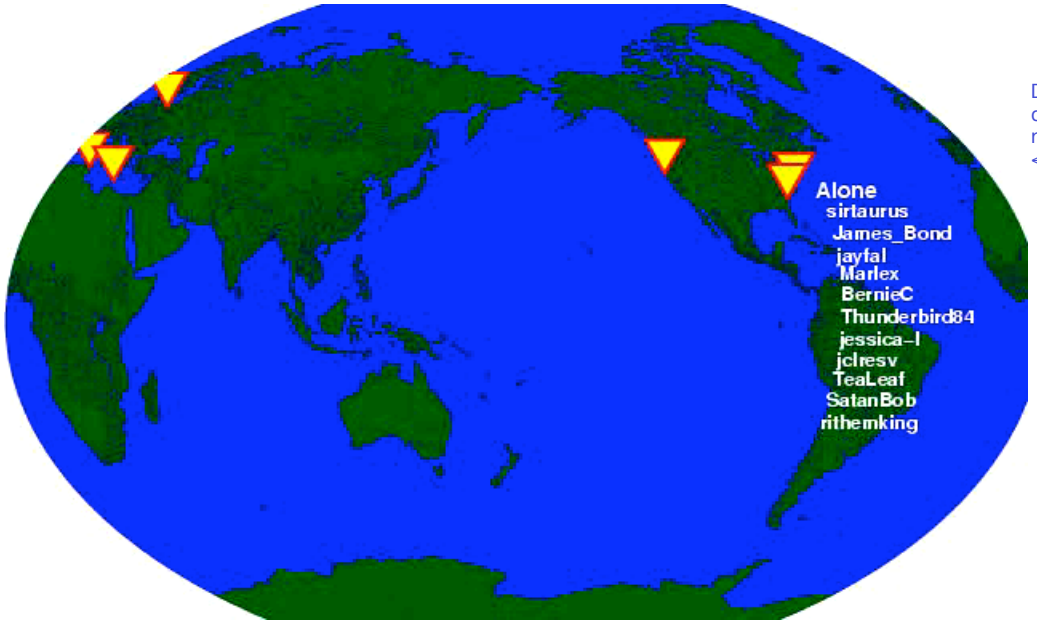
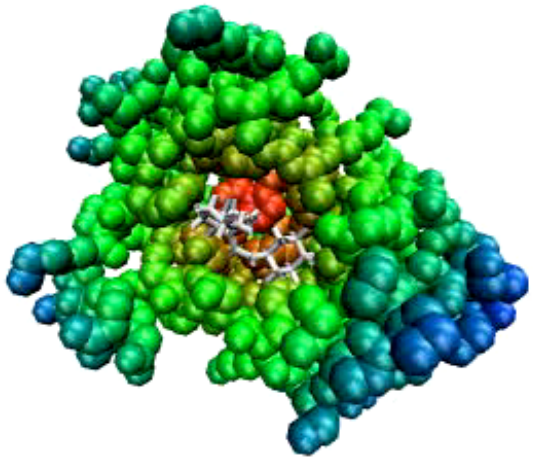


Distributed computing map ←

Parallel Computing Methods for Probing Biomolecular Kinetics and Thermodynamics



Guha Jayachandran
Stanford University
February 12, 2007

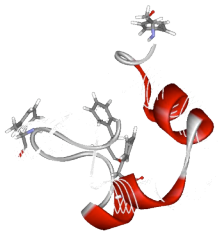
GuhaJ@cs.stanford.edu

Outline

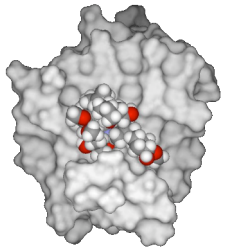
- Introduction and Problem
- Computational Infrastructure

Folding@home distributed computing

- Protein Folding: methods with application



- Protein-Ligand Binding: methods with application

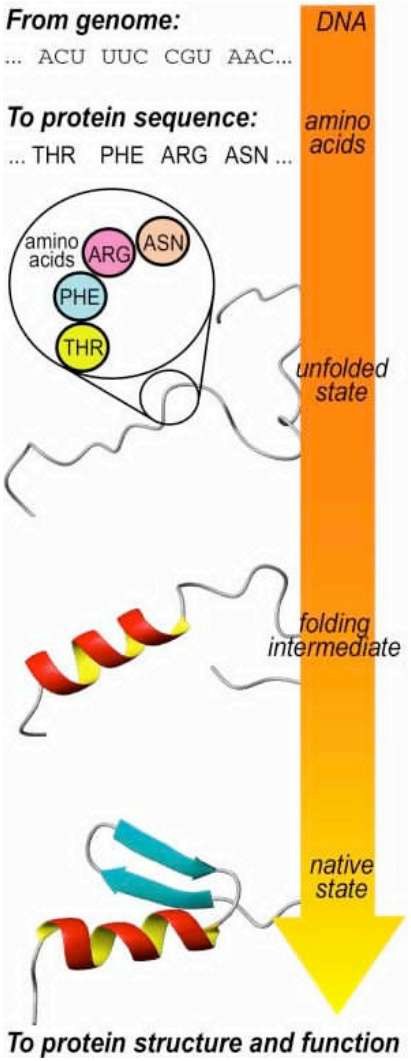


- Conclusion

Introduction and Problem

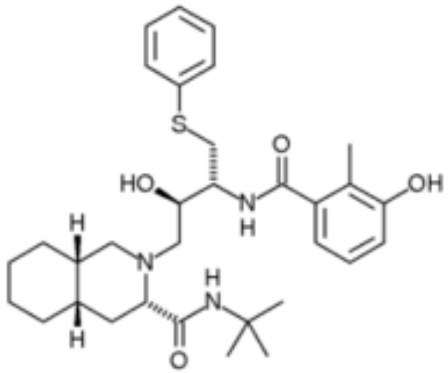
The Molecular Cast

- Protein
  - Carries out the functions of life at the cellular level
  - Paradigm for nanotechnology
  - Folds into a given 3D structure
    - Sequence → structure → function
    - Misfolding can cause disease



The Molecular Cast

- Small molecule ligand
  - Low molecular weight organic molecule
  - Can modulate proteins' activity by binding to them
  - Many drugs are small molecules that work by binding to a given protein
  - Probability of a ligand and protein being bound is a function of their binding free energy

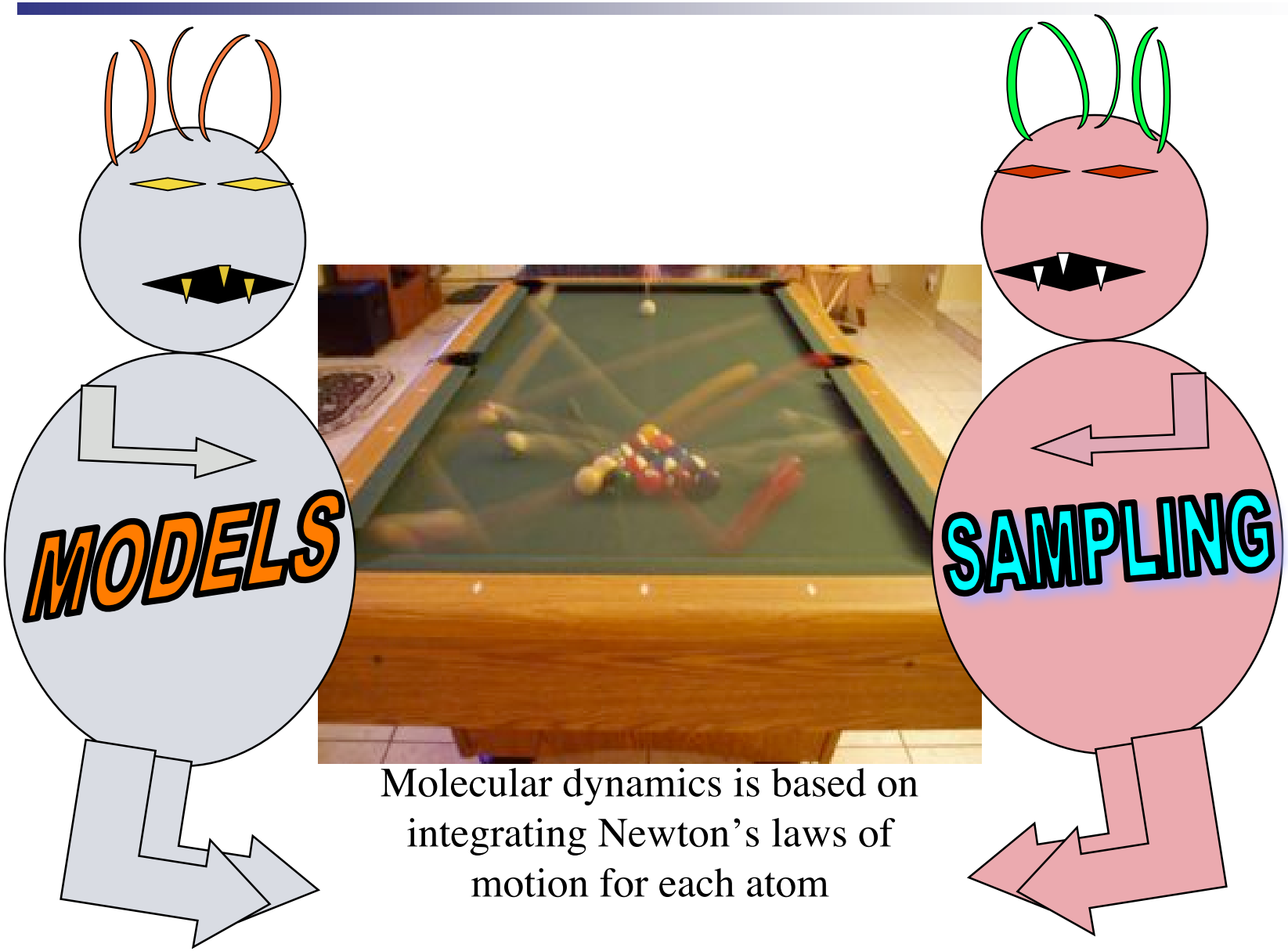


Nelfinavir: Binds to HIV-1 and HIV-2 proteases, inhibiting them from cleaving viral protein. Used in treatment of AIDS.

Grand Challenges in Biophysics

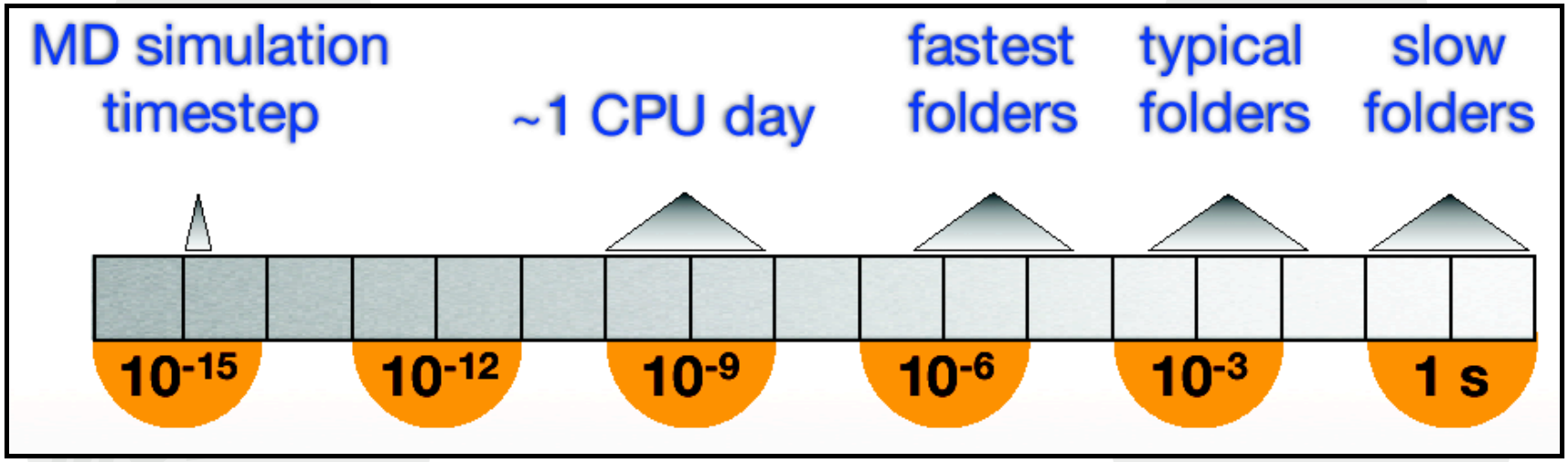
Biophysics	Biocomputation
▪ Protein structure and dynamics	
▪ Explain structure	Predict folded structure given sequence
▪ Elucidate folding mechanisms	Simulate folding from unfolded state… in detailed model, to long time
▪ Protein-ligand binding	
▪ Say how well a given ligand will bind to a protein	Accurately compute solvation and binding free energies

