9.55	7.32	7.97	9.02	9.24	~9[(b)]	57
$\langle RMSD$ (Å)$\rangle_{eq}$	3.60	4.00	1.88	2.59	7.85	7.68	5.13	5.31	-	-

(a) Calculated using 30° cutoffs as described in the text

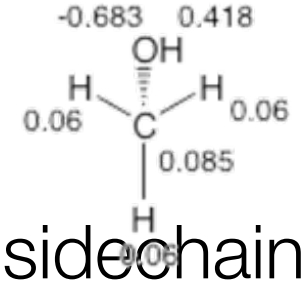
(b) Measured at ~283 K

Test 2: Solvation of Amino Acid Side Chain Analogs

- **Highly sensitive test of solvent-solute interactions**
  - **System:** side chain analogs (eg alanine → methane)
  - **Experiment:** Highly precise experimental data available for comparison (eg, Wolfenden)
  - **Protein Model:** CHARMM27, AMBER(ff94), and OPLS-AA; **Water model:** TIP3P
- **Novel computation aspect**
  - Highly precise results (0.05 kcal/mol)
  - Previous precision could not examine the error in the force field
- **Question: bias due to solvation ΔG error?**



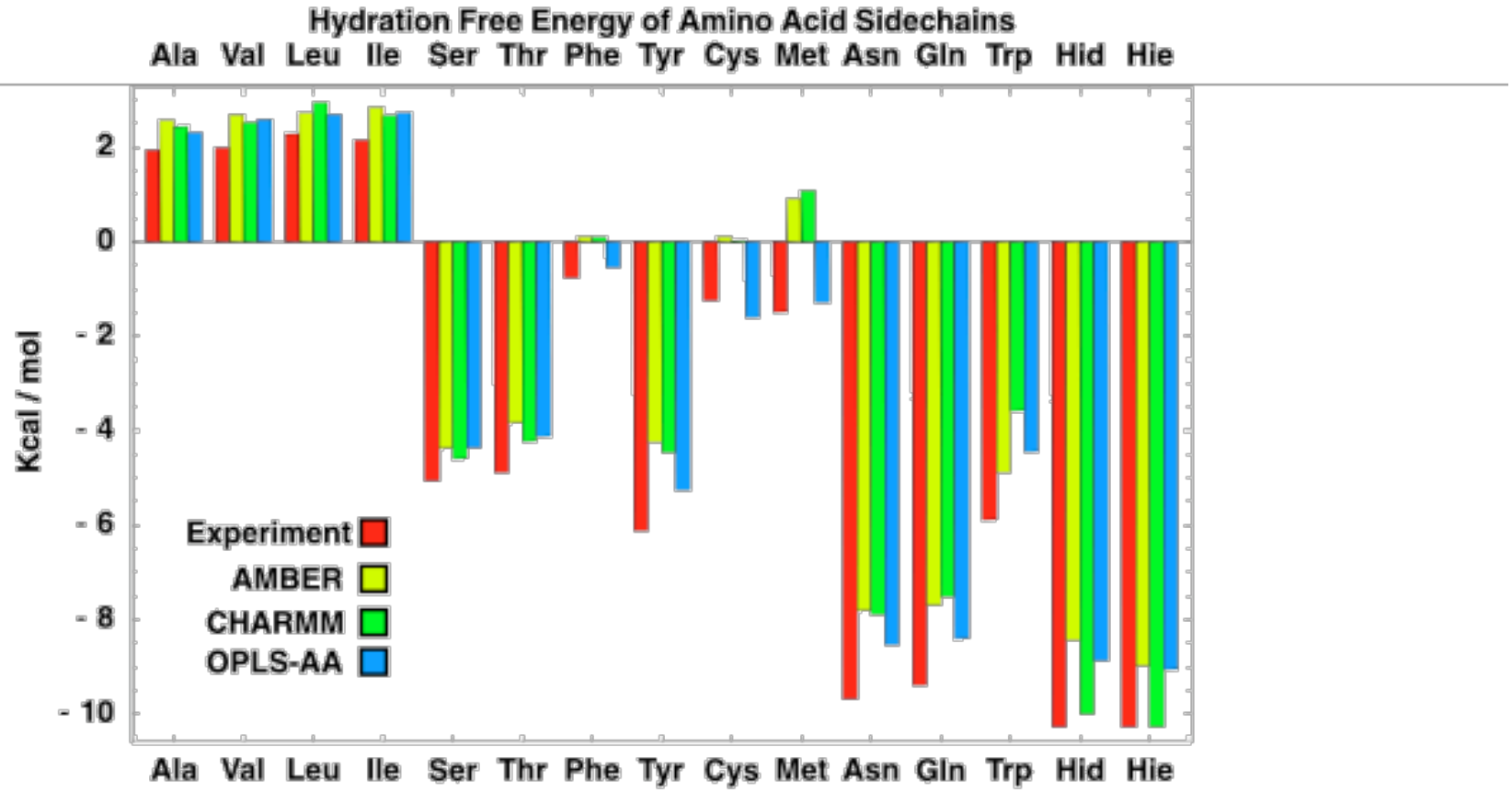
sidechain



sidechain analog

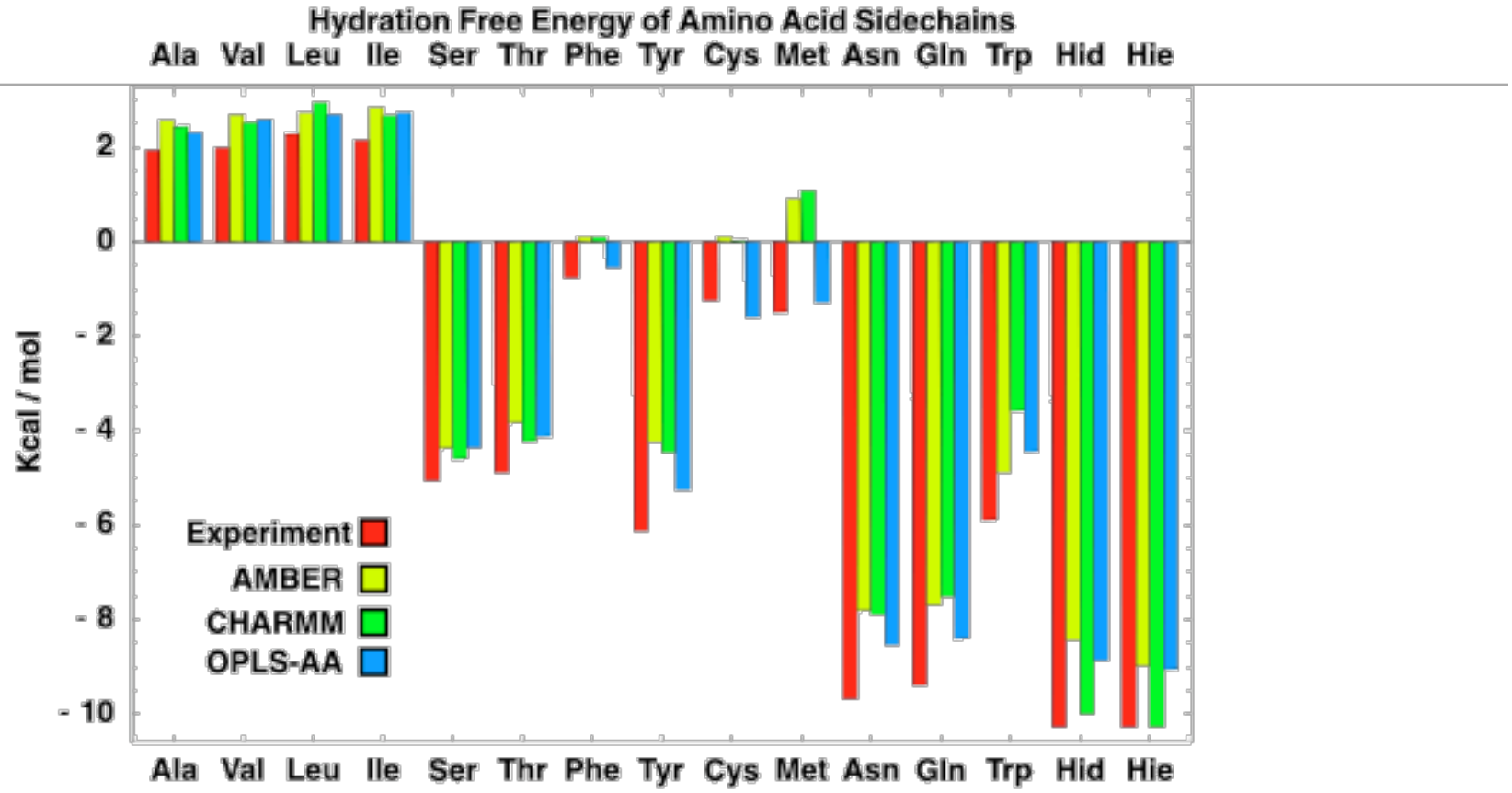
Shirts, Pitera, Swope, Pande, JCP, 117, 5564 (2003)

Comparison with experiment



- **Absolute RMS deviations from experiment (kcal/mol):**
  AMBER: 0.97 CHARMM: 0.84 OPLS-AA: 0.64

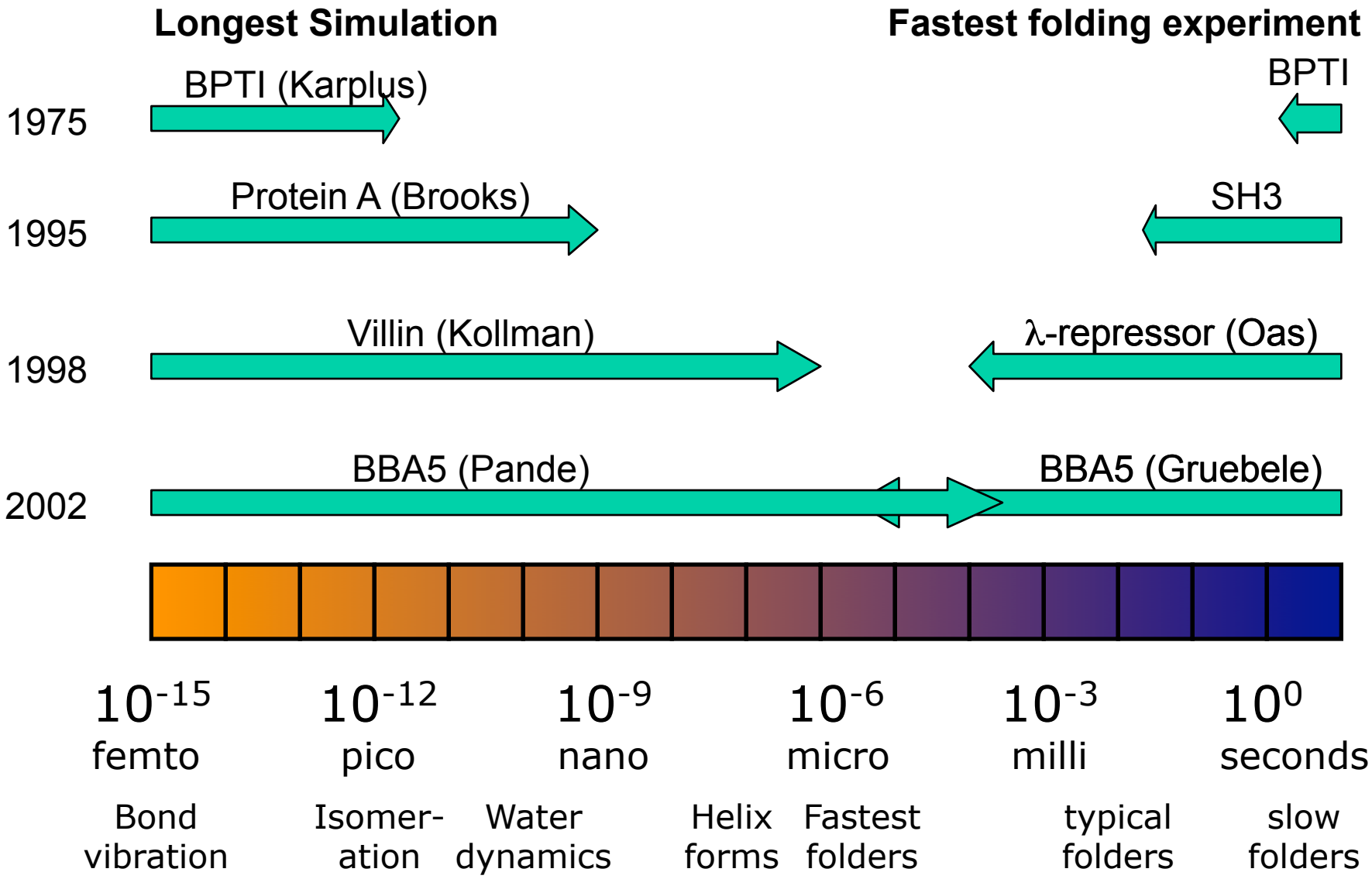
Comparison with experiment



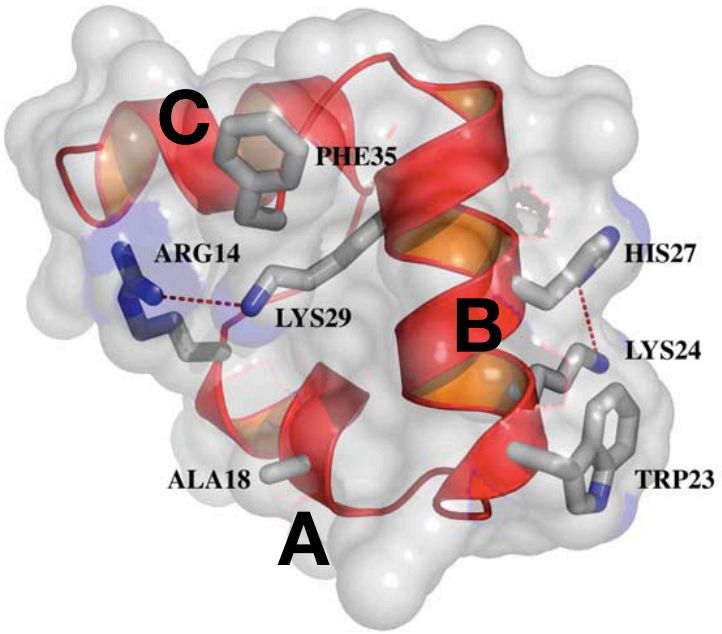
- **Absolute RMS deviations from experiment (kcal/mol):**
  AMBER: 0.97 CHARMM: 0.84 OPLS-AA: 0.64
- **Relative RMS deviations from fit (kcal/mol):**
  AMBER: 0.62 CHARMM: 0.58 OPLS-AA: 0.49

Case study: protein folding kinetics

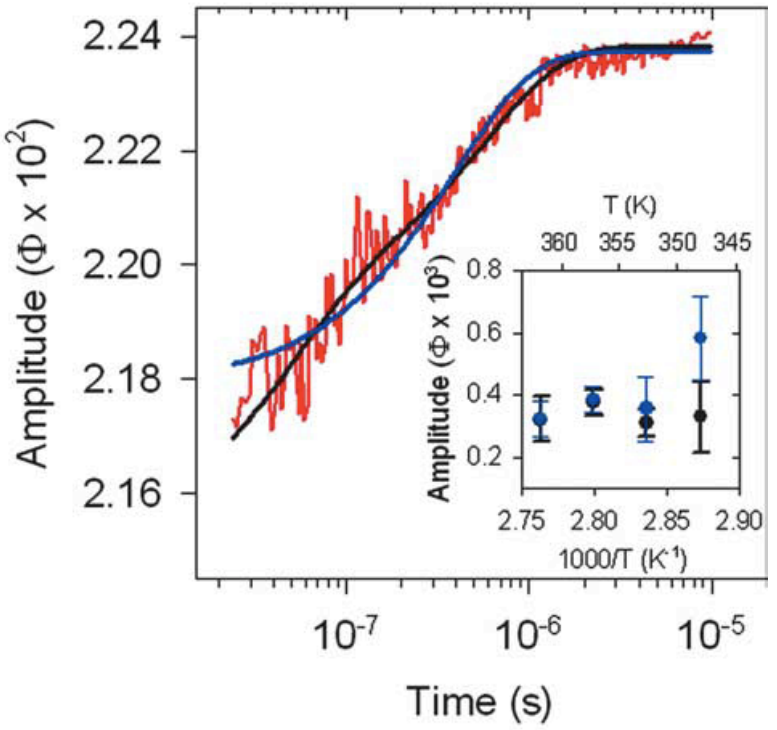
Progress of MD & experiment



A very fast folding protein: k_{fold} ~1/μs



structure



folding kinetics

villin headpiece
mutant designed by the Eaton Lab
(Kubelka *et al*, *JMB* 2006)

Let's look at a 1μs trajectory for villin: we see stochastic behavior

(Ensign, Kasson)

http://simtk.org

- **Simulation details**
  - villin headpiece (36 residues)
  - Eaton mutant (0.7μs folding time)
  - **explicit solvent**
  - 20,000 atoms total
  - AMBER2003 force field
- **MD Engine**
  - GROMACS 3.3.99 (CVS) code
  - SMP on FAH
- **Visualization (VMD)**
  - spacefill: aromatic resides
  - licorice: backbone