Difficulties in Computation



Molecular dynamics is based on integrating Newton's laws of motion for each atom

Difficulties in Computation



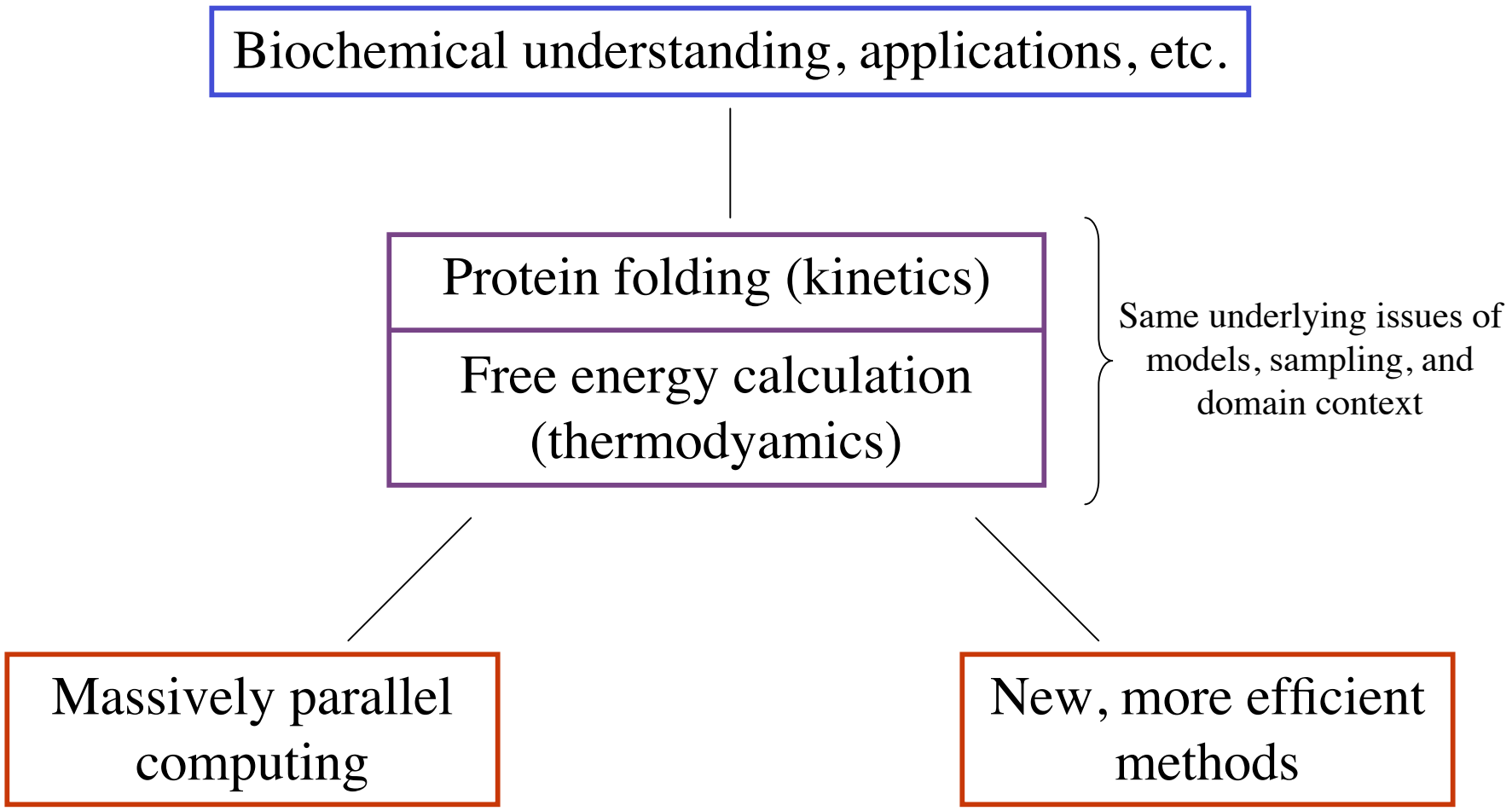
If you could simulate 1 ns/day, then a 1 ms simulation would take 3000 years!

Difficulties in Computation

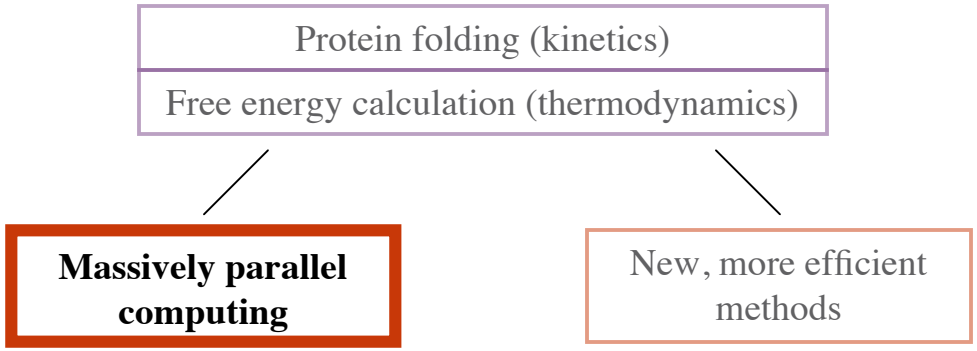
Need:

More computational power and better computational techniques

Means and Motivations

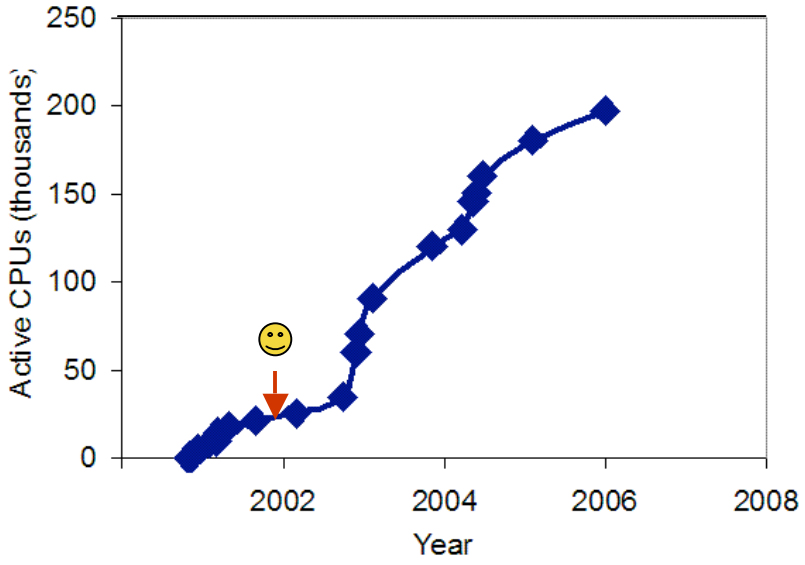


Distributed Computing

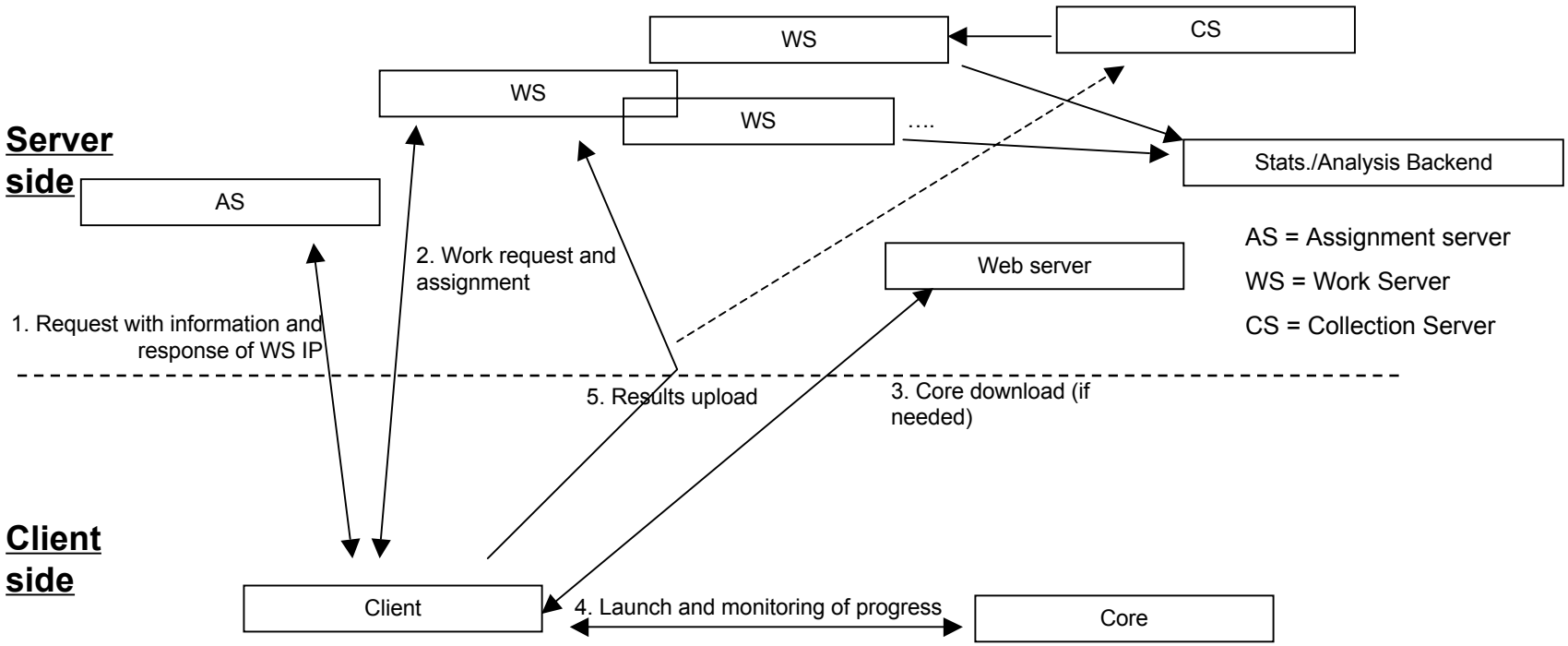


Folding@Home

- Distributed computing project (**http://folding.stanford.edu**): anyone can download our software (Linux/Windows/Macintosh)
- Clients run assigned simulations with unused processor cycles
- Steady growth in install base
- Current strength:
  - ~190,000 active clients
  - 223 TFLOPS sustained



Global Client/Server Architecture

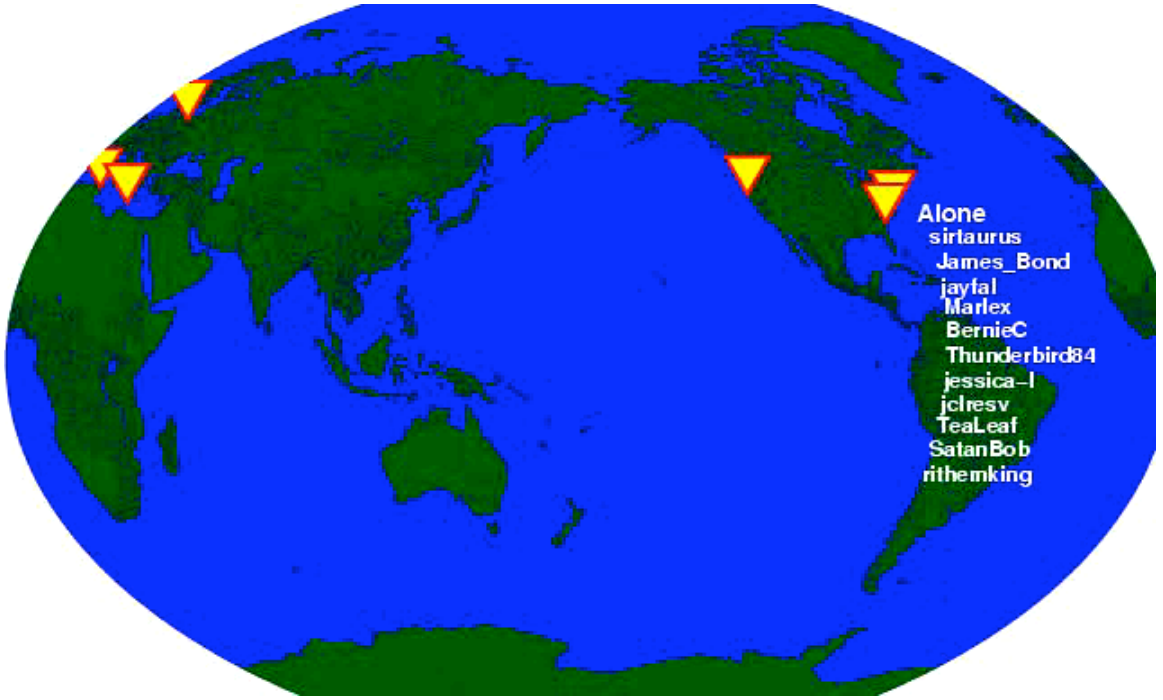


Global Client/Server Architecture

Server side

Servers are located in 3 campus buildings and are of varying ages, powers.

Client side



Client computers are in 6 continents and upgrade on their own.

Participants

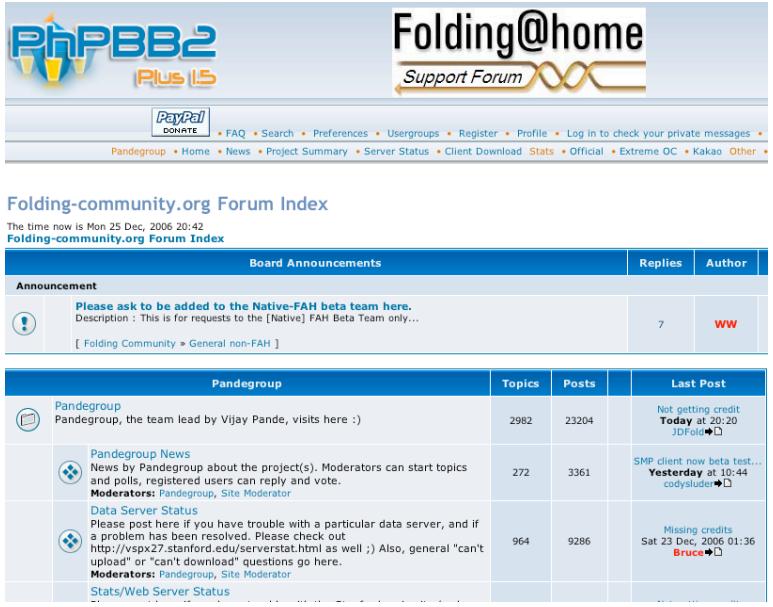
Online statistics let participants track production and compete for top rankings.
People also form teams and online community.

Donor Statistics — Last updated: Monday Decembe

Number	Name	Credit	Total	Te
search	is exactly	to	to	
	Top 100 users			
1	anonymous	108926492	1224577	0
2	eastms.edu	44979648	407172	17
3	OC-AMD	43431574	356322	30
4	1920x1080i-How???	41185582	228140	64
5	FLECOM	20326537	120354	33
6	Silicon_Wars_-_a_New_Hope	18200167	133777	24
7	Scott_H	16835547	137364	26
8	Christian_Bargmann	12145552	73996	19
9	nikhsub1	12114344	129562	32
10	Sceggs.nsw.edu.au	10371612	80322	24
11	newtownhigh	9406884	79979	37
12	KWSN_MillenniumKnight	9090336	81014	11
13	ulf.van.laak	8641758	38310	28
14	larrybpsu	8443546	35575	39
15	ITAP_RCAC	8060743	47344	15
16	[LEENFAT]dapo	7692747	39915	51
17	marty9876	7664570	38040	33
18	wayne	7649388	36977	46
19	Wedo	7637471	131927	32
20	KWSN_Sir_CADCAM	7509438	73622	11
21	belloq	7293691	61165	18
22	University_of_Alabama_Computer_Science	7257739	71821	10
23	DorXtar_(Team_Rage3D)	7233542	62773	64
24	Netsphere	7144986	88572	32
25	Bollocks	6869277	52238	24
26	ChasR	6622742	26733	32
27	Toasty[H]ardOCP	6363649	50035	33
28	fcapes	6315845	72204	4
29	Bhurak_AU	6313517	39477	24
30	Mustanley	6223423	58366	32
31	Trinedy	6094696	38303	24
32	Robert_Ameeti	6065403	32253	33

Rankings screenshot.

Some team logos.



User run support Forum.

Partial Change Log

5.04:
- Allows configuration of the amount of memory to indicate available, especially useful when running multiple "big unit" clients on a single machine.
- Allows for the setting of -advmethods as a persistent option through the Configuration.

5.02:
- Error checking on queue, related to big packet option.
- OpenBSD branding fix. (5.01)

5.00:
- Collection server support. If a client finishes a work unit from a Work server with this feature enabled, then if 2 upload attempts to that server fail, the client will return to the work unit to a Collection server. Pending credit will be given for the unit, becoming final when the Work server comes back up and verifies what the Collection server has received. All work servers will eventually be associated with a Collection server, and many already are.
- Configuration option to allow downloads and uploads greater than 5 MB in size, which some new projects will require due to the size of what they are simulating.
- G@H configuration option reworded to be "No deadline" option.
- Print executable name and launch directory to log.
- Fix for bug where queue slot could get stuck "Fetching"
- MyFolding links updated.
- -forceSSE flag removed, since even with older clients, the flag had become meaningless because recent Core_78's give SSE priority over 3DNow automatically.
- OpenBSD support: If give flag -openBSD, client will automatically make elf2olf system call on downloaded Cores.
- Win. only: Configuration option to pause work when the computer is using battery power (for laptops).
- Win. console: Built in service installation support through Configuration, to have the program automatically begin at startup.

4.00:
- Configurable minimum checkpointing frequency (3-30 minutes).
- If client run with verbosity of at least 5, FahCore_78 will print out a message whenever a timered checkpoint occurs.
- -configonly flag added to configure then exit.
--oneunit flag added to exit after completing one unit.
- -forceSSE flag added.
- Warnings printed if -forceasm/-forceSSE flags given.
- Benchmarking not done except at start of normal operation (for example, not if only configuring, listing queue, deleting unit, etc.).
- Warning if attempt to run client with Machine ID of already running client
- Web site printed updated.
- Upon finishing a unit, if fails to return the result but succeeds in getting new work, immediately tries sending all results again.
- Checks user name and team number for legality of use in statistics.
- Queue info properly outputted to FAHlog.txt.
- "Ask before connecting" change on Linux -- ignores typeahead and asks for "C"
- Proxy password written out in binary (upon reconfiguration)
- Assignment/work server request authentication
- Send all message altered ("sent x of y" if x not 0, or "no units...")
- CPU usage default of 100 printed out during Configuration.
- Tags in MyFolding page changed to better adhere to HTML standard
- Maximum machine IDs raised to 8.
- Upload and download bandwidths estimated to send to AS.
- If fails to post data, includes server and port in message. Also outputs HTTP status code (before this was only done if I.E. settings used).
- Bug fixed where performance fraction could exceed 1.
- Batch networking option for G@h (Network when less than 2 units ready to work on or more than 5 results to return, and when network, both get up to 10 units and send any results).
- If set G@h preference, reflected in acknowledgment message.

3.24 (Dec. 23, 2002)**:**
- Benchmarking improved in consistency and uniformity.
- Support for sending a measure to the assignment server of time taken on past work units (a weighted average). Can view this measure using verbosity 4 after finished units.
- If bad arguments given, will exit only after a key press.
- If upon startup, FAHlog.txt exceeds 50Kb and a is to be started, the old one is retained under the name FAHlog-Prev.txt.
- Deadline information included with work (Printed to unitinfo.txt and available in queueinfo). Active units checked for expiration periodically all others checked for expiration at autosend time. Configuration option of ignoring deadline information (to use if system clock is bad).
- MyFolding.html page written out by client with useful links (which Linux and Win. console users did not have easy access to before)
- If specify -freeBSD flag, downloaded cores will automatically be branded as Linux (should make usage seamless on FreeBSD platform).
- In making request to server for work, inform on whether have units done waiting to be uploaded to server. If so, servers running new code will give a new unit instead of giving same.
- If a work unit is rejected for any reason, server now returns reason to client, which prints it out. If WU fatally flawed, deleted from queue.
- Checks for bad URL, port on reception of work unit, deleting it before even starting work if bad.
- "Could not transmit unit" statement now also includes date unit was completed ("Could not transmit unit x (Completed xxxx)").
- Time stamp on every line of FAHlog.
- Arguments to client printed to log.
- Some log messages modified (like the requested, and deserved, "Thank you" added!).
- File "unitinfo.txt" printed and updated, with information and progress for current work unit. (Had been requested for use by 3rd party people.)
- Answers to Configuration questions now validated. One letter answers (i.e. "y" or "Y" instead of "yes") allowed.

3.14:
- fixed bug in configuration of core priority.
- -forceasm command line option supported to force assembly optimizations (could be offed again by core if problems) IF current work unit is a Core_78 WU.
- If fails to connect to assignment server, client will no longer just go back to previous work server, since that server may no longer be appropriate for this client.
- Fixed message where if run "-send all," user was incorrectly told that failed.
- If FAHlog.txt is locked by OS, FAHlog2.txt started.
- -advmethods option added for users wishing to try the newest cores and/or WU's
- Advanced Configuration option of disabling optimized assembly code added.
- Send/autosend conflict resolved.
- If CORE_OUTDATED situation arises, and remains even after a new core is downloaded, the current work unit is deleted.

3.12:
- Rather than in the /tmp directory, the User ID is stored in the same directory as the client, in a file called machinedependent.dat.

3.11:
- If Client-core communications error occurs, current work unit and work files deleted.
- If 5 consecutive cores that have been downloaded during the current execution of FAH fail to complete a work unit, the client will sleep for 1 day. When it wakes up, it give one more downloaded core a chance to run and complete a WU before sleeping again.
- FAHlog.txt is appended to (until it reaches 50 Kb) instead of overwritten.

3.1:
- Preference of project added to Configuration's Advanced tab.

Partial Change Log

Assignment/work server request authentication

Provide unitinfo and MyFolding pages

Batch networking option for G@h. Network when less than 2 units ready to work on or more than 5 results to return. Get 10 units at once.

Collection server support. If a client finishes a work unit and 2 upload attempts to the work server fail, the client will return to the work unit to a Collection server.

Support for sending a measure to the assignment server of time taken on past work units

-advmethods option added for users wishing to try the newest cores and/or work units

Send clients work unit completion deadline beyond which to give up

Configurable minimum checkpointing frequency

If a work unit is rejected, server now returns specific reason to client, which prints it out. If the unit is fatally flawed, then it's deleted from the client queue.

Scientific Results

Local structure formation in simulations of two small proteins
Guha Jayachandran, V. Vishal, Angel E. Garcıa and V. S. Pande. Journal of Structural Biology (2006)

Kinetic Definition of Protein Folding Transition State Ensembles and Reaction Coordinates
C. Snow and V. S. Pande. Biophysical Journal (2006)

Parallelized Over Parts Comp. of Absolute Binding Free Energy with Docking and Mol. Dynamics
Guha Jayachandran, M. R. Shirts, S. Park, and V. S. Pande. Journal of Chemical Physics (2006)

Folding Simulations of the Villin Headpiece in All-Atom Detail
Guha Jayachandran, V. Vishal, and V. S. Pande. Journal of Chemical Physics (2006)

Ensemble molecular dynamics yields submillisecond kinetics and intermediates of membrane fusion
P. Kasson, N. Kelley, N. Singhal, M. Vrjlic, A. Brunger, and V. S. Pande. PNAS, USA

Electric Fields at the Active Site of an Enzyme: Direct Comparison of Experiment with Theory
Ian T. Suydam, Christopher D. Snow, Vijay S. Pande, Steven G. Boxer. Science (2006)

A novel approach for computational alanine scanning: application to the p53 oligomerization domain
L.T. Chong, W. C. Swope, J. W. Pitera, and V. S. Pande. Journal of Molecular Biology (2006)

On the role of chemical detail in simulating protein folding kinetics
Young Min Rhee and Vijay S. Pande. Chemical Physics (2006)

Nanotube confinement denatures protein helices
Eric J. Sorin and Vijay S. Pande. JACS (2006)

The solvation interface is a determining factor in peptide conformational preferences
Eric J. Sorin, Young Min Rhee, Michael R. Shirts, and Vijay S. Pande. J. Molecular Biology (2006)

How large is alpha-helix in sol.? Studies of the radii of gyr. of helical peptides by SAXS and M.D.
Bojan Zagrovic, Guha Jayachandran, I. S. Millett, S. Doniach and Vijay S. Pande. J. Mol. Biology (2005)

Error Analysis in Markovian State Models for protein folding
Nina Singhal and Vijay S. Pande. Journal of Chemical Physics (2005)

A New Set of Molecular Mechanics Parameters for Hydroxyproline and Its Use in Molecular Dynamics Simulations of Collagen-Like Peptides
Sanghyun Park, Randall J. Radmer, Teri E. Klein, and Vijay S. Pande. J. Computational Chemistry (2005)

Solvation free energies of amino acid side chain analogs for common molecular mechanics water models
Michael R. Shirts and Vijay S. Pande. Journal of Chemical Physics (2005)

Foldamer dynamics expressed via Markov state models. I. Explicit solvent molecular-dynamics simulations in acetonitrile, chloroform, methanol, and water
Sidney Elmer, Sanghyun Park, & Vijay S. Pande. Journal of Chemical Physics (2005)

Unusual compactness of a polyproline type II structure
Bojan Zagrovic, J. Lipfert, E.J. Sorin, I. S. Millett, W.F. van Gunsteren, S. Doniach & V.S. Pande. PNAS (2005)

Empirical Force-Field Assessment: The Interplay Between Backbone Torsions and Noncovalent Term Scaling
Eric J. Sorin and Vijay S. Pande. Journal of Computational Chemistry (2005)

Exploring the Helix-Coil Transition via All-atom Equilibrium Ensemble Simulations
Eric J. Sorin and Vijay S. Pande. Biophysical Journal (2005)

Does Water Play a Structural Role in the Folding of Small Nucleic Acids?
Eric J. Sorin, Young Min Rhee, and Vijay S. Pande. Biophysical Journal (2005)

Dimerization of the p53 Oligomerization Domain: Identification of a Folding Nucleus by M.D. Simulations
Lillian T. Chong, Christopher D. Snow, Young Min Rhee, and Vijay S. Pande. J. Molecular Biology (2005)

Simulations of the role of water in the protein-folding mechanism
Young Min Rhee, Eric J. Sorin, Guha Jayachandran, Erik Lindahl, & Vijay S Pande. PNAS (2004)

Trp zipper folding kinetics by molecular dynamics and temperature-jump spectroscopy
Christopher D. Snow, Linlin Qiu, Deguo Du, Feng Gai, Stephen J. Hagen, & Vijay S Pande. PNAS (2004)

Does Native State Topology Determine the RNA Folding Mechanism?
Eric J. Sorin, B.J. Nakatani, Y.M. Rhee, Guha Jayachandran, V Vishal, & Vijay S Pande. J. Mol. Biology (2004)

Solvent Viscosity Dependence of the Folding Rate of a Small Protein: Distributed Computing Study
Bojan Zagrovic and Vijay S. Pande. Journal of Computational Chemistry (2003)

Insights Into Nucleic Acid Conformational Dynamics from Massively Parallel Stochastic Simulations
Eric J. Sorin, Young Min Rhee, Bradley J. Nakatani & Vijay S. Pande. Biophysical Journal (2003)

Multiplexed-Replica Exchange Molecular Dynamics Method for Protein Folding Simulation
Young Min Rhee & Vijay S. Pande. Biophysical Journal (2003)

The Trp Cage: Folding Kinetics and Unfolded State Topology via Molecular Dynamics Simulations
Christopher D. Snow, Bojan Zagrovic, and Vijay S. Pande. Journal of the American Chemical Society (2002)

Absolute comparison of simulated and experimental protein-folding dynamics
Christopher D. Snow, Houbi Ngyen, Vijay S. Pande, and Martin Gruebele. Nature (2002)

Native-like Mean Structure in the Unfolded Ensemble of Small Proteins
Bojan Zagrovic, C.D. Snow, Siraj Khaliq, M.R. Shirts, and Vijay S. Pande. J. Mol. Biology (2002)

Sim. of Folding of a Small Alpha-helical Protein in Atomistic Detail using Worldwide Distributed Computing
Bojan Zagrovic, Christopher D. Snow, Michael R. Shirts, and Vijay S. Pande. J. Molecular Biology (2002)