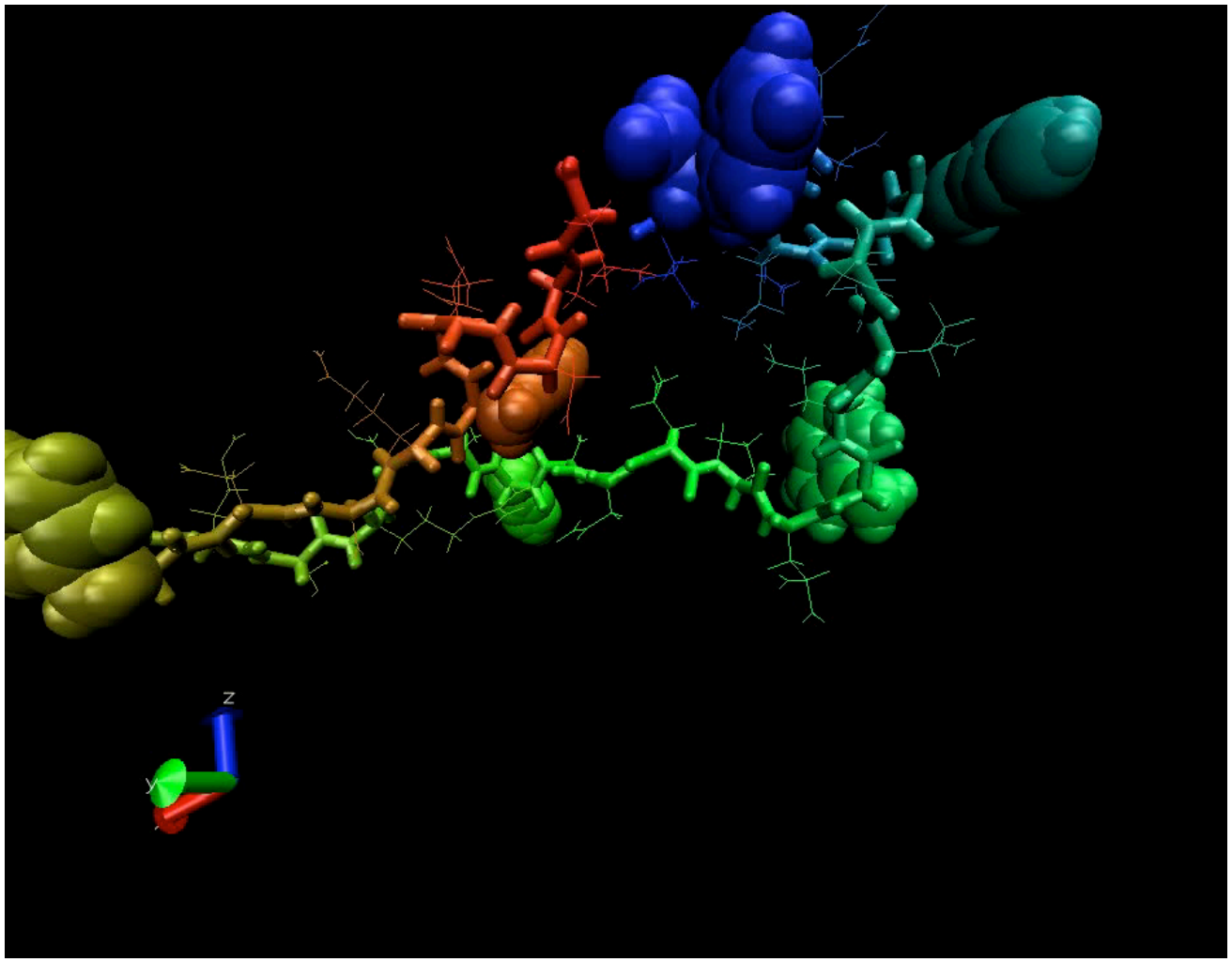
One trajectory of thousands, each on the >1 μs timescale

Ensign, Kasson, & Pande. *JMB* (2007)

Let's look at a 1μs trajectory for villin: we see stochastic behavior

(Ensign, Kasson)

http://simtk.org



One trajectory of thousands, each on the >1 μs timescale

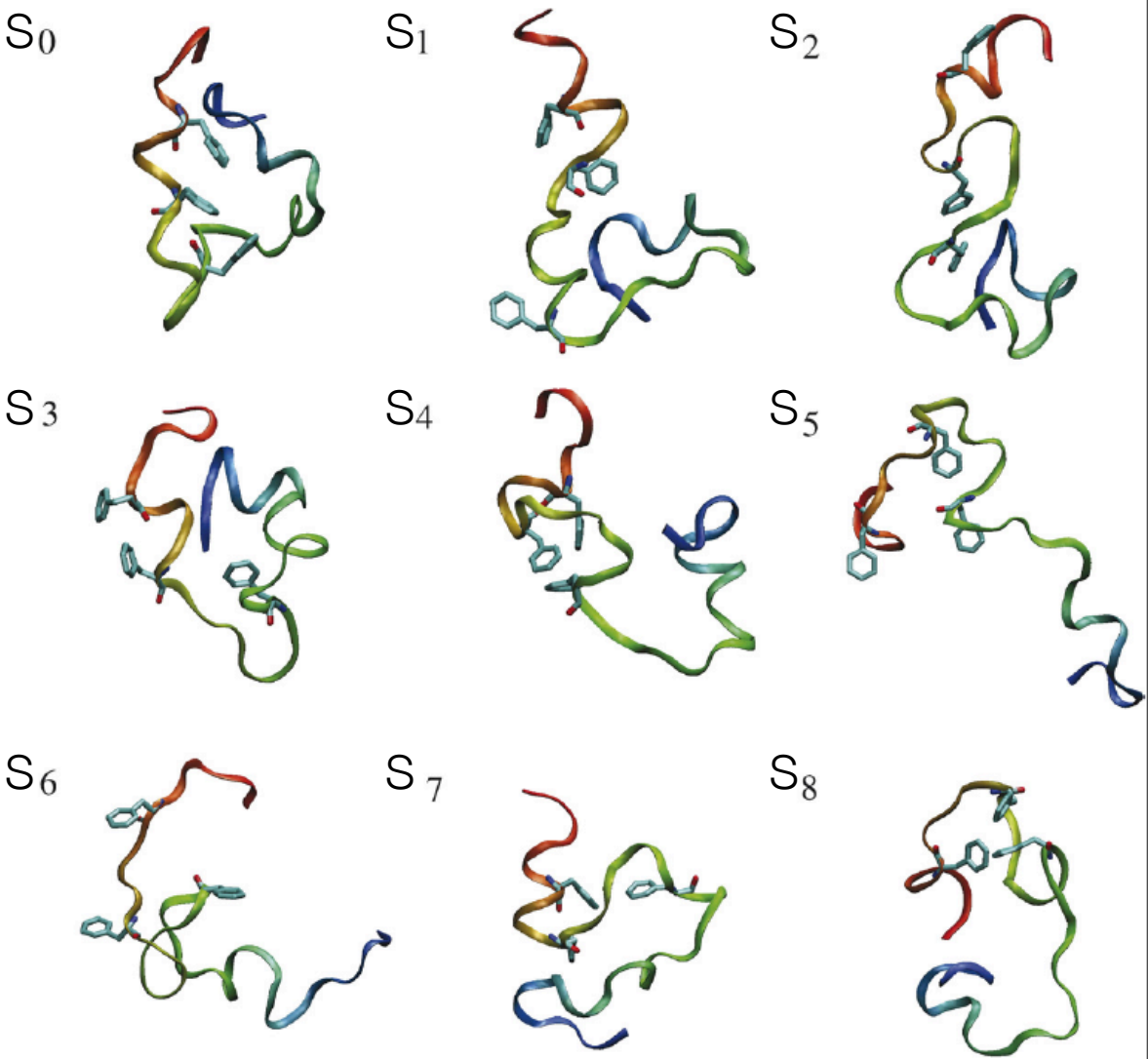
Ensign, Kasson, & Pande. *JMB* (2007)

- **Simulation details**
  - villin headpiece (36 residues)
  - Eaton mutant (0.7μs folding time)
  - **explicit solvent**
  - 20,000 atoms total
  - AMBER2003 force field
- **MD Engine**
  - GROMACS 3.3.99 (CVS) code
  - SMP on FAH
- **Visualization (VMD)**
  - spacefill: aromatic resides
  - licorice: backbone

(Ensign, Kasson)

Looking at ensembles of simulations

- **Starting structures**
  - 9 different structures
  - generated by high temperature unfolding
  - different degrees of native like structure
  - some have helices, other contacts
  - some have no native structure at all
- **Ensemble of trajectories**
  - hundreds to thousands of trajectories per structure
  - each trajectory ~1-2 μs timescale (longer than experimental folding

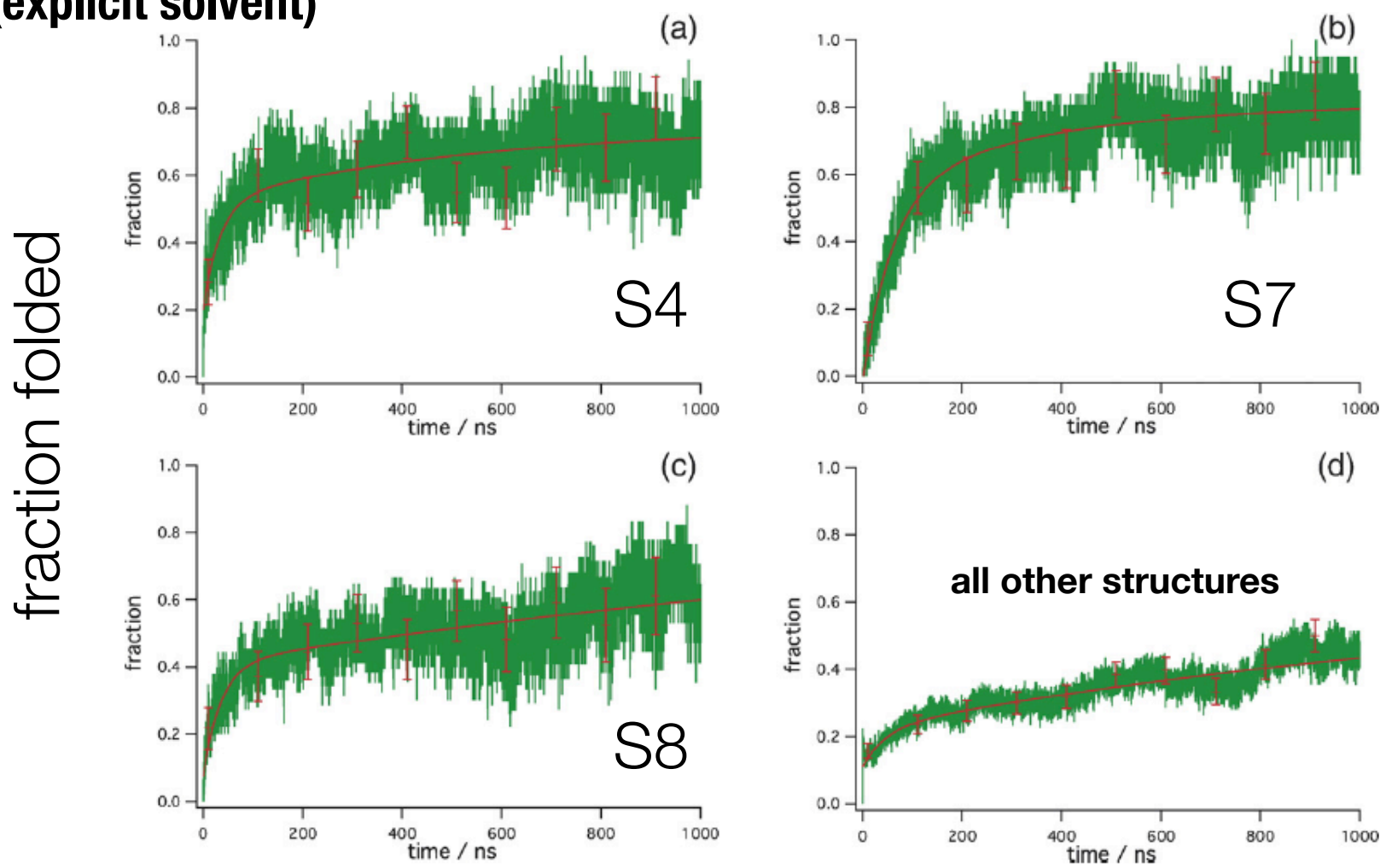


Ensign, Kasson, & Pande. *JMB* (2007)

(Ensign, Kasson)

Ensemble data agrees with experiment

(explicit solvent)