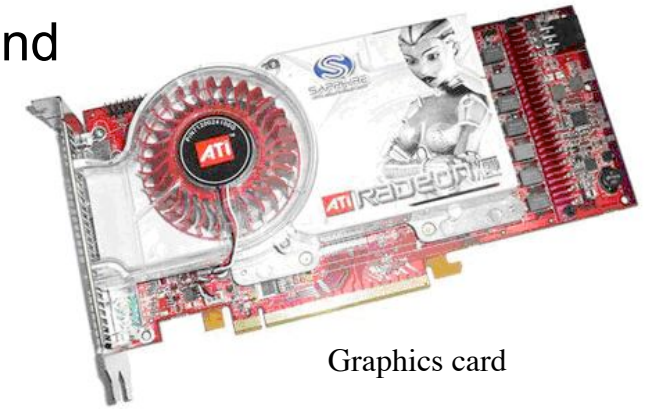
Atomistic protein folding simulations on the submillisecond timescale using worldwide distributed computing
Vijay Pande, et al. Peter Kollman Memorial Issue, Biopolymers (2002)

New Hardware Platforms

- Sony PlayStation 3
  - Cell processor
  - Client/core port by SCEA (Redwood City)

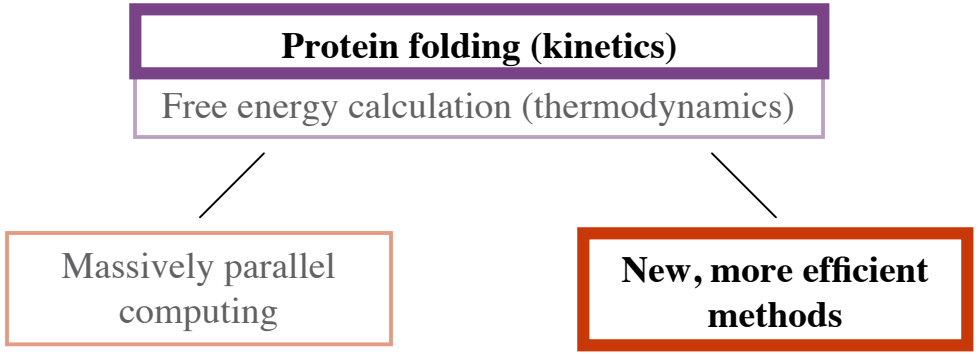
PlayStation 3

- Graphical processing unit (V. Vishal)
  - 20-30x speed up for certain simulations
  - Only fairly minor changes to client and server needed



Graphics card

Protein Folding



Aims and Topics

- Aims:
  - Achieve simulations of folding with all-atom detail
  - Develop analyses suited to data and parallelism
  - Gain biophysical insights on folding

- Contributions:
  - Examination of role of water in folding
  - Rate calculation methods
  - Fast sigmoidal dielectric implicit solvent model
  - Perturbation analyses
  - Markov models for protein dynamics
  - Local structure assessment
  - Algorithm for identifying crucial residue pairs

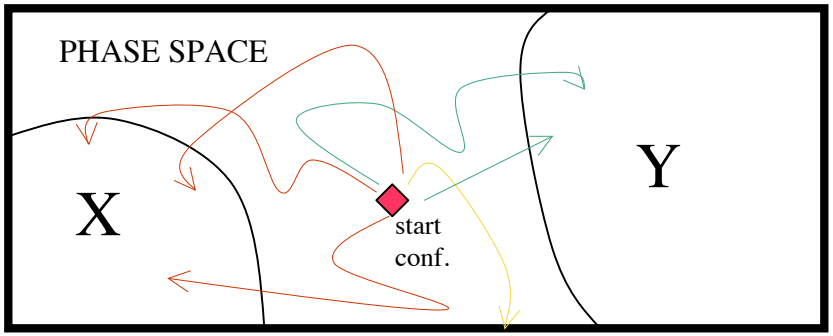
Perturbation Analysis

- Often interested in effect of a perturbation to a system
  - Perturbation can be physically meaningful (like changing temperature) or methodological

- Assessing such perturbations requires accounting for stochastic nature of dynamics
  - This requires much sampling

P(X,Y) Definition

P(X,Y) equals probability of reaching state X before state Y



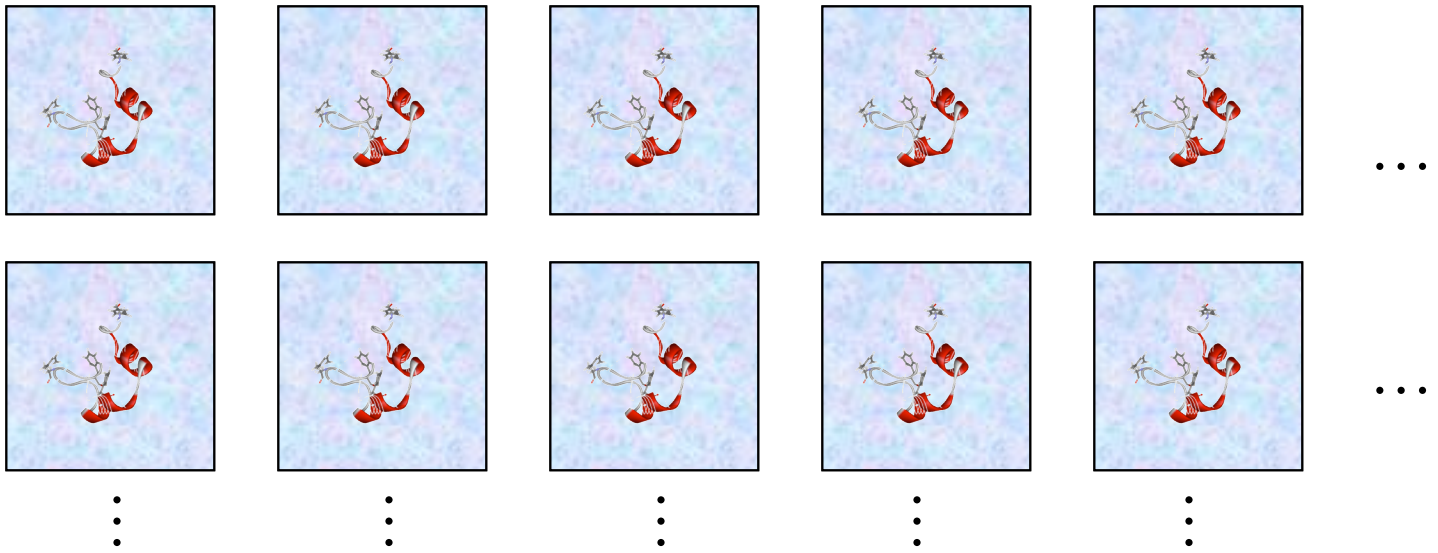
$$P(X,Y) = \frac{N(\rightarrow)}{N(\rightarrow) + N(\rightarrow)}$$

Is a generalization of P_{fold}:

P_{fold}=P(X,Y) where X='folded' and Y='unfolded'

Guha Jayachandran, V. Vishal, Vijay S. Pande. "Using massively parallel simulation and Markovian models to study protein folding..." J.C.P. (2006).

P(X,Y) Use

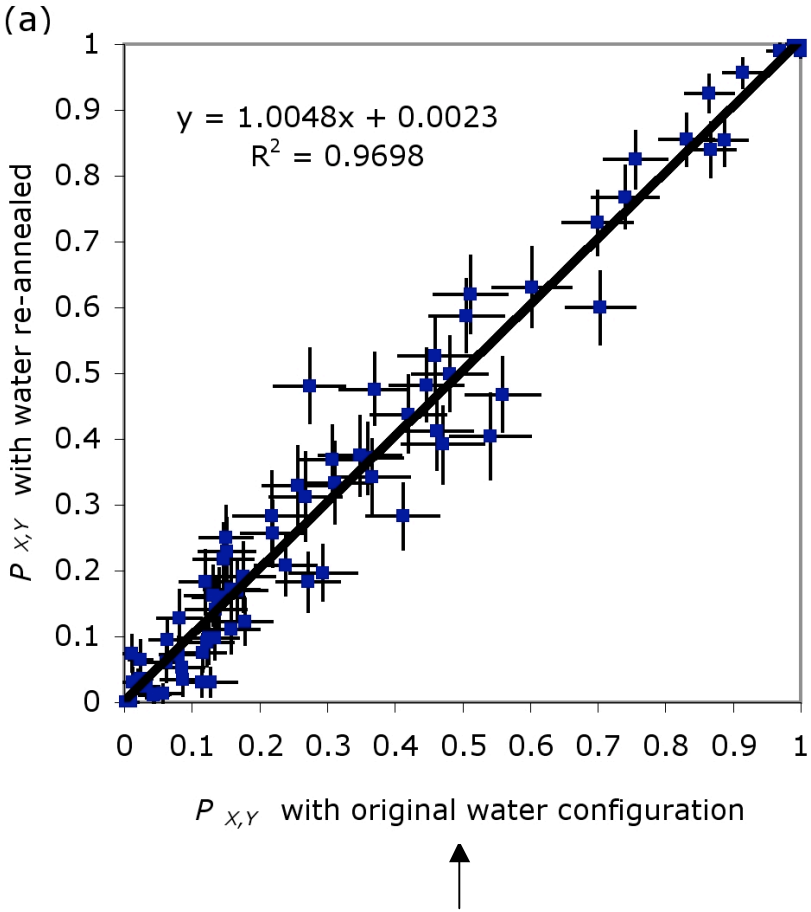


- Calculate value for a given conformation by running an ensemble of simulations and seeing what fraction reach state X before state Y
- Compute value again after a given perturbation and see if the values change, for various X and Y

Guha Jayachandran, V. Vishal, Vijay S. Pande. “Using massively parallel simulation and Markovian models to study protein folding...” J.C.P. (2006).

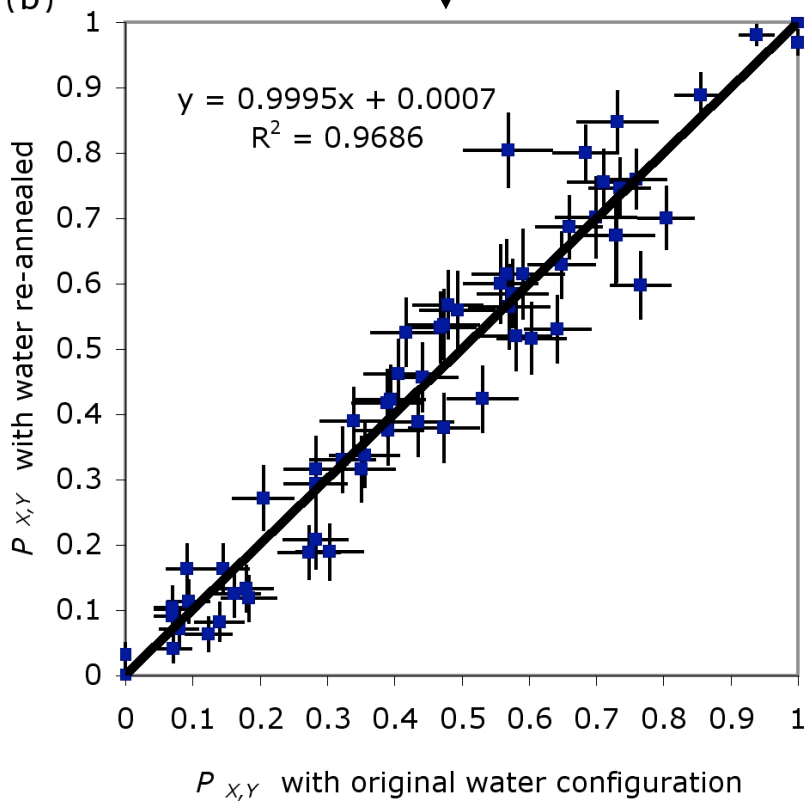
P(X,Y) and Water Configuration

Each point on graph is associated with a different protein conformation.



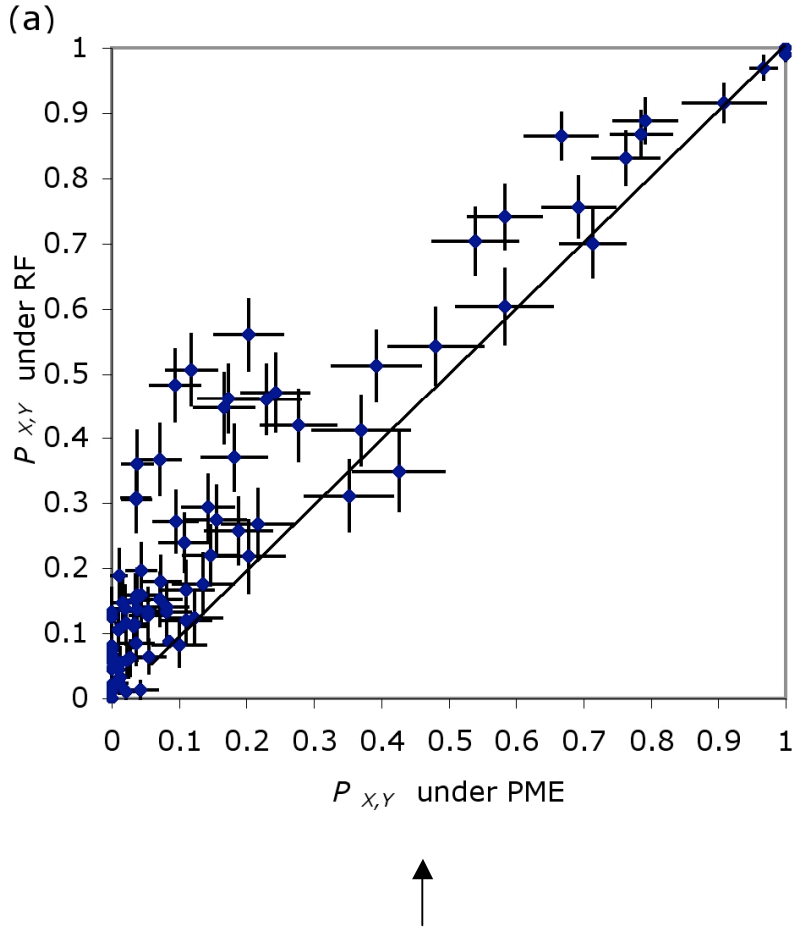
X = "<20% native contacts formed"
Y = ">40% native contacts formed"