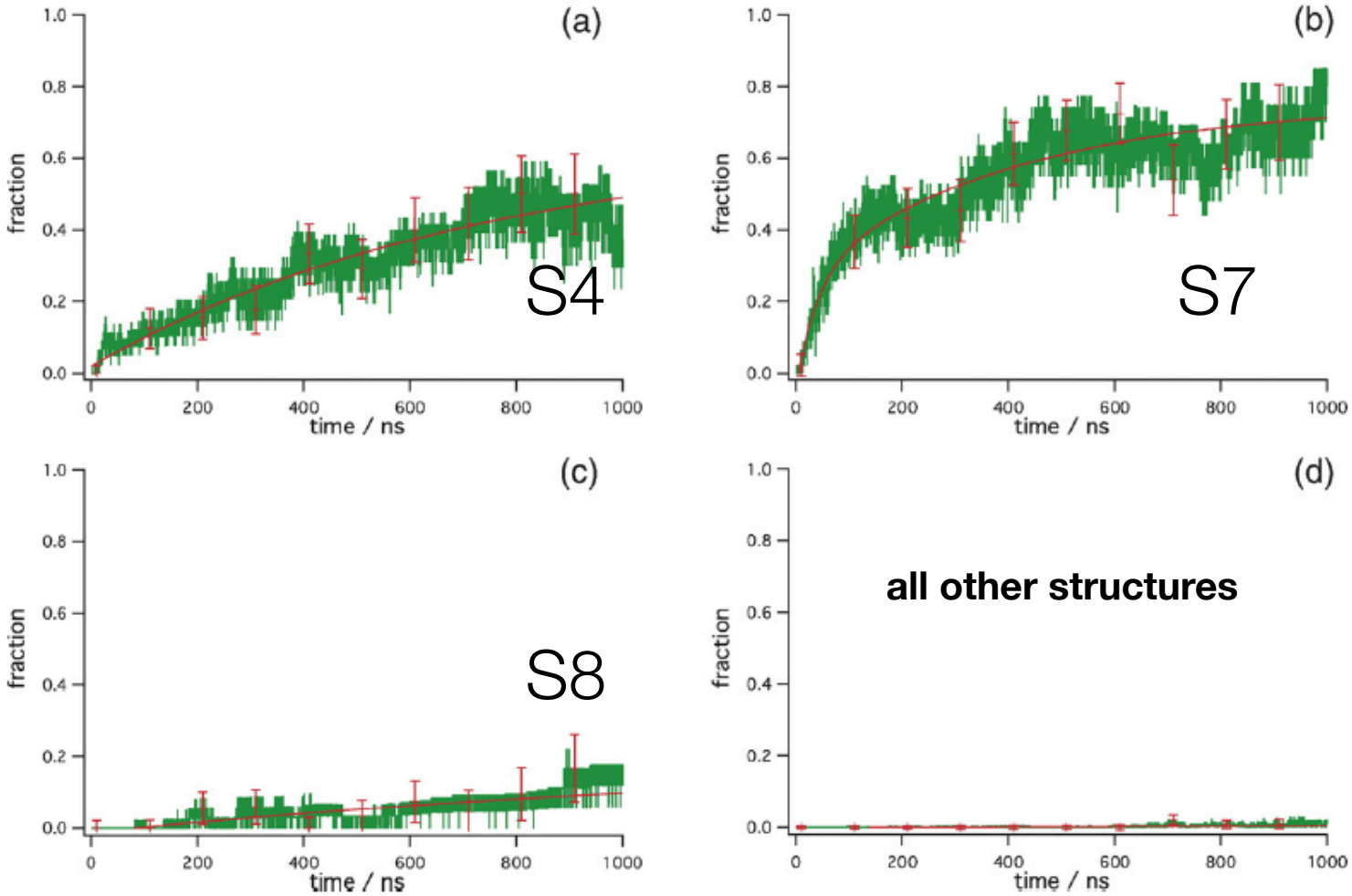
Fraction folded (via Trp-His distance) vs time

Ensign, Kasson, & Pande. *JMB* (2007)

(Ensign, Kasson)

But is the experimental assay looking at folding?

(explicit solvent)

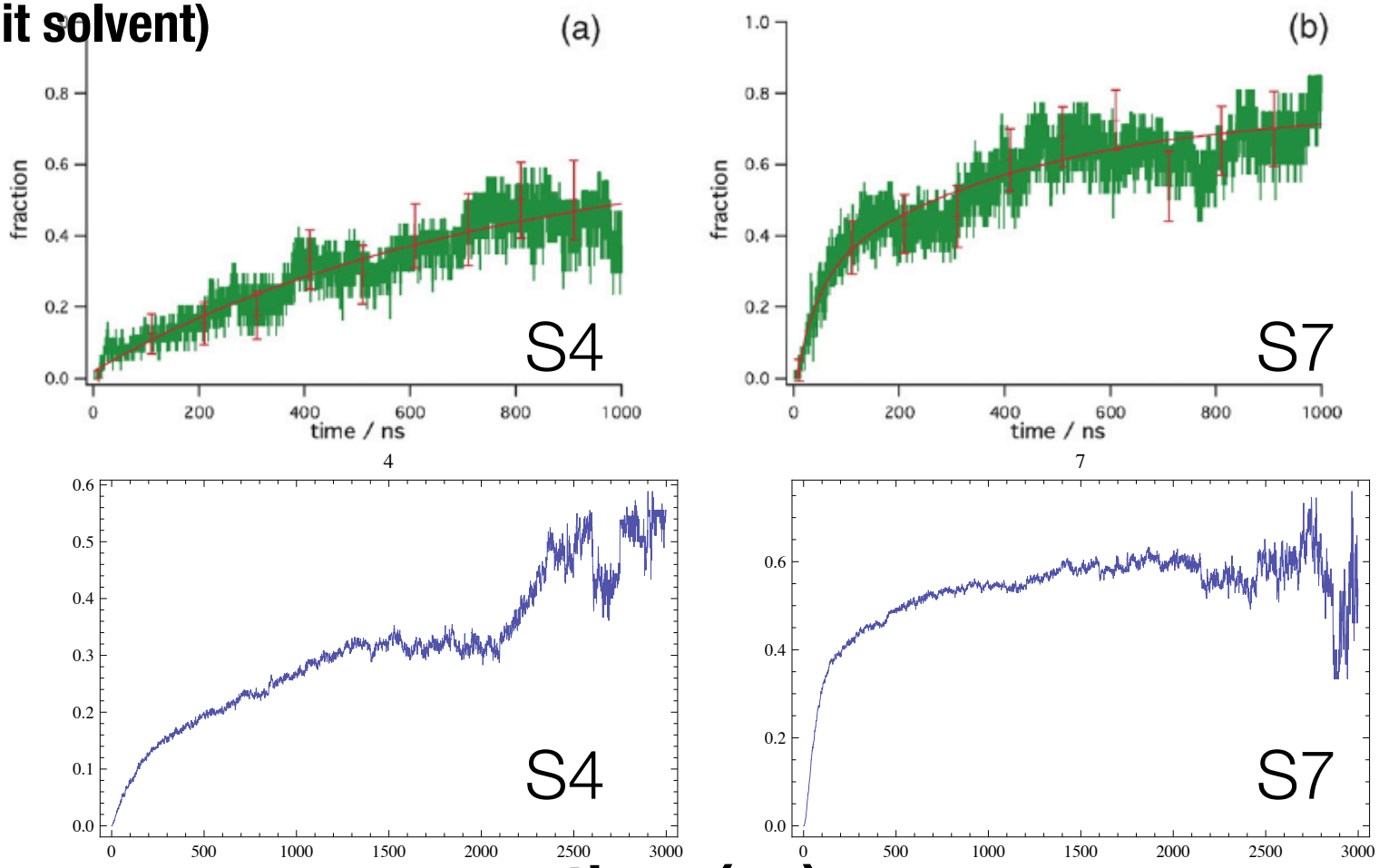


Fraction folded (via comparison to xray structure) vs time

Ensign, Kasson, & Pande. *JMB* (2007)

Comparison between explicit and implicit

(explicit solvent)