X = "Radius of gyration < 10Å"
Y = "Radius of gyration > 13Å"



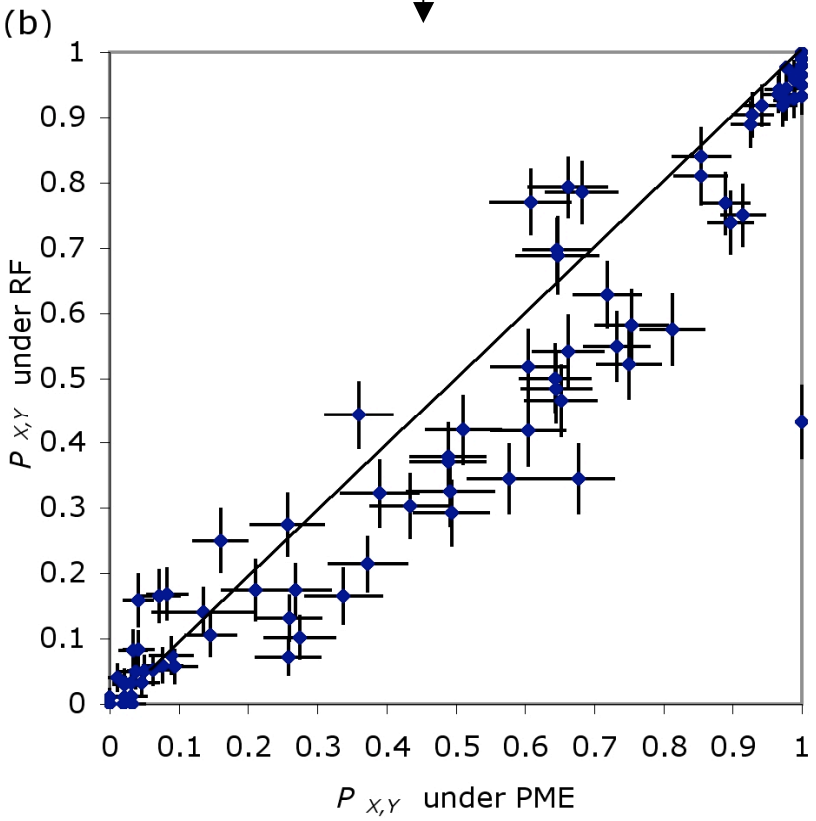
Guha Jayachandran, V. Vishal, Vijay S. Pande. "Using massively parallel simulation and Markovian models to study protein folding..." J.C.P. (2006).

P(X,Y) and Electrostatic Treatment



X = "<20% native contacts formed"
Y = ">40% native contacts formed"

X = "Radius of gyration < 10Å"
Y = "Radius of gyration > 13Å"

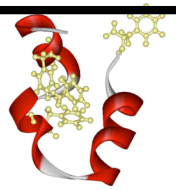


Guha Jayachandran, V. Vishal, Vijay S. Pande. "Using massively parallel simulation and Markovian models to study protein folding..." J.C.P. (2006).

Local Structure Formation

- Can local structure elements form without global structure?
  - Longtime question in protein folding and has experimental support for certain proteins
  - Related to question of “foldons”
- Can we identify the order in which elements form and what combinations are likely?

Villin and Protein A Simulations

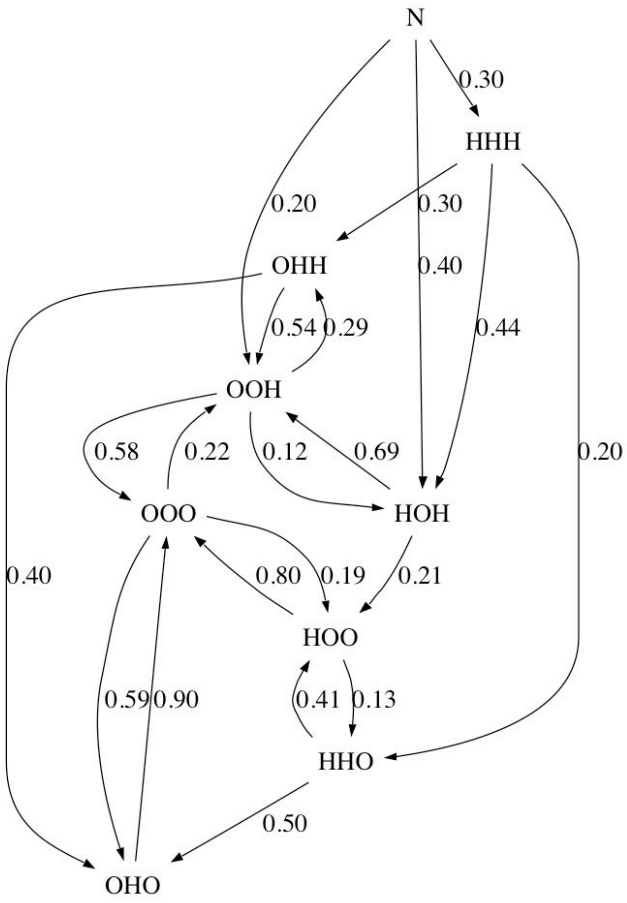
	Structure	Trajectories	Total sampling
Villin headpiece		>20,000	~500 μs
10-55 fragment B of protein A		>5,000	>100 μs

A trajectory (blue) of villin reaching the folded state from the unfolded state. The native structure is shown fixed in red.

Guha Jayachandran, V. Vishal, Angel García, Vijay S. Pande. "Local structure formation in simulations of two small proteins." J.S.B. (2006).

Protein A Helix Formation

- Fractions normalized by source

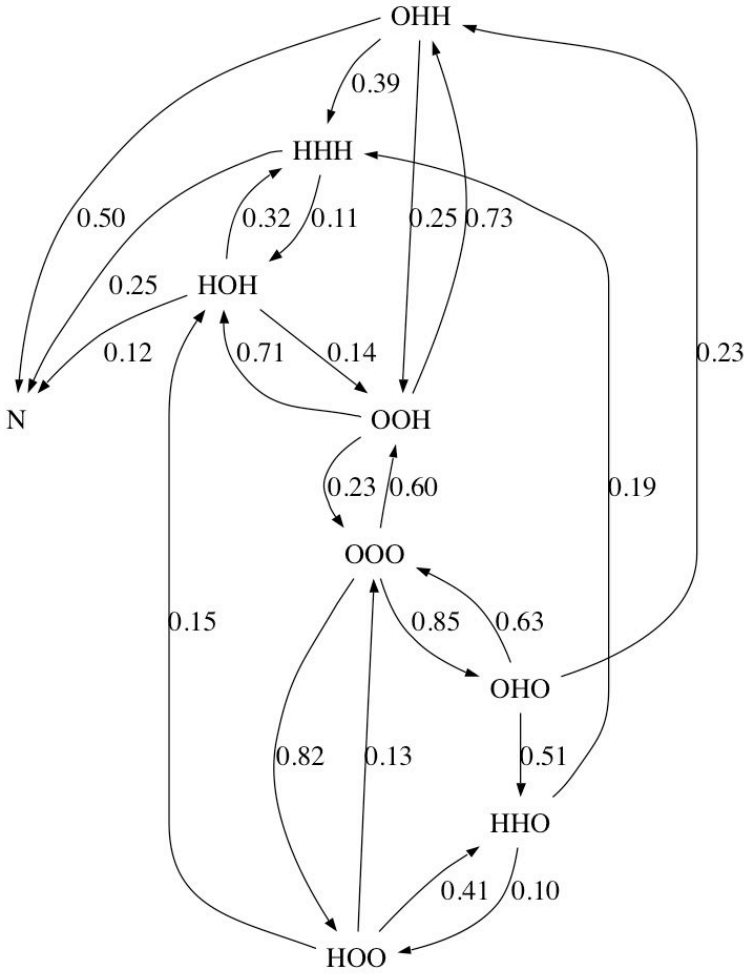


- Like for villin, observed the middle helix most commonly (approximately 60% of time) transiently forming first.
- If either the first or second helix existed in isolation, it was most often lost—in 80% and 90% of cases, respectively—before any additional helices formed. If the third helix existed in isolation, it gained the second helix in nearly 30% of cases.
- For each helix combination, the majority of observed transitions involved loss of a helix rather than gain of an additional one.

Guha Jayachandran, V. Vishal, Angel García, Vijay S. Pande. "Local structure formation in simulations of two small proteins." J.S.B. (2006).

Protein A Helix Formation

- Fractions normalized by destination

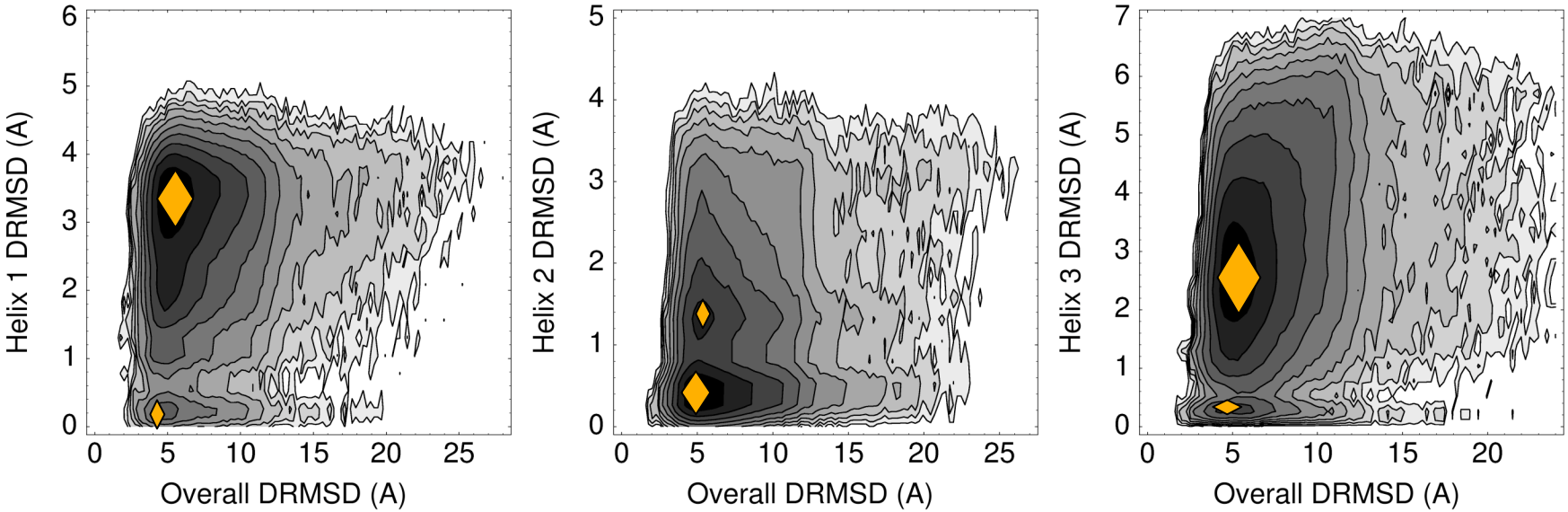


- When a state with the second two helices only was reached, 73% of the time the third helix had existed first and was then joined by the second.
- The native-like state (N) was reached 75% of the time from a state where the second two helices were present.
- Sato et al. (2004) concluded that the transition state, late in the folding process, includes a well formed second helix, agreeing with our simulation result.
- Also support from the results of García and Onuchic (2003), who claimed that the final stages of folding involved formation of helix 1 and its interactions with helix 2, and packing of an already formed helix 3 into the bundle.

Guha Jayachandran, V. Vishal, Angel García, Vijay S. Pande. "Local structure formation in simulations of two small proteins." J.S.B. (2006).

Helices in Unfolded Ensemble

- Villin



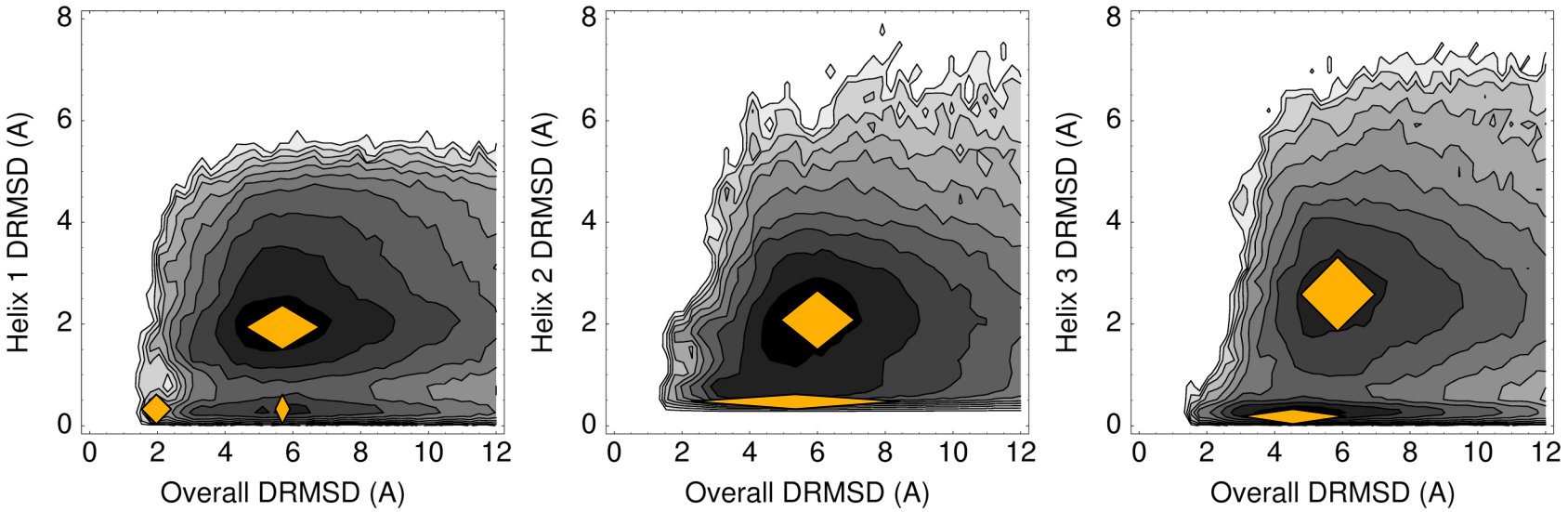
Lower DRMSD signifies more native-like.

Darker color indicates higher population.

Guha Jayachandran, V. Vishal, Angel García, Vijay S. Pande. "Local structure formation in simulations of two small proteins." J.S.B. (2006).

Helices in Unfolded Ensemble

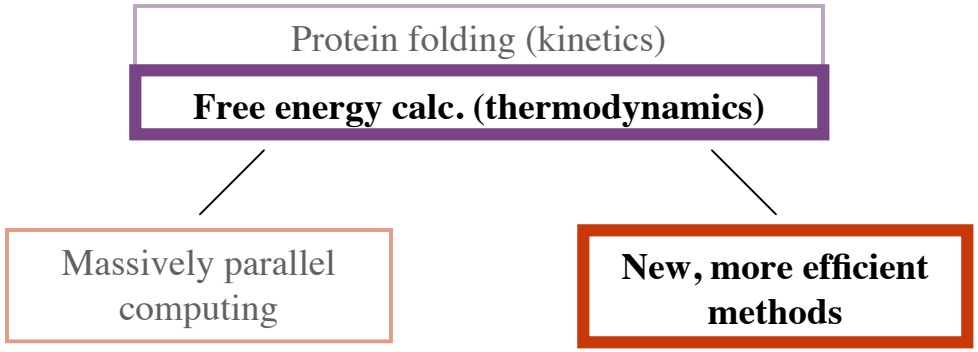
- Protein A



See low helix DRMSD even when overall structure highly nonnative.

Guha Jayachandran, V. Vishal, Angel García, Vijay S. Pande. "Local structure formation in simulations of two small proteins." J.S.B. (2006).

Free Energy Computation



Background: Free Energy

- $\Delta G = \Delta H - T\Delta S$

  Change in free energy — Change in enthalpy — Temperature — Change in entropy

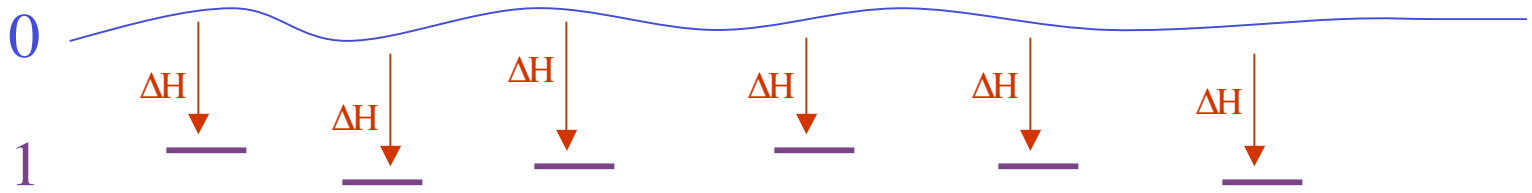
- Associated with probability
  - More negative the ΔG of a change, higher the probability of the change

- State function (pathway independent)

Background: ΔF Calculation

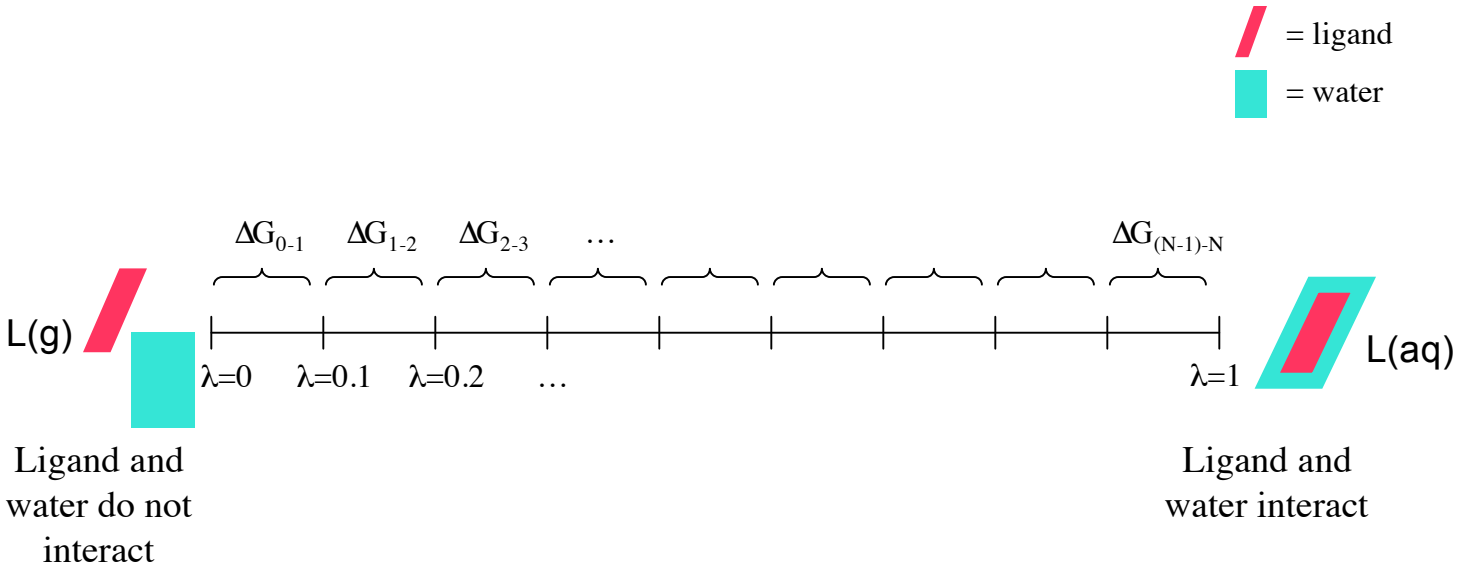
- Variety of algorithms:
  - Free energy perturbation (F.E.P.)
  - Thermodynamic integration (T.I.)
  - Slow growth

- In F.E.P., estimate equilibrium ΔG between reference and target states by exponentially averaging potential energy differences between the reference state sampled at equilibrium and the target state



$$\Delta G_{0\rightarrow 1} = -kT \ln \langle e^{-\Delta H/kT} \rangle_{eq.}$$

Background: Solvation Free Energy



$$\Delta G = \sum_{i=0}^{N-1} \Delta G_{\lambda_i \rightarrow \lambda_{i+1}}$$

Background: Absolute Binding Free Energy

$$L(aq) \rightarrow \cancel{L(g)} \qquad \text{(ligand desolvation)}$$

$$+ \quad P(aq) + \cancel{L(g)} \rightarrow P{:}L(aq) \qquad \text{(coupling)}$$

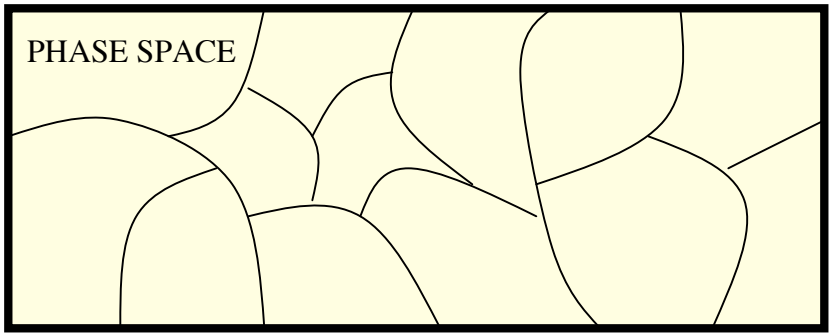
$$L(aq) + P(aq) \rightarrow P{:}L(aq) \qquad \text{(binding)}$$

New Method

- Aims:
  - Accurate
  - Fast
  - Work without known structure for bound complex

- Basis:

  Divide configurational space into states and find free energy change in each independently



Guha Jayachandran, Michael Shirts, Sanghyun Park, Vijay Pande. "Parallellized over parts computation of absolute binding free energy with docking and molecular dynamics." J.C.P. (2006).

Mathematical Foundation

Free energy of interest

Sum over all states

Limit from all states to just one state

Free energy for change within state s

$$\Delta G(0,U \rightarrow 1,U) = -kT \ln \sum_{s \in R} \left[\exp\left(\left(-\Delta G(0,U \rightarrow 0,s) - \sum_{i=0}^{N} \Delta G(\lambda_i, s \rightarrow \lambda_{i+1}, s) \right) / kT \right) \right]$$

U = All relevant phase space

k = gas constant

T = Temperature

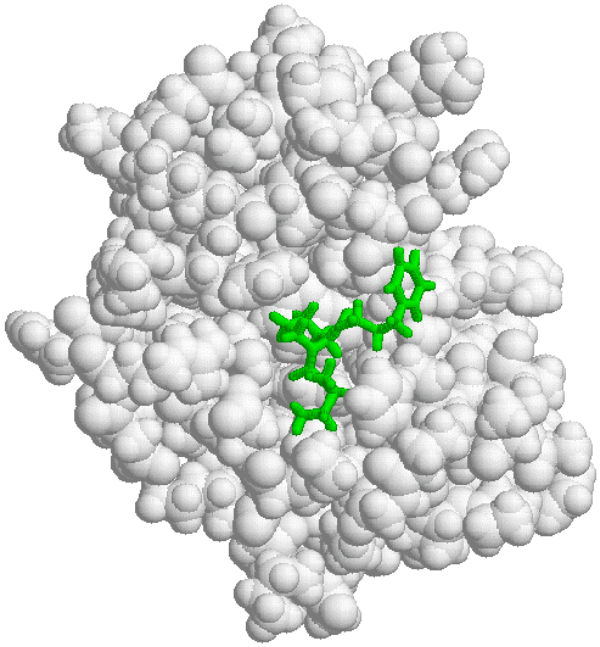
R = Set of mutually exclusive states s that collectively cover U

If parallelize over states,
at least a $|R|$ times decrease in wall clock time
over traditional F.E.P.
(in addition to N times decrease from parallelizing over lambda intervals)

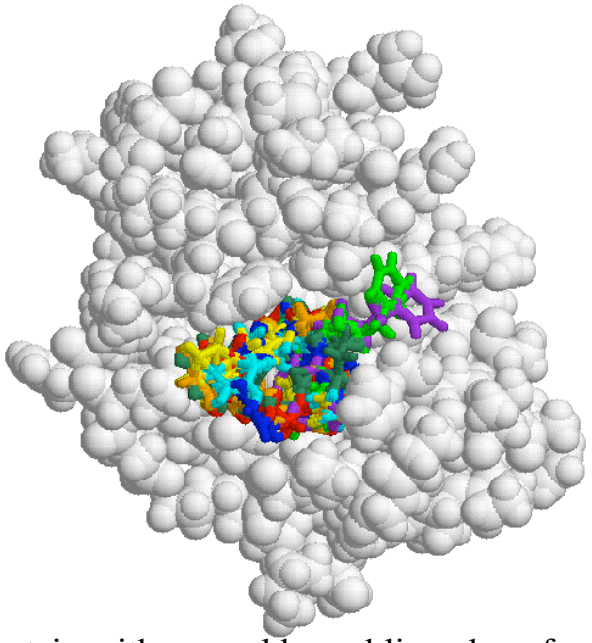
Guha Jayachandran, Michael Shirts, Sanghyun Park, Vijay Pande. "Parallellized over parts computation of absolute binding free energy with docking and molecular dynamics." J.C.P. (2006).