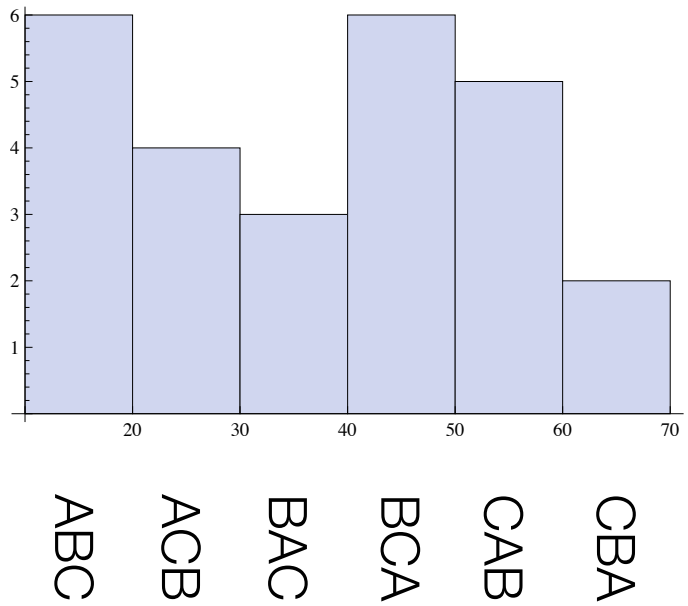


(implicit solvent)

Fraction folded (via comparison to xray structure) vs time

We find a heterogeneous set of folding pathways

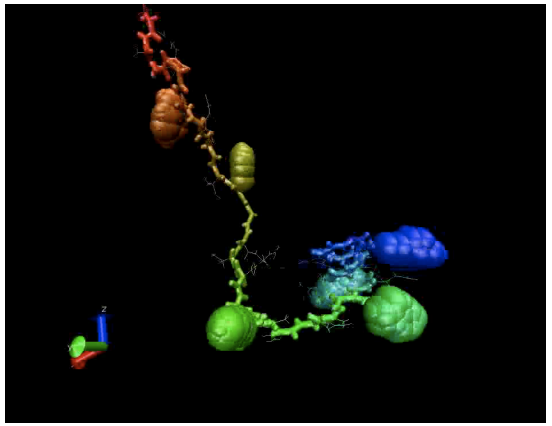
- **Do we see a single pathway or many different?**
- **Test this with a simple question: "Is the order of helix formation consistent between simulations?"**
  - for 3 helices (villin), there are 3! = 6 possible orderings
  - histogram shows a very wide variation of pathways seen
- **Other variations possible too**
  - which key core contacts form first?
- **A single trajectory (or even a few) would give a misleading picture of the folding dynamics**



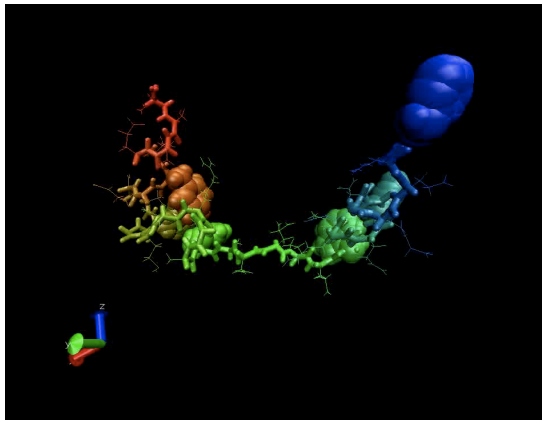
Histogram of folding kinetics: what is the order of formation of each helix A, B, C?

What have we learned about how proteins fold?

- **What did we see in that trajectory?**
  - starts with non-specific hydrophobic collapse
  - unfolds, breaks most contacts
  - refolds, with little native structure
  - some native persist over numerous folding/refolding cycles
  - eventually gets everything right



- **What about other trajectories?**
  - similar behavior in general, but different details
  - great heterogeneity in folding paths



- **General lessons?**
  - Folding is a stochastic process (if the folding time is 1ms, then it's not ½ folded at 0.5 ms)
  - Dynamics of even small molecules can be complex & **very heterogeneous**
  - **Even a few long trajectories aren't enough to inform us about the true nature of the complex phase space -- we need a statistical picture**

Challenges of Molecular Simulation

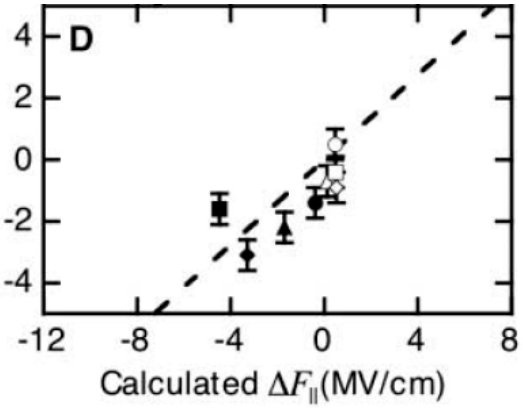
Models *vs* **Sampling**

Are our models sufficiently accurate to answer the questions we're asking?

Have we reached the appropriate equilibrium conditions?

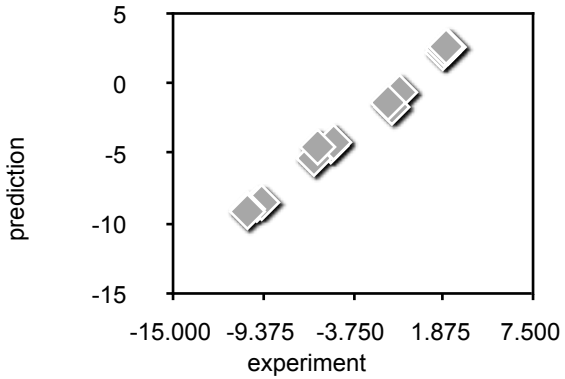
(Shirts, Snow, Zagrovic, et al)

How accurate are atomistic physical models?