Steps

- Step 1: Generate variety of initial configurations
  - For complex, use docking to come up with diverse binding poses



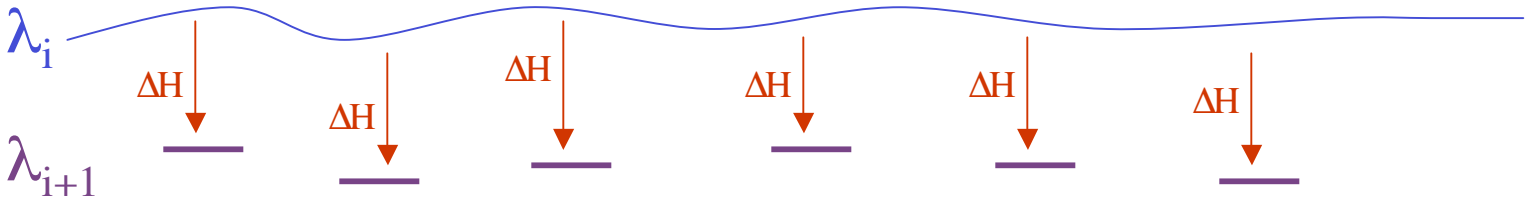
Protein with ligand (green) bound



Protein with several bound ligand conformations shown together (each different color)

Steps

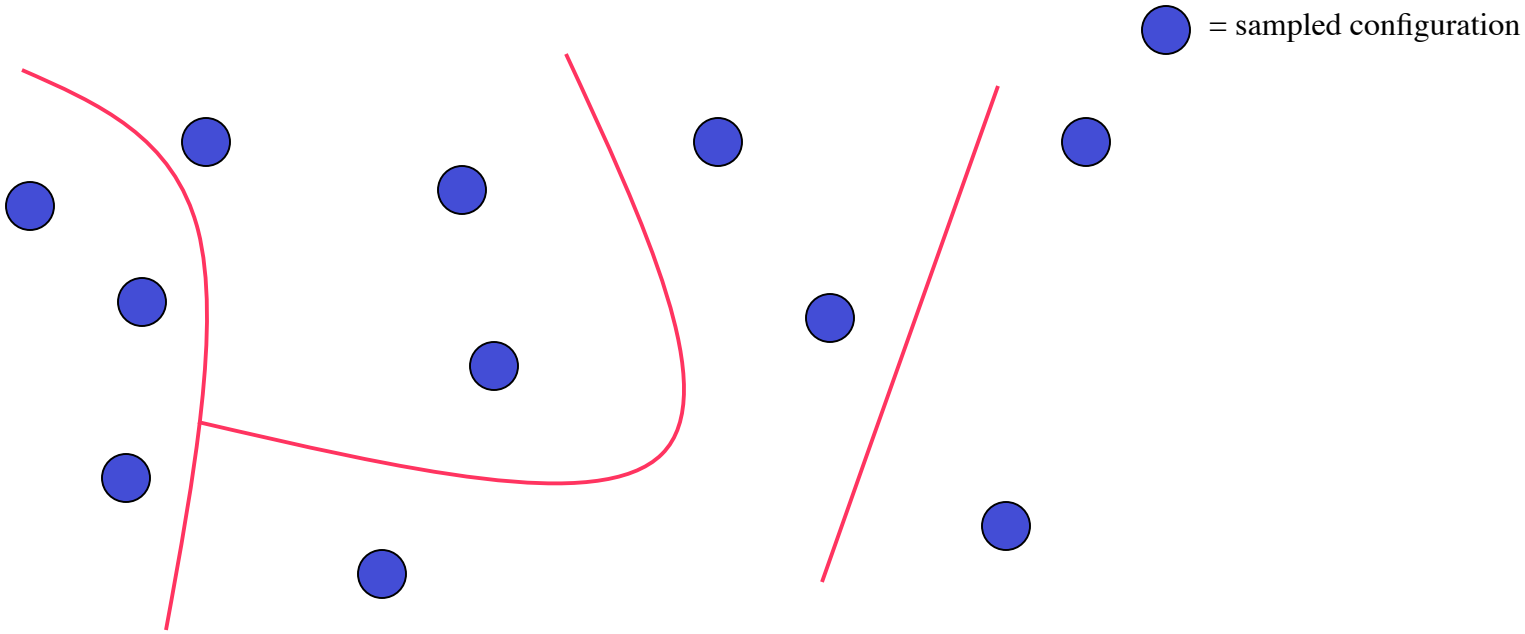
- Step 2: Molecular dynamics simulations from each initial configuration from Step 1
  - Do this at range of lambda values, measuring energy differences



Guha Jayachandran, Michael Shirts, Sanghyun Park, Vijay Pande. "Parallellized over parts computation of absolute binding free energy with docking and molecular dynamics." J.C.P. (2006).

Steps

- Step 3: Assign sampled configurations from Step 2 into configurational states
  - Hierarchical clustering along slow kinetic degrees of freedom



Guha Jayachandran, Michael Shirts, Sanghyun Park, Vijay Pande. "Parallellized over parts computation of absolute binding free energy with docking and molecular dynamics." J.C.P. (2006).

Steps

- Step 4: Determine restriction free energy change of restricting to each state *s*
  - $\Delta G(0,U \rightarrow 0,s)$
  - Best approach not clear
  - One way: stationary distribution of Markov model constructed with data at λ=0

Guha Jayachandran, Michael Shirts, Sanghyun Park, Vijay Pande. "Parallellized over parts computation of absolute binding free energy with docking and molecular dynamics." J.C.P. (2006).

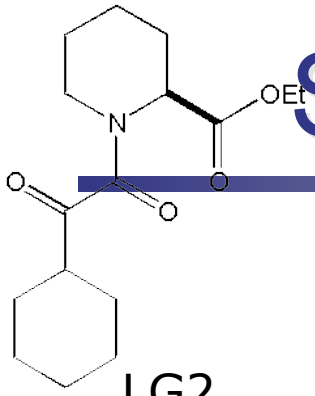
Steps

- Step 5: Calculate overall free energy with results of previous steps
  - Calculate free energies for each state/lambda interval pair
  - Combine together using

$$\Delta G(0,U \rightarrow 1,U) = -kT \ln \sum_{s \in R} \left[\exp\left(\left(-\Delta G(0,U \rightarrow 0,s) - \sum_{i=0}^{N} \Delta G(\lambda_i, s \rightarrow \lambda_{i+1}, s) \right) / kT \right) \right]$$

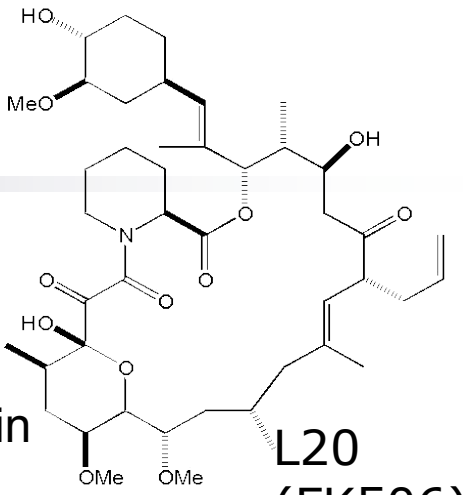
Guha Jayachandran, Michael Shirts, Sanghyun Park, Vijay Pande. "Parallellized over parts computation of absolute binding free energy with docking and molecular dynamics." J.C.P. (2006).

Sample System

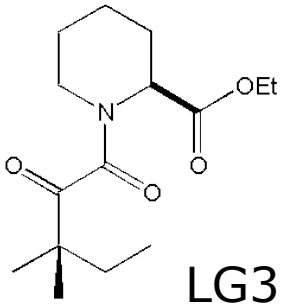


LG2

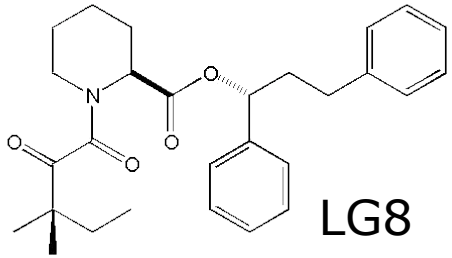
- FKBP12: a canonical system for binding calculations and experiments
- 107 residues, stable, well studied, involved in immune response
- Experimental data from Holt *et al*, *JACS* (1993)



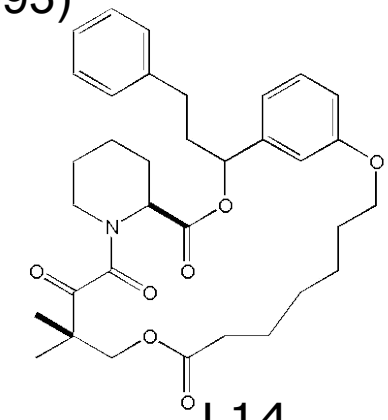
L20 (FK506)



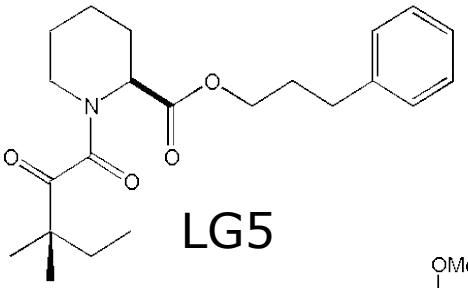
LG3



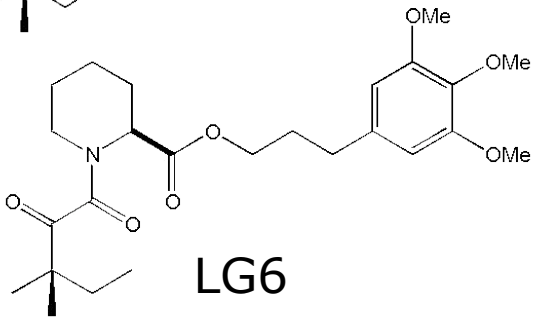
LG8



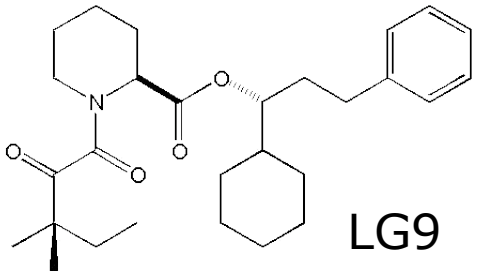
L14



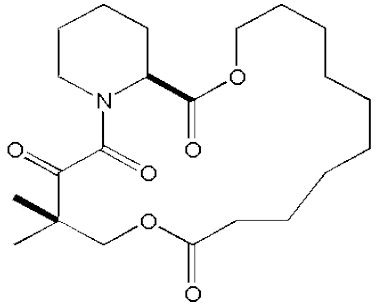
LG5



LG6

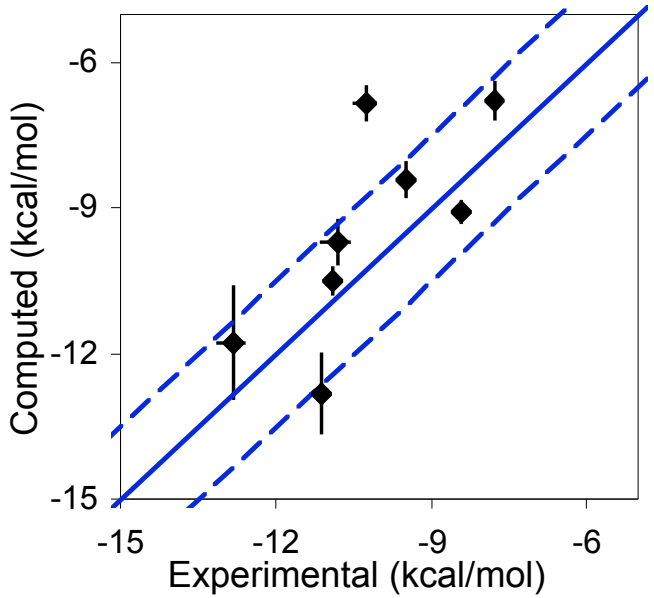


LG9



L12

Results



y=x and y=x±1.5
drawn as guides

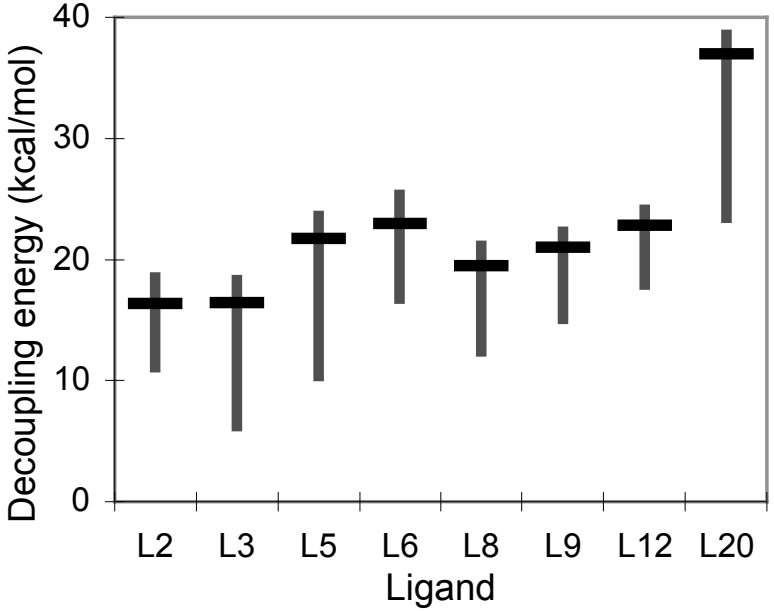
1.6 kcal/mol RMSD between experimental and computed values

1.1 kcal/mol excluding L12

Guha Jayachandran, Michael Shirts, Sanghyun Park, Vijay Pande. "Parallellized over parts computation of absolute binding free energy with docking and molecular dynamics." J.C.P. (2006).

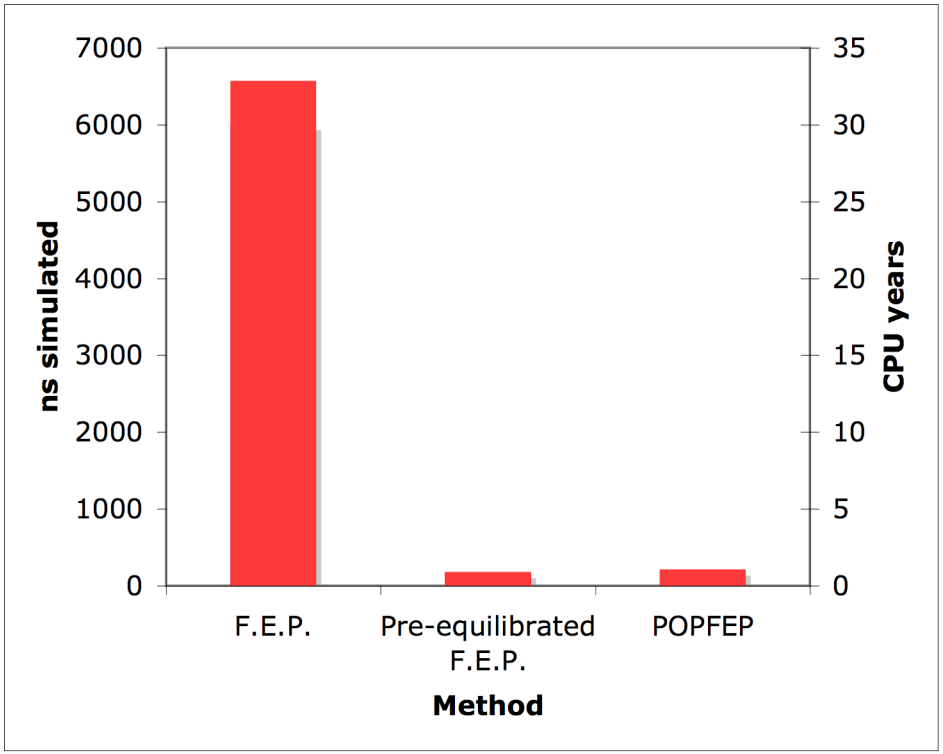
Per State Free Energy Ranges

Vertical lines go between minimum and maximum decoupling free energies measured for individual configurational states. Horizontal lines denote combined results.



Guha Jayachandran, Michael Shirts, Sanghyun Park, Vijay Pande. "Parallellized over parts computation of absolute binding free energy with docking and molecular dynamics." J.C.P. (2006).

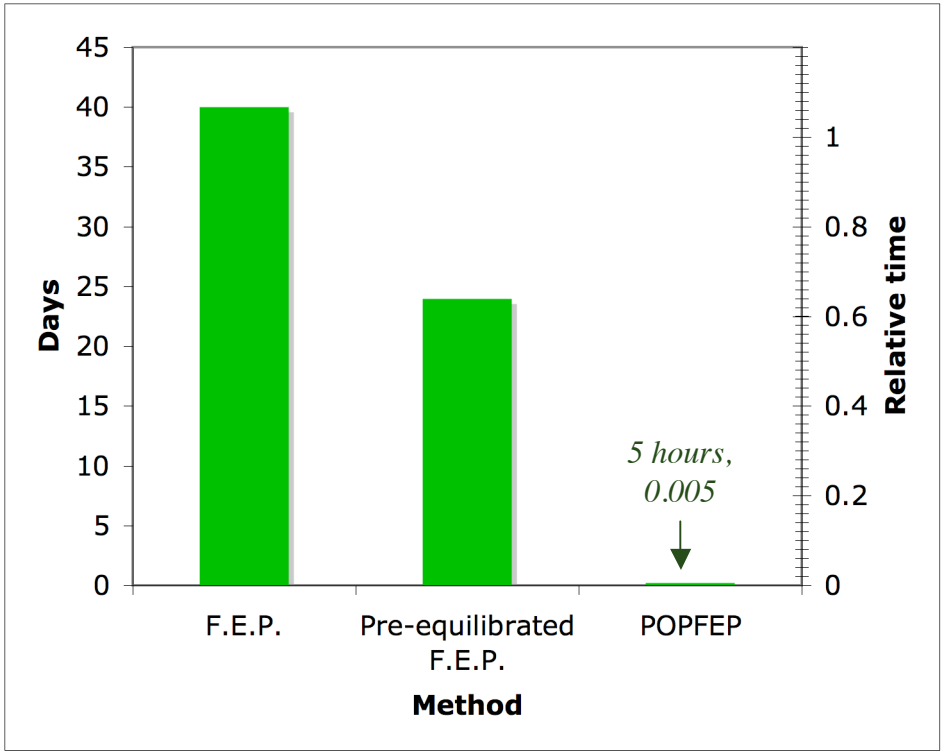
Simulation and CPU Time



For 1 binding free energy.

Guha Jayachandran, Michael Shirts, Sanghyun Park, Vijay Pande. "Parallellized over parts computation of absolute binding free energy with docking and molecular dynamics." J.C.P. (2006).

Rate Limiting Time



Limiting step sets a lower bound on wall clock time.

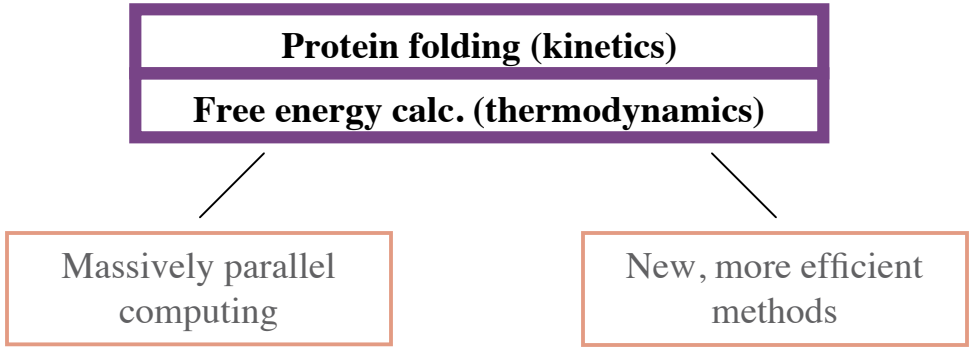
With 2000 processors, wall clock time approximately equal to above figures.

Guha Jayachandran, Michael Shirts, Sanghyun Park, Vijay Pande. "Parallellized over parts computation of absolute binding free energy with docking and molecular dynamics." J.C.P. (2006).

Limitations

- Results sensitive to ligand parameters
- Charged systems require still more sampling
- Complexes where the protein undergoes conformational change upon binding can be hard
  - One approach to this may involve defining configurational states including the protein degrees of freedom

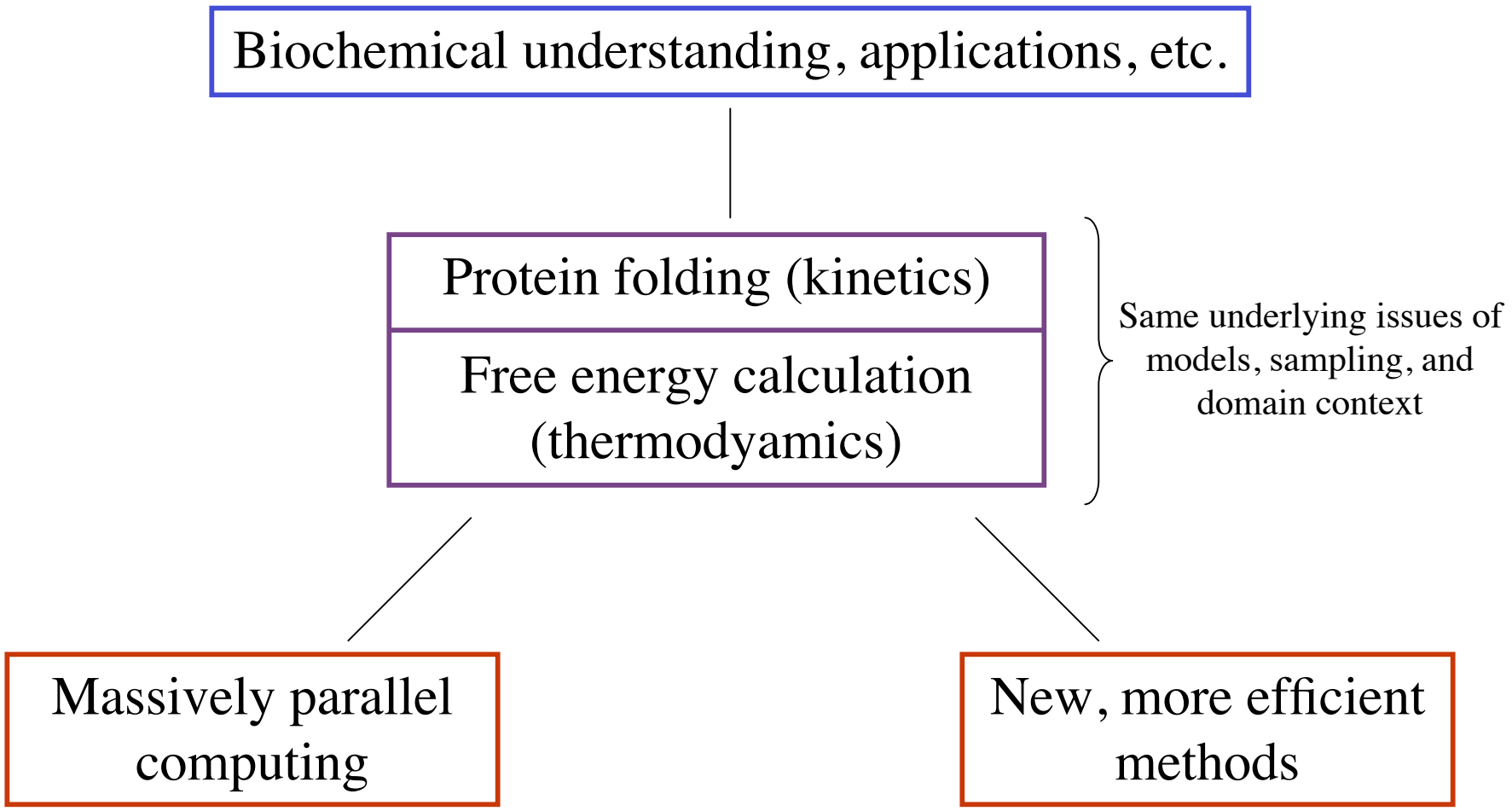
Conclusion


Protein folding (kinetics)
Free energy calc. (thermodynamics)
Massively parallel computing
New, more efficient methods

Summary

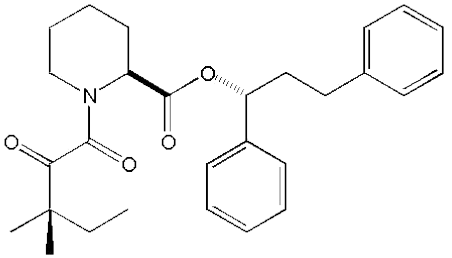
- Distributed computing
  - Several improvements to all components of worldwide distributed computing infrastructure
- Protein folding kinetics
  - Rate calculation methods
  - Fast sigmoidal dielectric implicit solvent model
  - Perturbation analyses
  - Construction and applications of Markov models
  - Local structure assessment
- Protein-ligand binding thermodynamics
  - New algorithm allowing computation of free energies faster and more accurately than prior methods

Means and Motivations



Future and Applications

- Modeling and simulation of varied disease/agent-relevant proteins (folding, aggregation, and binding)
- Large scale virtual screening of small molecule drugs
- Lead discovery and refinement

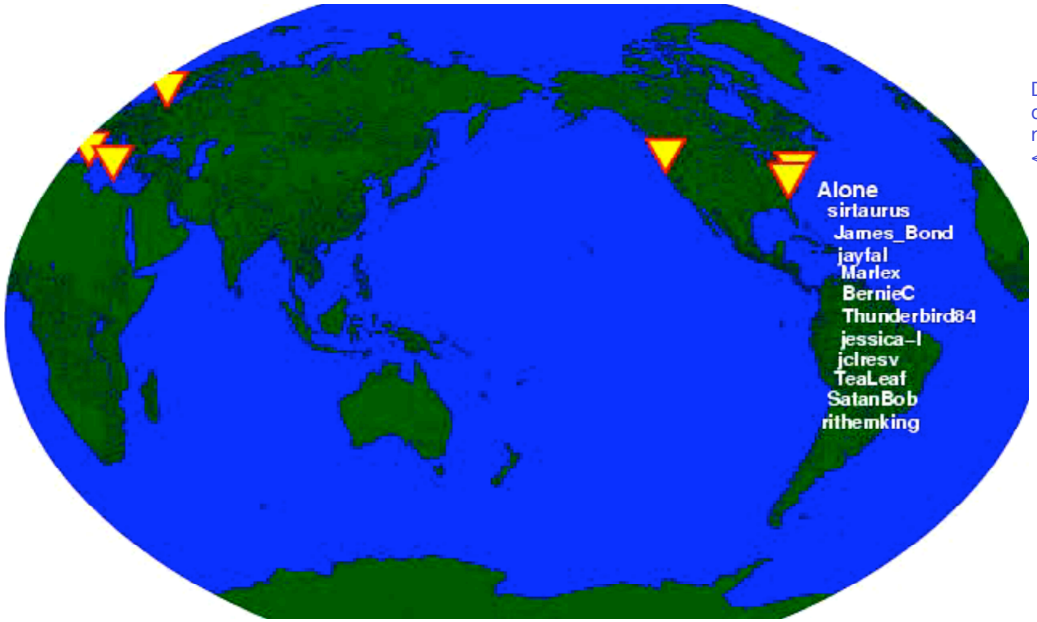


?

Acknowledgments

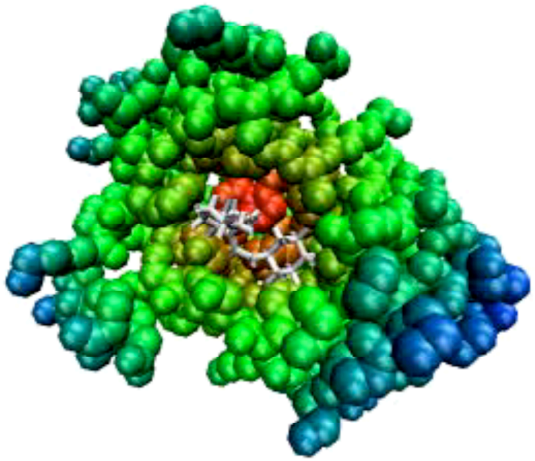
- Prof. Vijay Pande

- Prof. Jean-Claude Latombe
- Prof. Serafim Batzoglou
- Prof. Tom Wandless
- Prof. Ashish Goel
- Former fellow group members
  - Bojan Zagrovic
  - Christopher Snow
  - Eric Sorin
  - Lillian Chong
  - Mark Engelhardt
  - Michael Shirts
  - Sidney Elmer
  - Siraj Khaliq
  - Stefan Larson
  - Sung-Joo Lee
  - Young Min Rhee
- Current fellow group members
  - Adam Beberg
  - Dan Ensign
  - Del Lucent
  - Edgar Luttmann
  - Jason Wagoner
  - Jeremy England
  - John Chodera
  - Kim Branson
  - Nick Kelley
  - Nina Singhal
  - Paula Petrone
  - Relly Brandman
  - Sanghyun Park
  - Vishal Vaidyanathan
- Funding
  - N.D.S.E.G. Fellowship
  - N.S.F./N.I.H. grants
- Folding@Home participants worldwide
  - *Listing at folding.stanford.edu*



Distributed computing map ←

Parallel Computing Methods for Probing Biomolecular Kinetics and Thermodynamics



Guha Jayachandran
Stanford University
February 12, 2007

GuhaJ@cs.stanford.edu