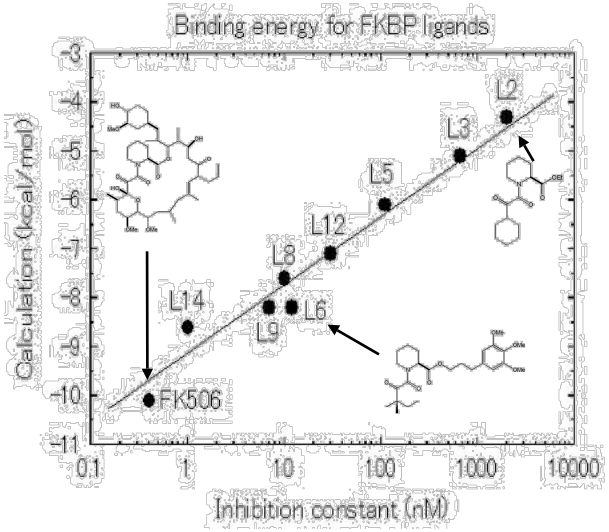
Science, 313 200-4 (2006)

ELECTROSTATICS



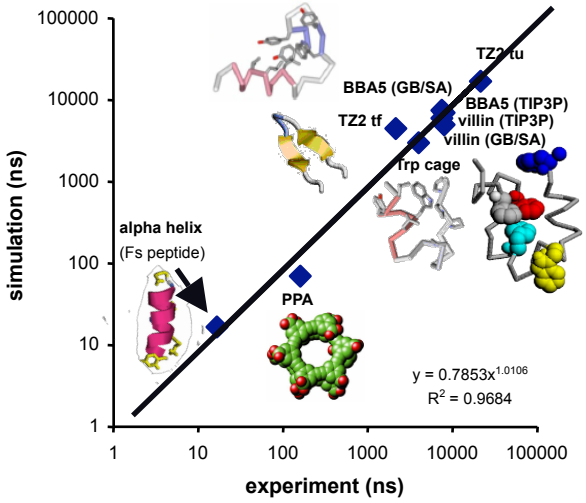
Journal of Chemical Physics, 119 5740-5761 (2003)

SOLVATION FREE ENERGY



Journal of Chemical Physics, 123 084108 (2005)

THERMODYNAMICS

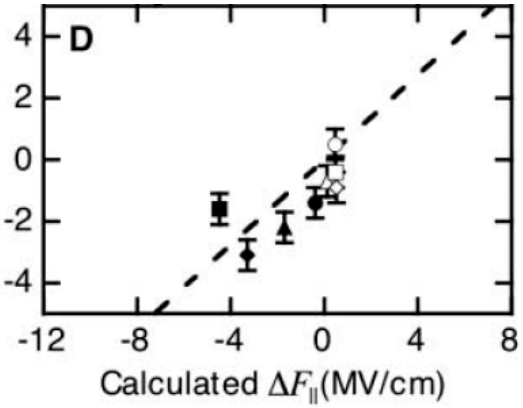


Annual Reviews of Biophysics 34 43-69 (2005)

KINETICS

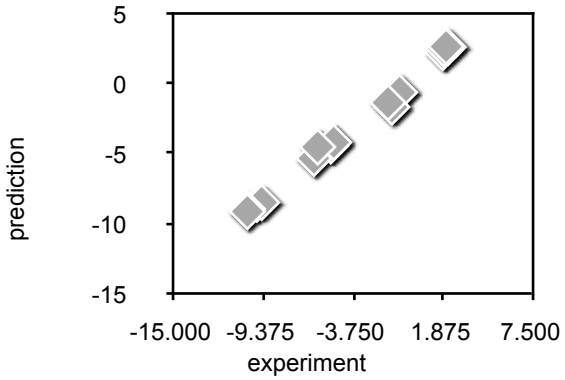
(Shirts, Snow, Zagrovic, et al)

How accurate are atomistic physical models?



Science, 313 200-4 (2006)

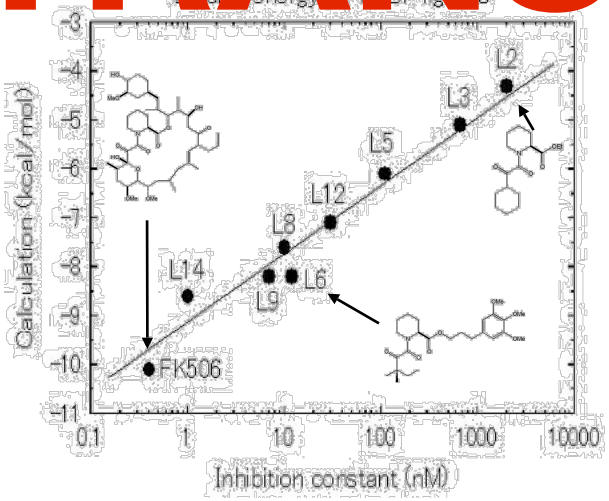
ELECTROSTATICS



Journal of Chemical Physics, 119 5740-5761 (2003)

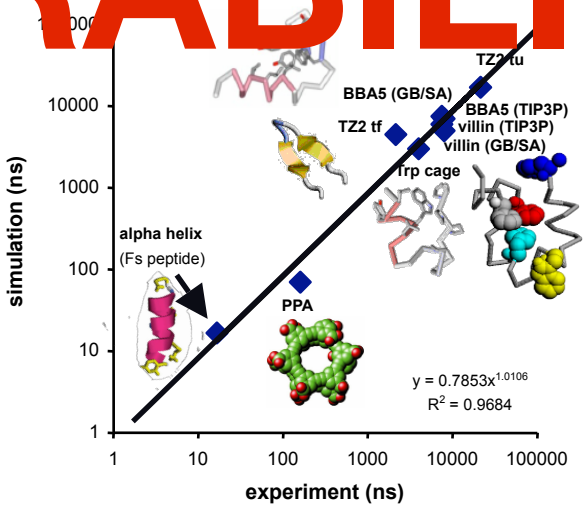
SOLVATION FREE ENERGY

TRANSFERABILITY



Journal of Chemical Physics, 123 084108 (2005)

THERMODYNAMICS



Annual Reviews of Biophysics 34 43-69 (2005)

KINETICS

Summary: What to watch out for

- **Sampling**
  - consider experimental timescales
  - did your results converge? Start from different conditions
- **Model**
  - sufficiently detailed?
  - force field can make a huge difference
- **Analysis**
  - Compare simulation to experimental observables, quantitatively
    - don't compare to experimental interpretation
    - must use numerical comparison
    - ideally compare multiple quantities
  - Understand the uncertainty in simulation and experiment

Where to learn more

- **Books:**
  - Leach, *Molecular Modeling*: Great first resource
  - Gromacs manual (http://gromacs.org): has full derivations and detailed explanations

- **Wikipedia**
  - believe it or not, it's pretty well written and has lots of information

- **Folding@Home: http://folding.stanford.edu**